**Uptake of ORCIDs Among Researchers**

Nicole Gray, Ali Krzton, Yee Sern Tan, and Yarid Timossini

**Abstract:** ORCID is a persistent-identifier system that is intended for use by the research community. The EU-funded THOR project is interested in the spread of ORCIDs throughout the world and within particular research disciplines. We analyzed data from 2012 to 2016 on the cumulative total number of authors with ORCIDs (persorcid) and publications linked to those ORCIDs (pubsorcid) to determine trends in ORCID use throughout the scholarly community. After cleaning and sorting the data using Excel and Tableau, we visualized the data using a combination of Tableau and Google Charts tools. We found that ORCID use is greatest in the applied sciences, physical sciences, and life sciences, and that the countries which were the earliest adopters of ORCIDs continued to maintain their lead in ORCID use through 2016.

**Keywords:** ORCID, scholarly communication, chloropleth map, stacked line graph

**Introduction**

An ORCID is a type of persistent identifier (PID), in this case for academic researchers. ORCIDs can provide each researcher with a unique, stable identifier that improves the ability to detect and analyze connections among both areas of research and researchers themselves under a common, trusted framework [1]. The use of ORCIDs or other types of persistent identifiers is voluntary and depends on researcher commitment to the system and its goals. The EU-funded THOR Project, our client, is interested in understanding what makes researchers participate in PID systems and identifying any trends in ORCID adoption, particularly with respect to geography and area of research.

The purpose of our analysis is to generate insights about patterns in ORCID uptake, such as geographic clusters, or an acceleration of the adoption rate in a particular field of study, that will help THOR and other initiatives for open science make decisions about how to encourage the use of PIDs. This has the potential to make searches for raw data and related research much easier, speeding up the pace of innovation.

**1. Background**

The adoption of ORCIDs has the potential to greatly reduce ambiguity in the identity of researchers across various fields. For instance, many names are common to hundreds or even thousands of people; the ORCID system can prevent one author being mistaken for another of the same name [2,3]. ORCIDs are also useful for improving the discoverability of all of a researcher’s work, including datasets that do not appear in a publication, or a preprint from a repository [3]. In contrast with other systems that have arisen to identify researchers, ORCID is not proprietary or tied to a specific company [4], which provides an additional level of openness and stability.

* Nicole Gray, School of Informatics and Computing, Indiana University Bloomington

E-mail: [nlmenser@umail.iu.edu](mailto:nlmenser@umail.iu.edu)

* Ali Krzton, School of Library and Information Science, Kent State University

E-Mail: [akrzton@kent.edu](mailto:akrzton@kent.edu)

* Yarid Timossini, School of Informatics and Computing, Indiana University Bloomington

E-Mail: [ytimossi@umail.iu.edu](mailto:ytimossi@umail.iu.edu)

* Yee Sern Tan, School of Informatics and Computing, Indiana University Bloomington

E-mail: [yeestan@indiana.edu](mailto:yeestan@indiana.edu)

Previous work concerning geographic influences on research community habits has focused on citation networks. For instance, in the USA, distance between co-authors was not increasing over time, despite the rapid adoption of the Web and its communications capabilities [5]. ORCID use may spread in a similar manner between researchers that are known to each other.

A study of the spread of influence and how opinions change also used maps to display a landscape of possible decision outcomes [6], but this was a map in variable space, rather than one with a geographic reference system. A map of ORCID uptake in physical space to understand how they are, or are not, spreading worldwide could be further analyzed using these methods to check for proximity effects.

**2. Discussion**

*2.1 Characteristics of the Dataset*

As the data associated with ORCIDs is publically available, we initially began by trying to characterize the entire dataset for the two most recent years in XML and JSON format. However, the size of this dataset was prohibitive (over 200 GB) and after speaking with the client it was determined that all of the data was not necessary to conduct a thorough analysis of their topics of interest. The client provided the group with an organized subset of the data in an Excel file. Two metrics, persorcid and pubsorcid, were listed as quarterly cumulative count data from the time period beginning in the fourth quarter of 2012 and ending in the third quarter of 2016. Persorcid signifies the number of authors who have signed up for ORCIDs, and pubsorcid signifies the number of publications listed in the Web of Science database that are associated with a registered ORCID. Records were listed according to country of origin and Web of Science topic ID, as well as seven broad client-assigned topical categories.

The spreadsheet provided by the client did not retain any specific author data, meaning that network analysis of collaborations was not possible using this dataset. However, listings of up to a hundred co-authors for a single article were one reason that the original, raw ORCID JSON file was so large. Our analysis therefore focused around trends in the count data for persorcid and pubsorcid by country and by topical category.

*2.2 Data Cleaning and Analysis*

The count data received from the client was already refined, but we performed additional cleaning to correct minor errors (for instance, data for the Netherlands could not be properly aggregated until a misspelling, ‘Neitherlands’, was corrected for all rows in which the variant occurred) and then subdivided the data. Pivot tables in Excel and category breakdowns in Tableau were used to aggregate counts by country and by topical category; these tables were transferred to Google Sheets to facilitate sharing and to allow them to be linked to Google Charts. For analyses that compared persorcid and pubsorcid across years instead of quarters, the total for the quarters pertaining to that year were used (the fourth, eighth, twelfth, and sixteenth columns of the cumulative data, respectively).

We ranked countries by yearly persorcid and pubsorcid totals, deriving top ten lists over all topical categories and by individual category (see Appendix). Since the top ten countries tended to remain the same, with only a few changing rank over the course of the four-year period, and because there was no sharp discontinuity between the top ten countries and the rest of the data, we opted to use the full dataset in subsequent analyses. We also animated a plot of the change in the top ten countries’ the number of persorcids vs. the number of pubsorcids (Fig. 1). The relationship was roughly linear, indicating that the overall trends in growth of persorcids and pubsorcids were not very different.

As growth in persorcid was less stochastic and provided greater contrast than pubsorcid, we chose to use persorcid in our final geographic visualizations. We also opted not to try to combine the geographic and the topical analyses, as we could not find a suitable way to do this, nor any obvious trends/interactions that would justify the added complexity.

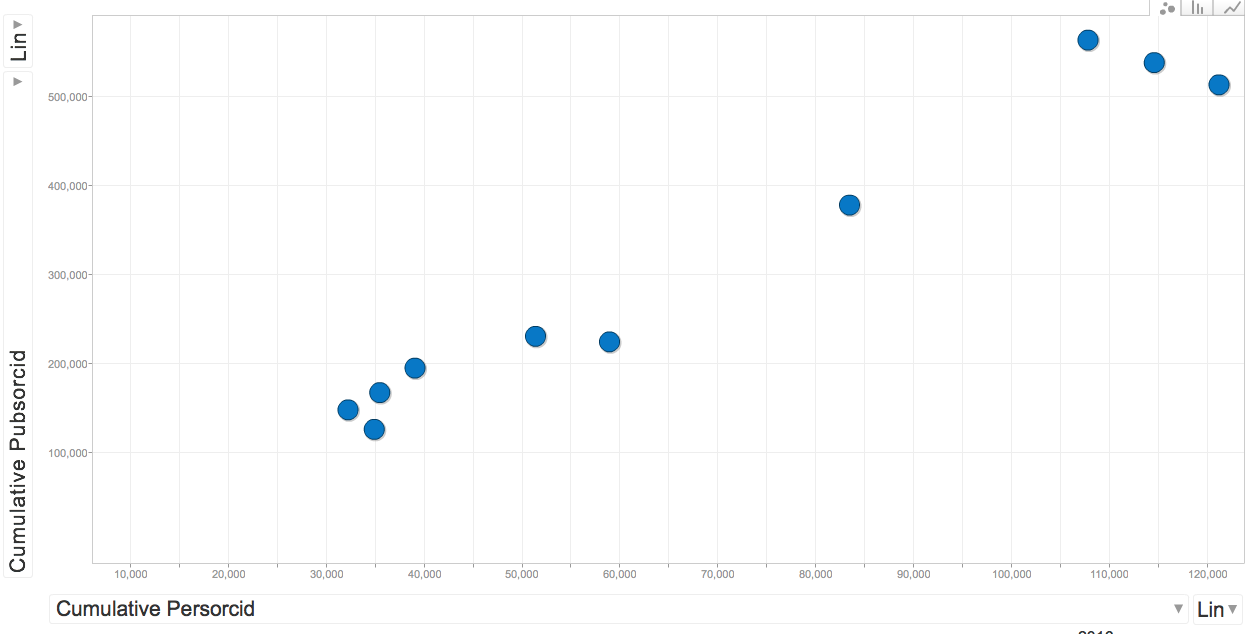


Figure 1. A scatterplot of the values for persorcid vs. pubsorcid in the top ten countries in 2016 (the final frame of animation covering the change from 2013 to 2016).

*2.3 Visualizations*

Initially, lat longs were assigned to represent the geographic center of each country so that a proportional symbol map could be generated in Sci2 for the country analysis. However, this visualization was not appropriate for our analyses by country because the symbols often overlapped and obscured country borders (Fig. 2); another downside was that these maps lacked interactivity. We switched to Google Charts for our geographic visualizations, although we continued to use a mixture of tools (Excel, Tableau, and Google) for other analyses and topical visualizations.

To make the final geographic visualization in Google Charts, we used a Google Sheet with the cumulative annual persorcid totals for 2013-2016 by country. We standardized country names to ensure they would be recognized by the Google algorithm, then embedded javascript to query the sheets and draw the resulting charts into HTML (see Appendix). This resulted in interactive maps that allow the user to see values by country in a tooltip. Further, an animated GIF was made to show the change in persorcid worldwide over the entire period.

We decided that interactivity was not required for the topical analysis and opted for a stacked line graph showing the contribution of each category to the total for both persorcid and pubsorcid. We also retained the more fine-grained quarterly analysis for this visualization, as opposed to year-over-year. The graph was created in Tableau Desktop and exported as a PDF.

The final project visualizations are available at:

<http://visualizations.thoughtsofakat.com/animap.html>

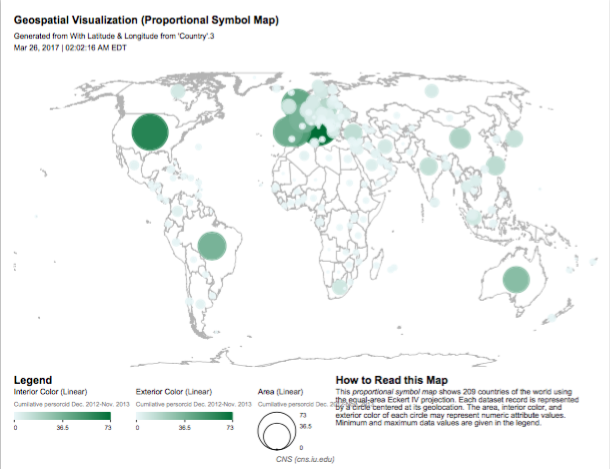


Figure 2. Proportional symbol map generated by Sci2, unsuitable due to its obscuring many European countries.

**3. Conclusion**

The visualizations we produced reflect the trends in the data we judged to be the most important to the client, based on their needs. We elected to retain a basic design to allow for maximum readability, built interactivity into our geographic analysis, and broke down topical trends while still presenting the totality of the author and publication data.

With regards to worldwide patterns of ORCID adoption by authors, the changes in the map over the course of 2013-2016 reflect a sharp inequality in the adoption of ORCIDs worldwide. Although nearly every country in the world gains registered authors, the countries that have the most registered authors in 2013 have widened their lead over the rest of the world by 2016. This trend is apparent despite the fact that ORCIDs would be expected to reach a saturation point in the countries with the most researchers at some future time. Researchers based in countries with a mature, rather than a developing, scholarly community may be expected to adopt ORCIDs at a greater rate. On the other hand, there is no guarantee that ORCIDs will not be “orphaned” or that none of the growth is due to duplication. It remains to be seen how many of the registered authors contribute to real growth in linked publications.

From the perspective of research categories, growth is dominated by the applied sciences, followed closely by the physical and life sciences, in both authors and publications. The growth rate in both appears largely homogenous, with the exception of a small spike in publications added in late 2015 and the beginning of 2016. This may be the first burst of publications from research commitments made within the study timeframe, as opposed to before it. The European data, being the most active overall continent for ORCIDs and potentially of interest to our client, matches the broader trends observed worldwide.

Our analysis shows that ORCIDs are most established in the sciences (broadly) and in countries that already have extensive scholarly networks. Future analyses could fruitfully pursue questions of institutional affiliation or professional seniority to determine whether these characteristics influence a researcher’s propensity to adopt an ORCID.

**References**

[1] “Our Mission.” ORCID. <https://orcid.org/about/what-is-orcid/mission>

[2]L.L. Haak, M. Fenner, L. Paglione, E. Pentz, and H. Ratner. “ORCID: A system to uniquely identify researchers.” *Learned Publishing*, vol. 25, no. 4, pp. 259-264, Oct. 2012. doi: 10.1087/20120404

[3]A. Anstey. “How can we be certain who authors really are? Why ORCID is important to the British Journal of Dermatology.” *British Journal of Dermatology*, vol. 171, no. 4, pp. 679-680, Oct. 2014. doi: 10.1111/bjd.13381

[4]A.Y. Gasparyan, N.A. Akazhanov, A.A. Voronov, and G.D. Kitas. “Systematic and Open Identification of Researchers and Authors: Focus on Open Researcher and Contributor ID.” *Journal of Korean Medical Science*, vol. 29, no. 11, Nov. 2014. doi:10.3346/jkms.2014.29.11.1453

[5]K. Borner, S. Penumarthy, M. Meiss, and W.M. Ke. “Mapping the diffusion of scholarly knowledge among major U.S. research institutions.” *Scientometrics,* vol. 68, no. 3, pp. 415-426, Dec. 2006.

[6]M. Moussaid, J.E. Kammer, P.P. Analytis, and H. Neth. “Social influence and the collective dynamics of opinion formation.” *PLoS ONE*, Nov. 2013. doi:10.1371/journal.pone.0078433

**Appendix**

HTML code for the final pages:

animap.html

<!doctype html>

<html lang="en">

<head>

<meta charset="utf-8">

<title>Trends in ORCID Adoption, 2013-2016</title>

<style>

#wrapper { width: 90%; margin-left: auto; margin-right: auto; }

</style>

</head>

<body>

<div id="wrapper">

<h2 id="top">Growth in Authors with ORCIDs by Country, 2013-2016</h2>

<figure>

<img src="orcidauthorsanimated.gif" alt="Animated World Map" width="900" height="500">

<figcaption>Number of Authors</figcaption>

</figure>

<br>

<p>View the component maps:</p>

<ul>

<li><a href="AuthorID2013.html">Interactive 2013 Map</a></li>

<li><a href="AuthorID2014.html">Interactive 2014 Map</a></li>

<li><a href="AuthorID2015.html">Interactive 2015 Map</a></li>

<li><a href="AuthorID2016.html">Interactive 2016 Map</a></li>

</ul>

<p>The changes in the map reflect [...]<p>

<br>

<h2>ORCID Authors and Linked Publications by Topic, 2013-2016</h2>

<figure>

<a href="Orcid\_topical\_final.pdf"><img src="topicalviz.png" alt="Stacked Line Graph by Topic" width="938" height="461"></a>

<figcaption>Click the figure to view the PDF.</figcaption>

</figure>

<br>

<p>Above, both cumulative author ORCIDs and [...]<p>

<br>

<br>

<br>

<a href="animap.html#top">Back to Top</a>

</div>

</body>

</html>

AuthorID2016.html (all four Google map pages are the same except for the columns queried)

<!doctype html>

<html lang="en">

<head>

<meta charset="utf-8">

<title>Author Adoption of ORCIDs 2016</title>

<script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>

<script type="text/javascript">

google.charts.load('current', {'packages':['geochart']});

google.charts.setOnLoadCallback(draw2016Map);

function draw2016Map() {

var query = new google.visualization.Query('https://docs.google.com/spreadsheets/d/1Ju1dlHJhuK5t855g-odLHgYn1Sdv0FdmSpeVxxehGJQ/gviz/tq?tq=select%20A%2C%20E');

query.send(handleQueryResponse);

}

function handleQueryResponse(response) {

var data = response.getDataTable();

var options = {

colorAxis: {minValue: 0, maxValue: 121215, colors: ['#ffffff', '#008888']}

};

var chart = new google.visualization.GeoChart(document.getElementById('authorid2016'));

google.visualization.events.addListener(chart, 'ready', function () {

document.getElementById('png').outerHTML = '<a href="' + chart.getImageURI() + '">Printable version</a>';

});

chart.draw(data, options);

}

</script>

</head>

<body>

<h2>Number of Authors with ORCIDs by Country, 2016</h2>

<div id="authorid2016" style="width: 900px; height: 500px;"></div>

<div id="png"></div>

<a href="animap.html">Return to Main Page</a>

</body>

</html>

Persorcid growth by year:

<https://docs.google.com/spreadsheets/d/1aIZn9XhCzvKrNRLMEaKE6AK6ACcXyORHpL4MrP9_3AI/edit?usp=sharing>

Top 10 countries cumulative totals of publications and researchers:

<https://docs.google.com/spreadsheets/d/14WRjPuJVNM_IbaNgmrT21oWjQlsa4ll0-NVjjSFcGlI/edit?usp=sharing>

Cumulative totals of persorcid and pubsorcid with motion charts:

<https://docs.google.com/spreadsheets/d/1nc3bOV9WGkBCzW4f8Tz9ws4ahlUF6_yXbstDTqZIHAQ/edit?usp=sharing>