Project Acronym: **THOR**
Grant Agreement no: **H2020-EINFRA-2014-2 654039**
Project Title: **Technical and Human Infrastructure for Open Research**

Study of ORCID Adoption across Disciplines and Locations

Document Information

Deliverable Type: OTHER
Dissemination Level: PUBLIC
Date: 19/12/2017
Authors: Robin Dasler (CERN, orcid.org/0000-0002-4695-7874)
Adenike Deane-Pratt (ORCID, orcid.org/0000-0001-9940-9233)
Artemis Lavasa (CERN, orcid.org/0000-0001-5633-2459)
Laura Rueda (DataCite, orcid.org/0000-0001-5952-7630)
Sünje Dallmeier-Tiessen (CERN, orcid.org/0000-0002- 6137-2348)
Abstract: As a basis for improving the global PID infrastructure, the THOR project studied ORCID adoption to identify trends and barriers. The study focuses on regional and disciplinary trends in the use of ORCID identifiers.
DOI n/a

Coordinated by the British Library, the THOR project is funded by the European Union.
Visit http://project-thor.eu for more information.

Revision History

Version	Status	Name, organisation	Date	Comments
1.0	FInal	Robin Dasler (CERN)		Submitted version

Review and Approval

Action	Name, organisation	Date
Reviewed by	Angela Dappert	13/12/2017
Approved by	Adam Farquhar (BL), Project Coordinator	

Report template version: 1.9, 22-03-2017

Project Summary

The **THOR** project establishes a sustainable international e-infrastructure for persistent identifiers that enables long-term access to critical information about the life cycle of research projects. It enables seamless integration between articles, data, and researcher information creating a wealth of open resources. This will result in reduced duplication, economies of scale, richer research services, and opportunities for innovation.

The project has four concrete aims:

1. Establishing interoperability
2. Integrating services
3. Building capacity
4. Achieving sustainability

The project will meet these aims by defining relations between contributors, research artefacts (including data), and organisations. We will incorporate these relationships into the ORCID and DataCite systems. We will also expand existing linkages between different types of identifiers and versions of artefacts to improve interoperability across platforms and integrate ORCID iDs into production systems for article and data submission services in pilot communities and beyond.

The consortium will develop systems to embed new PID resolution techniques into existing services to support seamless direct access to artefacts, and in particular data. We will create services to allow associations between datasets, articles, contributors and organisations at the time of submission. Building on these, we will deliver the means to integrate trans-disciplinary PID services in community-specific platforms, focussing on cross-linking, claiming mechanisms and data citation (guided by the FORCE 11 data citation principles[1]).

For more information, visit http://project-thor.eu or contact info@project-thor.eu

Copyright Notice

[1] https://www.force11.org/group/joint-declaration-data-citation-principles-final

Executive Summary:
Study of ORCID Adoption across Disciplines and Locations

In order to pursue its mission of improving the global PID infrastructure, the THOR project had previously developed the means to aggregate information on basic high-level trends in PID adoption from our primary infrastructural partners, ORCID and DataCite (Dasler, 2016). While this was useful for determining overall trends, we were limited in our ability to gain deeper understanding of factors affecting the adoption of PIDs globally and the service implications of those factors. We therefore undertook a deeper study of adoption of the PID services provided by ORCID. This amounted to an analysis of the trends and gaps evident in the statistical measures kept by ORCID, with a special focus on adoption across regions and disciplines, as those two factors were particularly relevant to future outreach and service development. We also provide a broader context for the discovered trends by identifying reference datasets to use for comparison.

In addition to pointing out areas of PID adoption success, the analysis also focussed our attention on areas for growth, in particular areas that could benefit from a further tailoring of outreach approaches and messages.

Contents

1 Background

The THOR project seeks to improve the interoperability of persistent identifiers (PID) and foster their integration and uptake. Targeted activities to support these efforts rely on knowledge of the current PID landscape, i.e. its evolution, drivers and barriers to integration and adoption. A first step towards obtaining and displaying such information was the development of the THOR dashboard, which helped determine overall trends in uptake of PIDs for contributors, data and papers. The nature and purpose of the dashboard was to be a simple aggregator of existing harvestable data. Any deeper understanding of specific questions still necessitates a stand-alone study and extensive data analysis. Therefore, in order to better understand the developments and potential drivers for the overall trends observed in the dashboard, this initial effort was followed up by the more detailed analysis presented here.

As a first step in this process, the dataset needed for the analysis was identified and studied. This dataset was derived from the ORCID (Deane-Pratt, 2017) registry. This is the first detailed analysis of this kind and therefore was also able to help us understand the limitations of the dataset, what kind of questions can be explored, and what is not yet possible. We ascertained that it is possible to obtain a significant amount of information about the use of ORCID iDs by discipline and location.

Within this report, gaps and limitations of the dataset underline that future work is still needed in order to get a more complete picture of the PID landscape.

Two recent studies underline the interest in understanding and exploiting such datasets better. A recent analysis by Bohannan (2017) investigated the mobility of researchers based on ORCID data. Klein and van de Sompel (2017) studied ORCID data on researchers, subjects, and locations to discuss potential discoverability of orphan works through such services.

2 Analysis

In this section of the report, we describe analyses done using available data from the ORCID registry. We were particularly interested in understanding:

- ORCID use according to discipline and region
- How this usage compares to the distribution of scholarly activity in general
- Whether ORCID use has changed over the lifetime of the registry
- Whether there are any identifiable factors that correlate with trends in ORCID usage

A detailed description of the methods used can be found in the following sections. Briefly, these were:

- Metadata extraction from the publicly available records in the ORCID registry
- A multi-step process of mapping people's records to scientific discipline based on their works, allowing for the assignment of multiple disciplines to both people (records) and works
- Mapping records and their works to a country and world region
- Obtaining reference data from an external source with which to compare the ORCID data

- Examining data for the four-year period 2012-2016, from ORCID's launch to the most recent public data file: totals, cumulative growth, and also on a quarter-by-quarter basis

Results from this part of the study, described in detail below, show that:

- ORCID users in the natural, health and applied sciences are best represented in the available data
- ORCID users in the humanities add fewer publications to their records than researchers in other disciplines
- ORCID users in Europe are better represented in the available data than other regions
- The rate of increase of ORCID usage is punctuated by peaks in activity that appear to be correlated with specific events such as national mandates and new integrations

2.1 Pre-processing of the ORCID Public Data File

2.1.1 Initial Sample

Data were extracted from the ORCID 2016 public data file[2]. There were 2,528,935 records in total, of which 1,070,218 contained at least one piece of public information in addition to a name and the ORCID iD itself. These 1,070,218 comprise the initial dataset used for this analysis, which will be referred to in this document as the *initial sample* (Figure 1). Table 1 contains a summary of the information attached to the records in this *initial sample*.

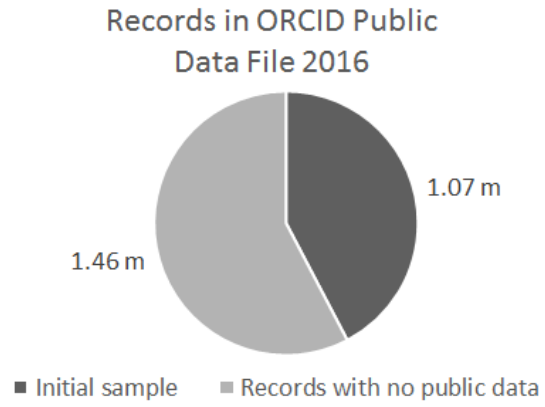

Figure 1: Of a total 2.53 million records in the ORCID Public Data File 2016, 1.07 million included public information in addition to name and ORCID iD. These comprised the initial sample used in the subsequent analyses.

[2] http://dx.doi.org/10.6084/m9.figshare.4134027.v1

Table 1: Summary of the contents within records in the initial sample.

Record type	**No. of records**	**% of *initial sample***
With country code	357,633	33
With employment	486,752	45
With education	557,428	52
With affiliation	660,516	62
With works	487,471	46
With funding	52,347	5
With works and country code	166,188	16
Works	**No.of works**	
Total works	13,942,859	

For the purposes of our analysis this *initial sample* was further processed into three subsets (Table 2, samples 3-5) to assess disciplinary uptake, geographical uptake and the interaction of the two. This required additional work to map the ORCID records in the *initial sample* to disciplines and locations.

Table 2: ORCID records selected for the discipline and location analyses.

Sample name	**No. of records in the sample**
1. ORCID public data file 2016	2,528,935
2. Initial sample (of records containing at least one public piece of metadata)	1,070,218
3. Discipline only mapping sample (of records containing at least one work)	487,471
4. Location only mapping sample (of records containing a country code or affiliation country)	785,020
5. Discipline + location mapping sample (of records containing both a country code or affiliation country and one or more works)	166,188

2.1.2 Discipline Mapping

Before we could analyse ORCID uptake across disciplines, we needed to consistently categorise records according to discipline. ORCID provides tools that support connections between researchers and their works and affiliations. In addition, researchers may add biographical details and keywords to their record, both of which can give an indication of the field of research in which they work. However, at present there is no globally-used discipline ontology, and the constantly evolving nature of research is such that disciplinary affiliation can have a large degree of fluidity. Together, this means that user-supplied information about disciplinary affiliation is not helpful for the purposes of this study. Instead, we chose to use publications as a proxy for discipline. Article/journal title and/or book title was available in the ORCID record metadata and could be reliably mapped to a discipline.

For this task, we utilised the Clarivate Analytics Web of Science (WoS) classifications. WoS is a database that indexes almost 30,000 academic journals, conference proceedings, books and patents. As of April 2017, there were more than 65 million records in the WoS database. WoS classifies records according to one or more subject areas from a predefined list of 251 fields. There is no limit on the number of subject areas that can be assigned to a WoS record, but most have only been assigned one or two subject areas.

Out of the 1,070,218 records in the *initial sample*, we selected those ORCID records that contained one or more works (487,471 records). These became the *discipline only* sample used in the discipline analysis. Table 2 shows a breakdown of each selected sample.

With the help of Clarivate Analytics, we matched the works in the *discipline only* sample to records within the WoS database. This matching was done using external identifiers in the ORCID works metadata as search terms. These variously included DOI, PubMed, ISSN or WoS internal identifiers. A full list of external identifier types is in Appendix B: External identifiers supported by ORCID. Works that could be matched to a record within the WoS database were assigned to the subject areas associated with that WoS record, preferentially according to the article title, otherwise to the journal title. Of 13.9 million unique ORCID works in the *discipline only* sample, 9 million were successfully mapped to a discipline in this way.

Works in ORCID records associated with a valid ISBN external identifier were first assigned a Library of Congress top-level shelfmark (Appendix C: Discipline map used in the ORCID analyses), retrieved using the Classify API (OCLC Research)[3]. By 'valid' we mean entries of the correct 10- or 13-character length for an ISBN number. Shelfmarks were then manually mapped to WoS subject area. This step increased the size of the *discipline only* sample to 9.2 million works. A schematic of the matching process is shown in Figure 2.

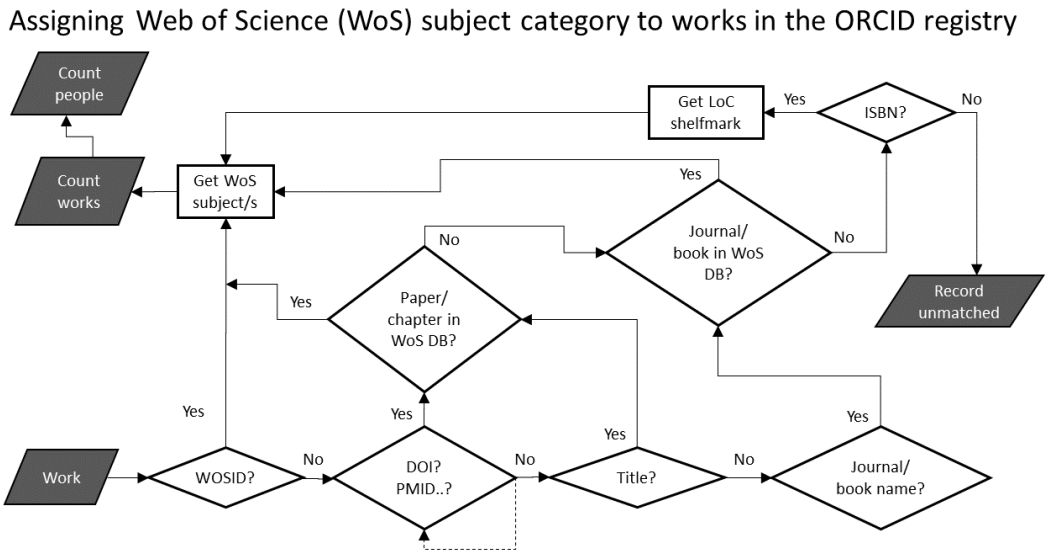

Figure 2: Schematic of the process used to assign scientific subject area/s from the 251 Web of Science categories to works in the ORCID registry samples.

[3] http://classify.oclc.org/classify2/

After mapping to the 251 WoS subject areas, ORCID works in the *discipline only* sample were also manually assigned to categories based on the top two levels of a three-level hierarchical subject classification tree developed by Science Metrix[4]. This was done to enable a broad overview of the results by major academic field, since WoS subject areas on their own were too granular to make such comparisons. At its most granular level, the Science Metrix system contains 176 categories. They are easily relatable - and in many cases identical - to categories in the WoS list of 251 subject areas.

The four top-level categories of the Science Metrix system and their 20 subcategories are as follows:

1. Applied sciences
   - 1.1. Agriculture, fisheries & forestry
   - 1.2. Built environment & design
   - 1.3. Engineering
   - 1.4. Information & communication technologies
   - 1.5. Technology & other applied sciences
2. Arts & Humanities, economic & social sciences
   - 2.1. Arts & general humanities
   - 2.2. Communication (language & literature)
   - 2.3. Economics & business
   - 2.4. Historical studies
   - 2.5. Philosophy & theology
   - 2.6. Social sciences
3. Health sciences
   - 3.1. Biomedical research
   - 3.2. Clinical medicine
   - 3.3. Psychology & cognitive sciences
   - 3.4. Public health & health services
4. Natural sciences
   - 4.1. Biology
   - 4.2. Chemistry
   - 4.3. Earth & environmental sciences
   - 4.4. Mathematics & statistics
   - 4.5. Physics & astronomy

The full mapping between WoS categories and the top two levels of the Science Metrix hierarchy is shown in Appendix C: Discipline map used in the ORCID analyses. In the Science Metrix taxonomy, arts & humanities is distinct to the top-level category for economics & social sciences. However, the two were combined here to produce approximately equal numbers of subcategories under the four major fields.

2.1.3 Location Mapping

To enable analysis of ORCID uptake by location, the 785,020 records in the *initial sample* with country information given in the biographical details or affiliation at September 2016 were selected as the *location only* sample. In addition, 166,188 records in the *initial sample* contained both works and country information, and formed the *discipline + location* sample. Works in the *location only* and

[4] http://www.science-metrix.com/en/classification

discipline + location samples were also assigned to one of the following five world regions, chosen to align with ORCID's existing reporting: Asia Pacific, Europe, Middle East & Africa, Latin America, USA/Canada.

2.2 Reference Data (CWTS)

To enable a comparison of the public ORCID Registry to the distribution of researchers and publications by region and discipline in the wider world, a proxy for such a worldwide distribution was needed. A reference dataset was therefore obtained from the Centre for Science & Technology Studies (CWTS) at Leiden University. CWTS holds an internal copy of the WoS database for their own research purposes from which they provided us with the total numbers of disambiguated authors (Caron & van Eck, 2014) and their publications as indexed in WoS for the period 2012-15 (Table 3).

This disambiguated author estimate was obtained using a rule-based computer algorithm that clusters publications by author according to a method developed at CWTS (Caron & van Eck, 2014). An estimate of the number of authors was necessary because of the inherent difficulty in uniquely identifying different authors who share a name.

Table 3: Summary of the reference dataset from CWTS

Measure	Unique count	Full count	Scale of the full count as a proportion of the unique count
Estimated no. of authors	11,704,957	38,134,913	3.26
No. of their publications	7,880,584	19,684,224	2.46

Note that the estimated number of authors in the '**full count**' column of Table 3 is based on counting each author once for every subject area assigned to their publications in the WoS database. For example, a person with publications in both physics and acoustics is counted as two distinct authors, once for physics and once for acoustics. Similarly, the full count of publications includes each publication once for every subject area assigned to its WoS record. A single publication indexed as belonging to both acoustics and physics research was counted twice, once for each subject. A similar approach was also described and used in Ruiz-Castillo and Costas (2014).

This full method of counting is 'multiplicative' in that it multiplies the original author estimate by the number of different subject areas in which the authors have published, and multiplies the number of unique publications by the number of subjects per publication as assigned in WoS. In this case, the estimated full count of authors in the reference dataset is 3.26 times the size of the original data. In other words, authors in this reference dataset publish in 3.26 different subject areas, on average.

One alternative method of counting would be to assign a single subject area to each author according to the field in which they are the most active (see for example Larivière & Costas, 2016), however this would potentially lose information about the full breadth of research activity. ORCID data were therefore also counted in the same multiplicative way in the following analyses, and are referred to as a 'full count' of the data.

2.3 Observations

In our analysis of the available data, we were primarily interested in the following general trends with regard to disciplinary and regional adoption of ORCID iDs:

Disciplines

- How are ORCID iDs adopted across disciplines?
- Is this comparable to the "usual" distribution of researchers/scholarly output across disciplines?
- How has ORCID use developed over time?

Regions

- How does ORCID adoption vary across regions?
- Have there been significant changes over time?
- How does this compare to known national initiatives (e.g., mandates)?

Our key findings related to these trends are presented below.

2.3.1 Discipline

Most ORCID records with works associated with them can be mapped to the sciences

A full count of the mapped data, grouped by the four top-level Science Metrix categories, shows that the most populous field of research in the *discipline only* dataset (Figure 3a) is the natural sciences with 1.39m ORCID iDs. This is 38% of the total share of 3.7m iDs across all four categories. The applied sciences represent 28% of the sample (1.04m), health sciences 25% (0.92m) and arts, humanities, economic & social sciences 9.6% (0.36m). The number of works added to ORCID records by this last research community is relatively fewer, however (Figure 3b) at 5.5%.

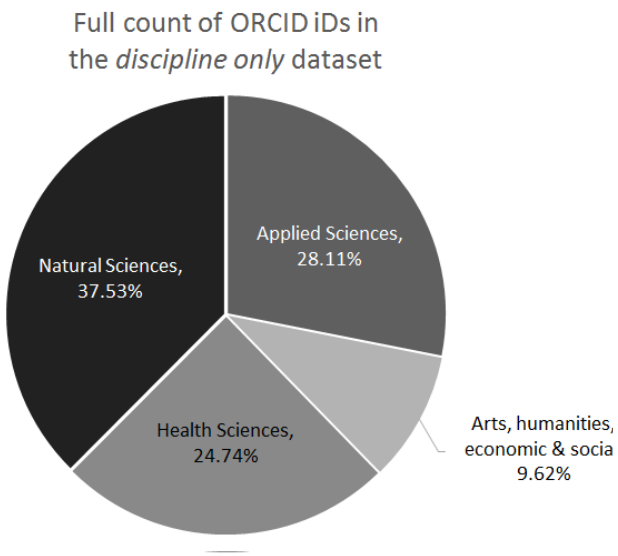

Figure 3a: Full count of the number of ORCID iDs in the *discipline only* dataset after mapping the four top-level areas of academic research.

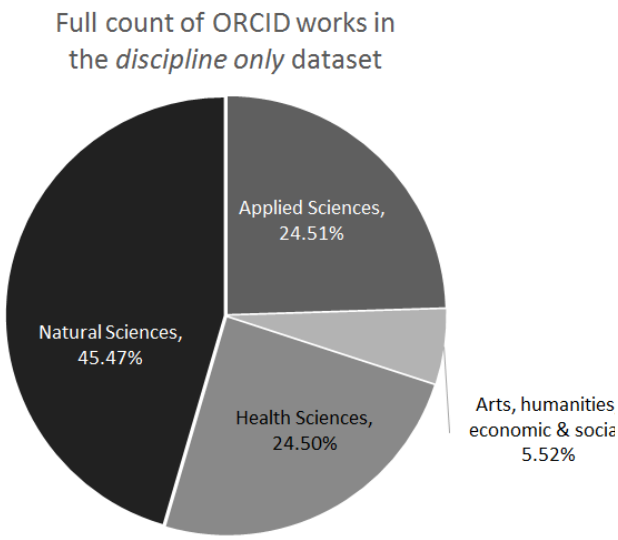

Figure 3b: Full count of the number of ORCID works in the *discipline only* dataset after mapping the four top-level areas of academic research.

Looking at these data in more detail, the count of ORCID iDs in the *discipline only* dataset (Figure 4) varies widely across the 20 categories, even within each top-level grouping. Three areas stand out as being the best-represented: clinical medicine, technology & other applied sciences, and biology, with 507k (13.69%), 469k (12.65%) and 454k (12.27%) of the share of ORCID iDs respectively. Within the second top-level category: arts, humanities, economic & social sciences, the social sciences subcategory contains the bulk of ORCID iDs, whereas philosophy & theology and the arts & general humanities are relatively poorly represented.

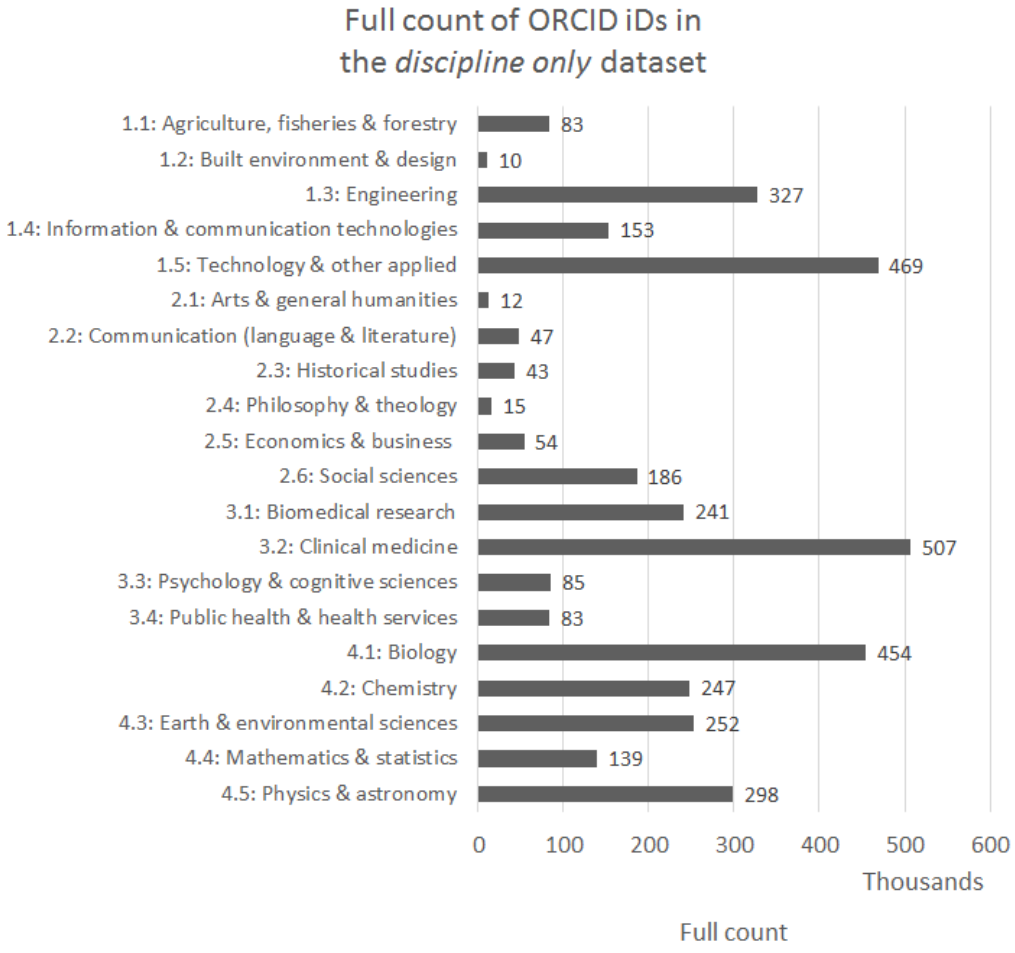

Figure 4: Full count of the number of ORCID iDs in the *discipline only* dataset after mapping the 20 second-level categories of academic research, grouped by the four top-level categories: 1) applied sciences; 2) arts, humanities, economics & social sciences; 3)

In the reference dataset (Figure 5) the distribution across the 20 categories is broadly similar to the ORCID data. Clinical medicine is the most common field in which authors have published, which is in line with the ORCID Registry data. The main differences seem to be in the applied sciences, in which ORCID has relatively more iDs in the engineering and technology & other applied sciences subcategories, as well as in the social sciences.

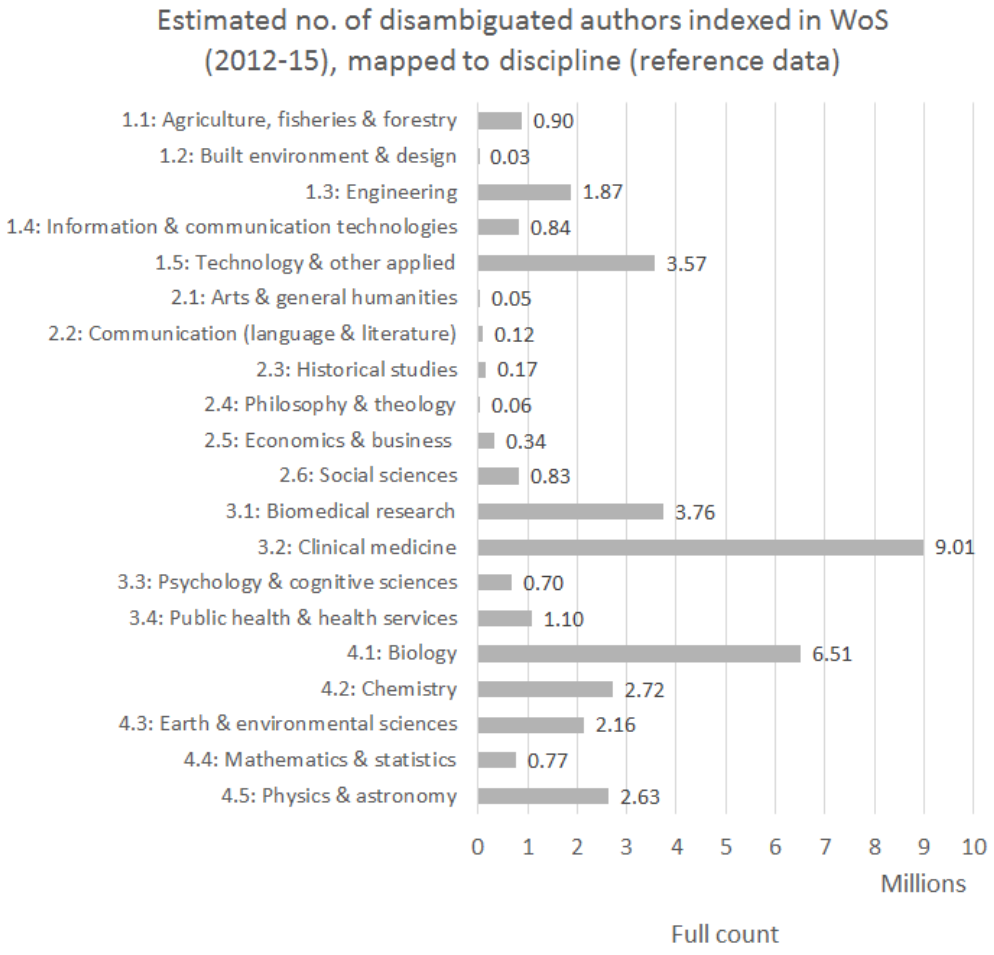

Figure 5: Full count of the estimated number of disambiguated authors in the reference data after mapping the 20 second-level categories of academic research, grouped by the four top-level categories: 1) applied sciences; 2) arts, humanities, economics & social

ORCID data is better represented in the arts, humanities, economic & social sciences and in the applied sciences when compared to the CWTS reference data

In the CWTS reference dataset, the estimated number of authors is also greatest in the natural sciences at 14.79m (Figure 6). This accounts for 39% of the 38.1m authors across the four categories and is approximately equal to the 38% share in the ORCID *discipline only* data. However, in the reference data, the health sciences take second place with 14.56m authors (38%) and the applied sciences are third, with 7.21m (19%), which is a reversal of rankings from the ORCID *discipline only* data set.

Both the ORCID *discipline only* and CWTS reference datasets have the fewest people publishing in the arts, humanities, economic and social sciences. However, at this top-level of analysis the ORCID dataset is less skewed towards the health sciences, and the *discipline only* dataset has more than double the percentage of arts, humanities, economic and social sciences iDs (9.6%) compared to the same category of authors in the reference data (4.1%; 1.57m).

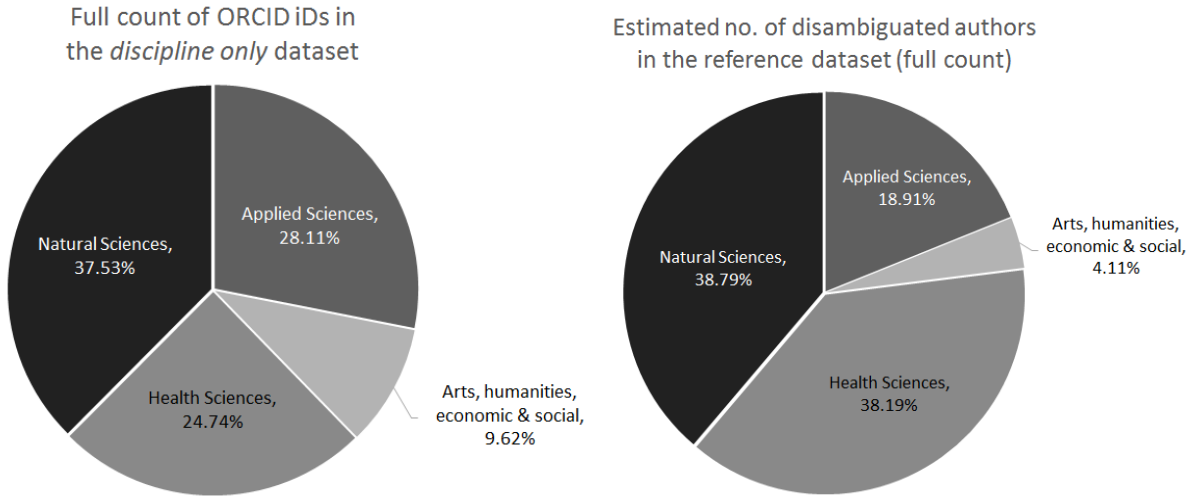

Figure 6 : Comparison of the full count of ORCID iDs (representing authors) in the *discipline only* dataset and the full count of the estimated number of disambiguated authors in the reference data after mapping to the four top-level areas of academic research.

ORCID record-holders publish in multiple disciplines

After discipline mapping, there were 0.58 million unique ORCID records and 9.2 million works in the *discipline only* sample to which at least one discipline could be assigned (Table 4). With full counting, the number of records in this mapped *discipline only* dataset increased six-fold (6.16 times the number of unique records) to 3.7 million. This ratio suggests that authors with ORCID iDs are generally publishing in multiple disciplines. The change in the number of works was smaller, increasing by a factor of 1.82 from 9.2 to 16.7 million. This smaller increase factor implies that individual publications fall under a narrower set of disciplines than would a researcher's total output, as one would expect. It is interesting to compare full count to unique count proportions of the ORCID datasets (Table 4) with the those of the reference dataset (Table 3) and note that the proportion reported for the ORCID datasets is higher. Further investigation would be required to determine whether this is indicative of a tendency among ORCID registrants or whether this is an artefact of the datasets used for the analysis.

Table 4: Summary of the ORCID datasets after discipline and location mapping.

Measure	Unique count	Full count	Full count as a proportion of unique count
No. of records in the *discipline only* data	583,707	3,703,958	6.16
No. of works in the *discipline only* data	9,192,363	16,684,224	1.82
No. of records in the *location only* data	785,020	785,020	1

2.3.2 Trends over Time across Disciplines

The number of ORCID iDs associated with works is increasing steadily in all field categories

A trend analysis of the cumulative total full count of ORCID iDs in the *discipline only* dataset (Figure 7) shows that in all four of the top-level categories, the number of iDs has increased fairly steadily since the

end of 2013, one year after ORCID was launched. Growth is slowest in the arts & humanities, but the rate of growth in this area shows a slight increase in the final quarter of 2015. (See Figure 9a and Figure 9b.)

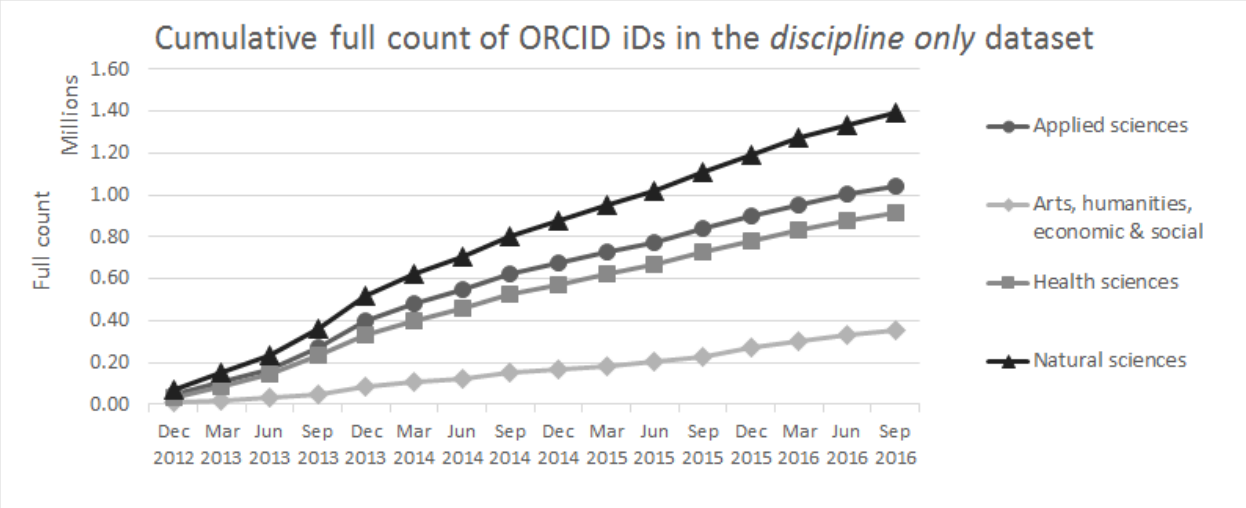

Figure 7: Trend analysis of ORCID iDs in the mapped *discipline only* dataset according to creation date in the ORCID registry, for each of the four top-level categories.

The number of works connected to ORCID records has grown more rapidly in the past 2 years

In order to understand more about the way ORCID is being used, we also looked at the rate at which works have been added to records in the registry. Figure 8 shows a trend analysis of the cumulative full count of works in the *discipline only* dataset, according to their creation date, for each of the four top-level disciplines.

As expected, given the relatively small number of ORCID iDs matched to the arts, humanities, economic & social sciences, the number of works in this area is fewest. However, unlike the analysis of iDs, the rate of growth in works added to records shows a marked increase from the second quarter of 2015 onwards. This change is most apparent in the natural sciences, but is also visible in the health sciences and applied sciences, with the latter two growing in step with each other.

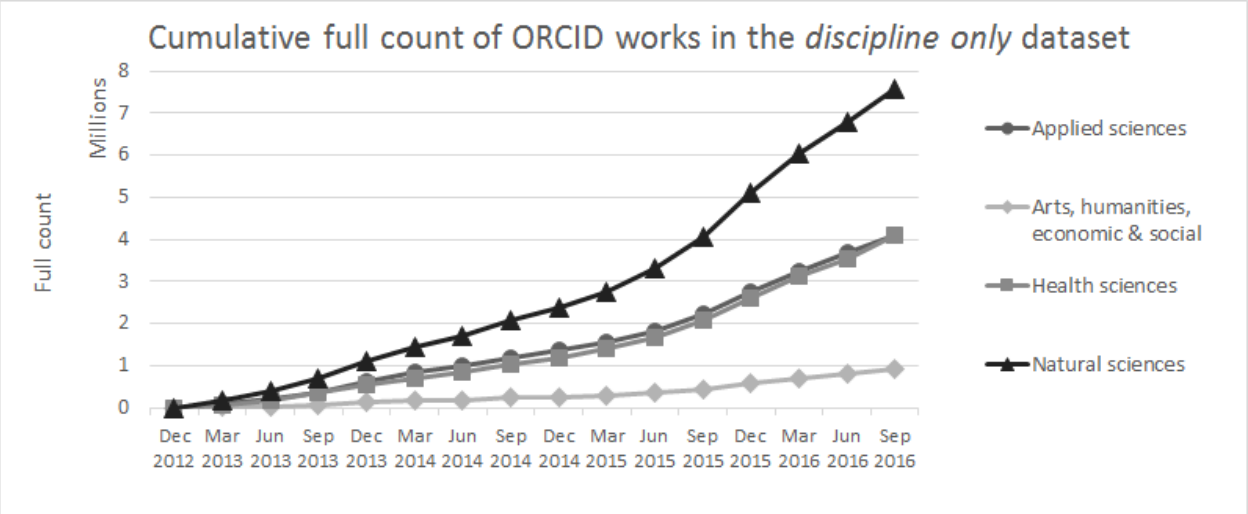

Figure 8: Trend analysis of ORCID works in the mapped *discipline only* dataset according to creation date in the ORCID iD-holder's record, for each of the four top-level categories.

ORCID uptake shows peaks of activity across all scientific disciplines

To examine the reasons for changes in growth rate of ORCID iDs and ORCID works, we looked at the full count of both iDs and works in the *discipline only* dataset, quarter by quarter. The figures below show the quarterly additions of ORCID iDs (Figure 9a) and ORCID works (Figure 9b) for each of the four top-level subject categories.

The data show three clear peaks in activity for both ORCID iDs and ORCID works in the quarters ending in December 2013, September 2014 and December 2015. These peaks are apparent in all four subject areas. In the case of ORCID iDs, the December 2013 peak is the most prominent and especially in the natural sciences and applied sciences. For ORCID works, the greatest peak is in December 2015 and most apparent in the natural sciences.

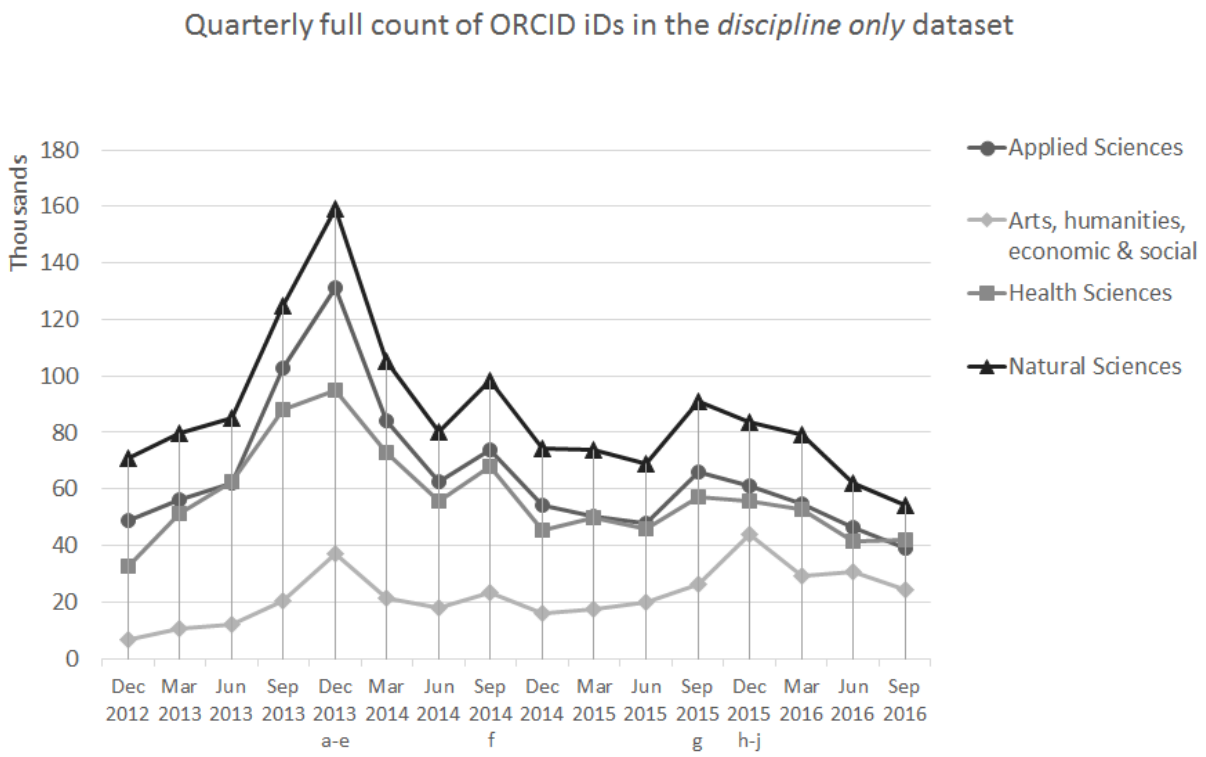

Figure 9a: Trend analysis of the full count of ORCID iDs from the *discipline only* dataset by quarter, according to the creation date in the registry.

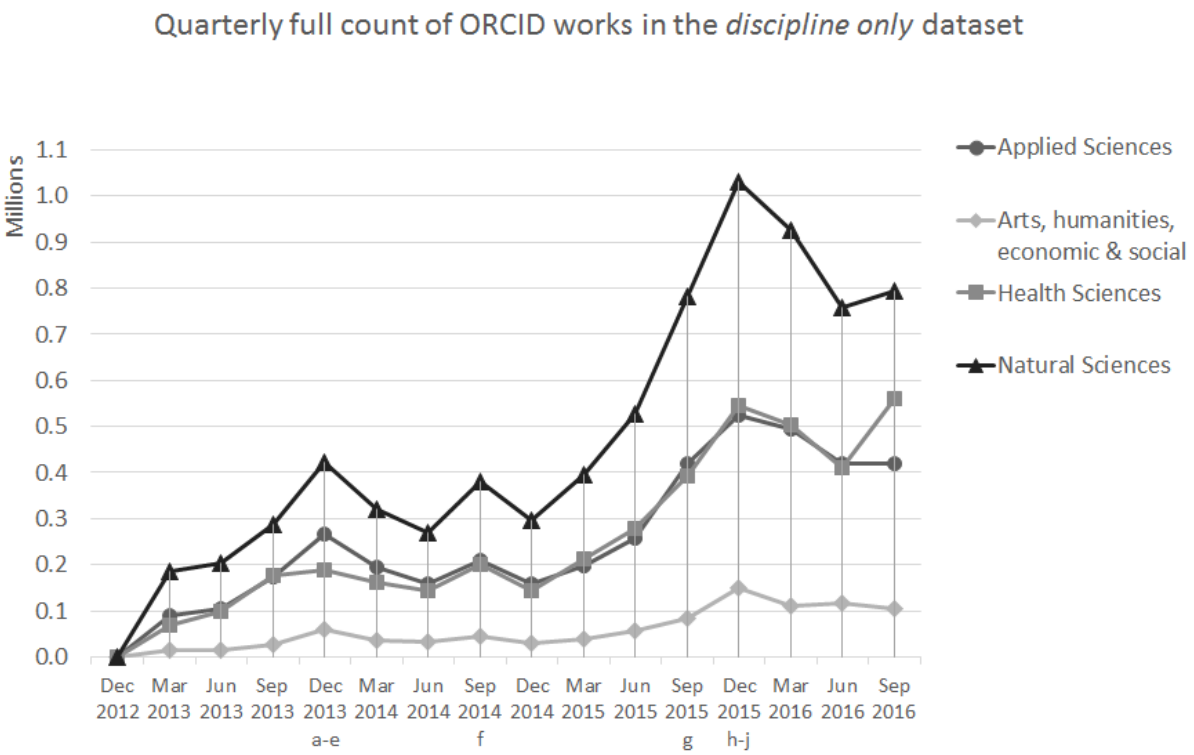

Figure 9b: Trend analysis of the full count of ORCID works from the *discipline only* dataset by quarter, according to the creation date in the registry.

Peaks in ORCID uptake correspond to specific updates, service launches, integrations and mandates

These peaks across all subject areas suggest that specific actions or events may have increased awareness of ORCID or facilitated its use. Looking back at ORCID's communications and outreach activities over the past four years we can see a number of likely causes for the marked increases in uptake. Table 5 lists a number of such events and activities, and these are marked along the horizontal axes in Figure 9a and Figure 9b with the corresponding letters from the 'Key' column. It is important to note that many of these events involve integrations in partnership with organisations external to ORCID.

Table 5: Events and activities that are likely drivers of increases in ORCID uptake across disciplines.

Key	Quarter ending	Event
a	Dec 2013	UI update: ORCID works manual entry
b	Dec 2013	New integrations: including CERN, Elsevier PURE, Atlas and Flysheet
c	Dec 2013	UI update: ORCID works display
d	Dec 2013	Service launch: Scholar One (Clarivate Analytics)
e	Dec 2013	Regional mandate implemented: Portugal (via Scopus)
f	Sep 2014	Service launch: BibTeX import
g	Sep 2015	Service launch: Jisc ORCID Consortium
h	Dec 2015	Milestone event: 1.5m iDs (Haak, 2015)
i	Dec 2015	Regional mandate implemented: Italy (via CINECA)
j	Dec 2015	Service launch: auto-update implemented by Crossef and DataCite

ORCID uptake varies widely between disciplines within each top-level academic field

While peaks in ORCID uptake are apparent at the broadest level of academic discipline we were also interested to find out what the picture looks like at a more granular level within the four categories. Figure 10a, Figure 10b, Figure 10c, and Figure 10d show the quarterly additions of ORCID iDs for each of the four top-level subject categories broken down into their sub-categories. Figure 11a, Figure 11b, Figure 11c, and Figure 11d show the same breakdown for ORCID works.

It is clear from this analysis that the events and activities listed in Table 5 have had a markedly greater effect in some fields compared to others. The major new integrations in the quarter ending December 2013, together with improvements to the ORCID user interface, the launch of Scholar One and the national mandate in Portugal notably correlate with uptake of ORCID iDs. This is most apparent in the fields of technology & other applied sciences; engineering; social sciences; clinical medicine; and in all of the natural sciences, particularly in biology.

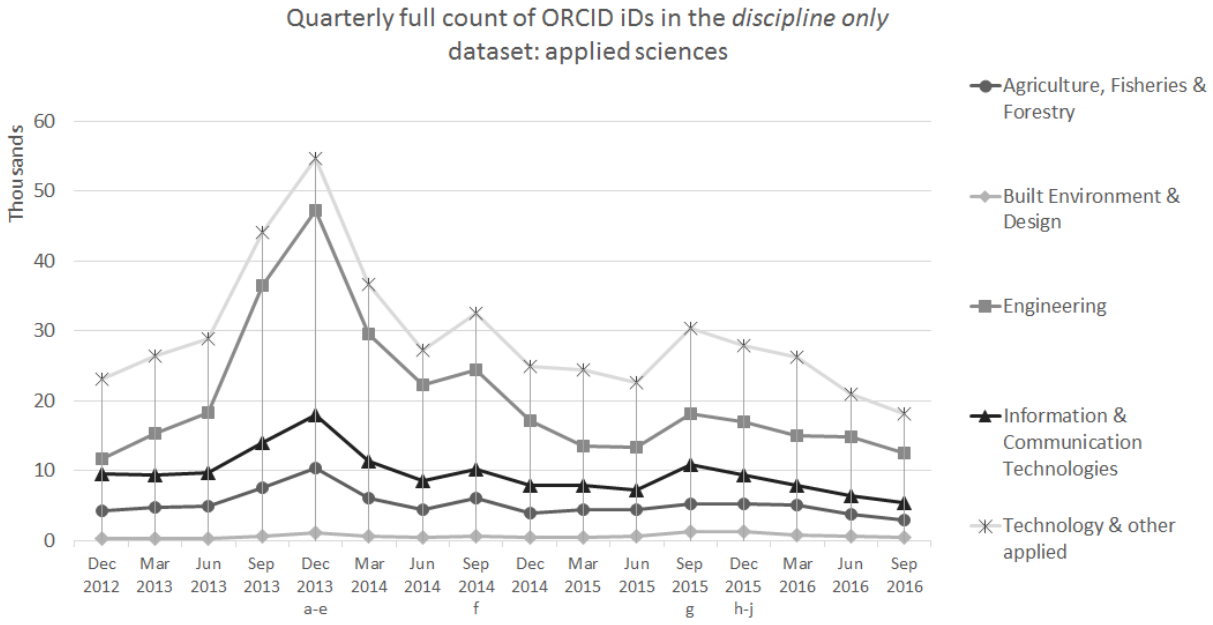

Figure 10a: Trend analysis by quarter, according to creation date in the registry, of the full count of ORCID iDs within the category of applied sciences from the *discipline only* dataset.

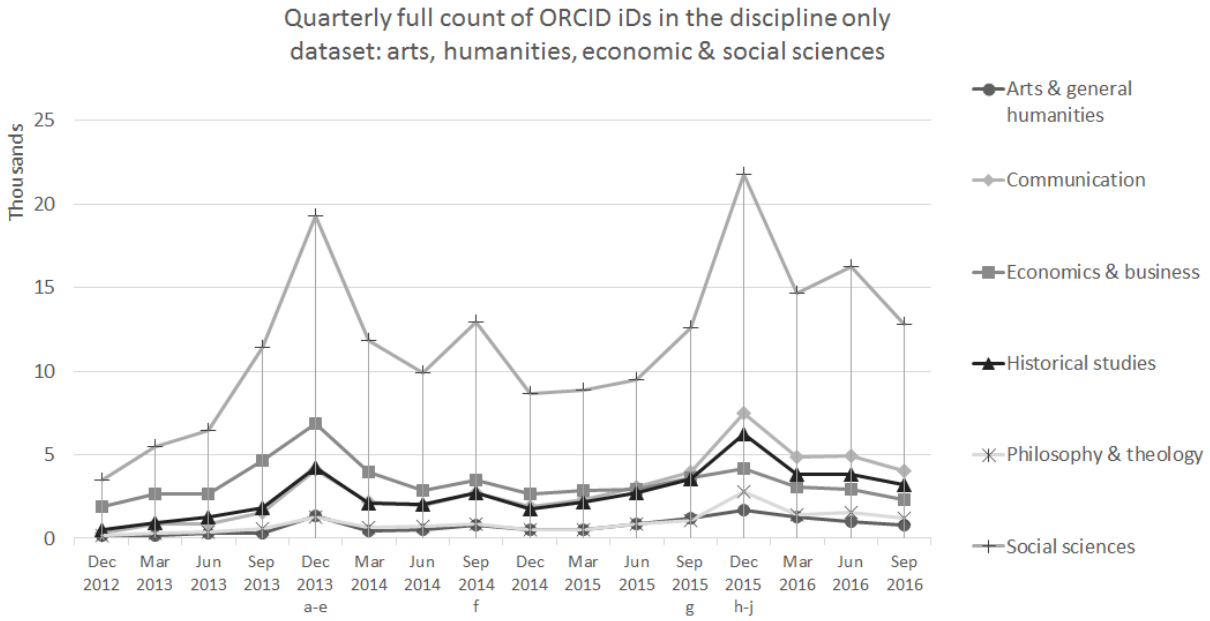

Figure 10b: Trend analysis of the full count of ORCID iDs within the category of arts, humanities, economic & social sciences from the *discipline only* dataset by quarter, according to the creation date in the registry.

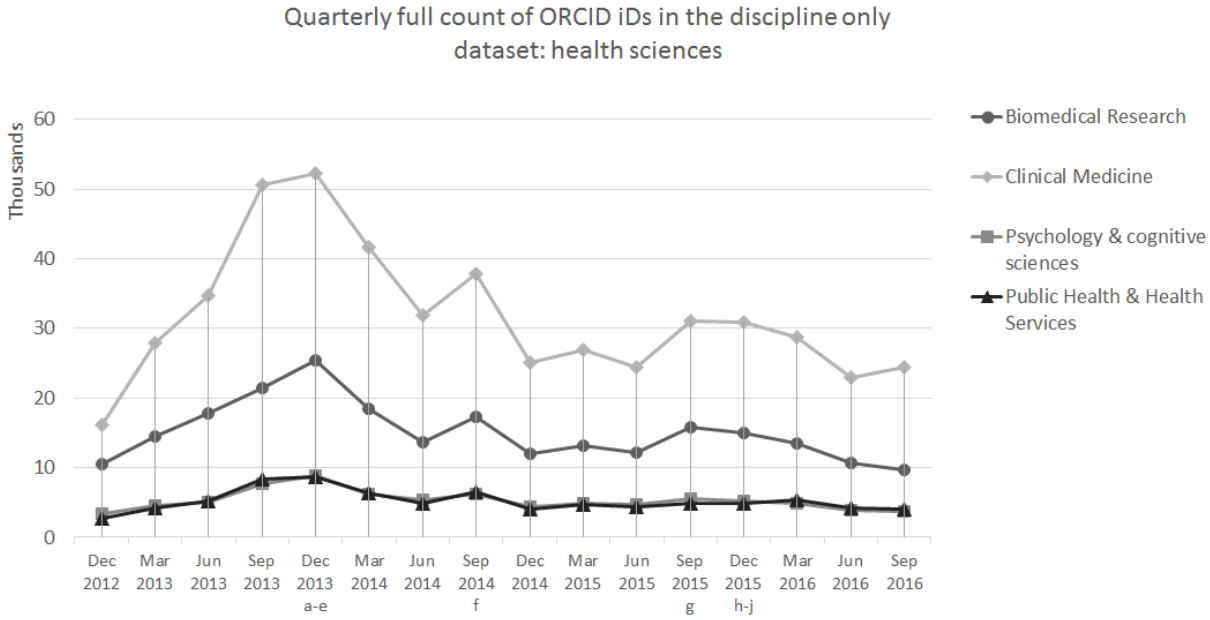

Figure 10c: Trend analysis of the full count of ORCID iDs within the category of health sciences from the *discipline only* dataset by quarter, according to the creation date in the registry.

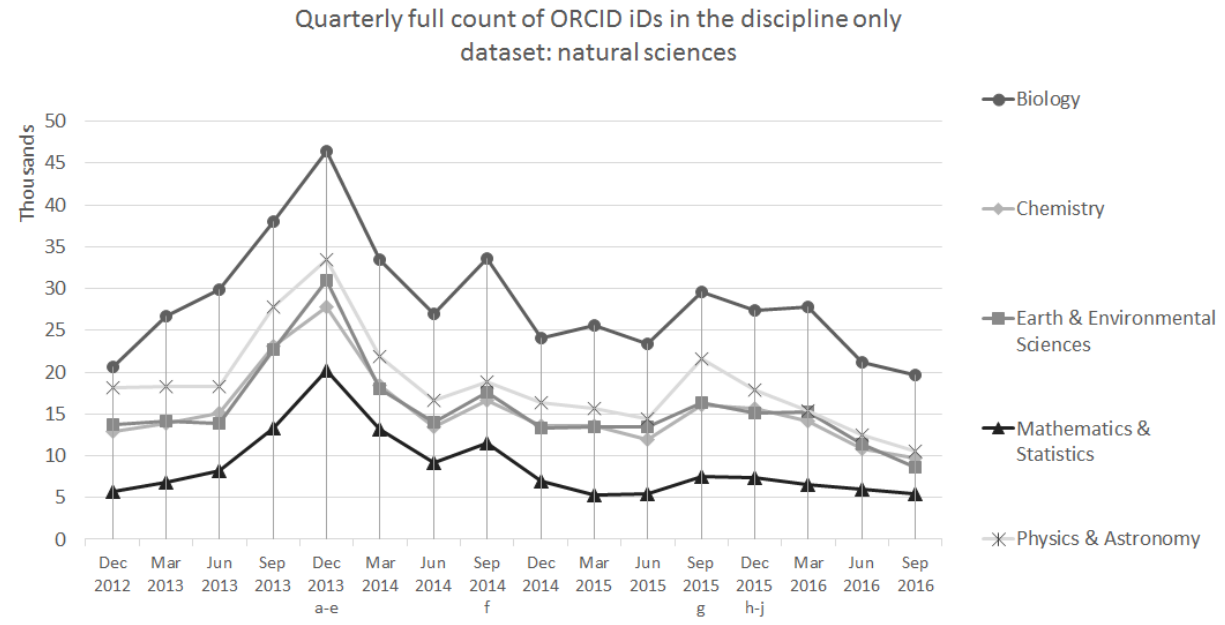

Figure 10d: Trend analysis of the full count of ORCID iDs within the category of natural sciences from the *discipline only* dataset by quarter, according to the creation date in the registry.

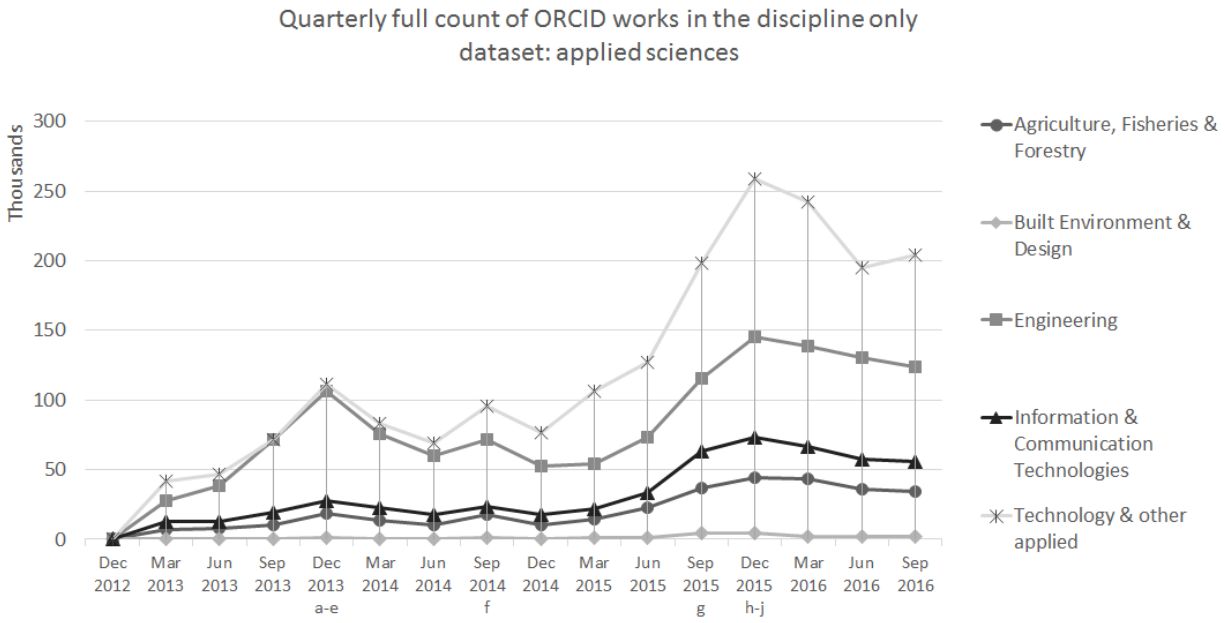

Figure 11a: Trend analysis of the full count of ORCID works within the category of applied sciences from the *discipline only* dataset by quarter, according to the date they were added to ORCID records.

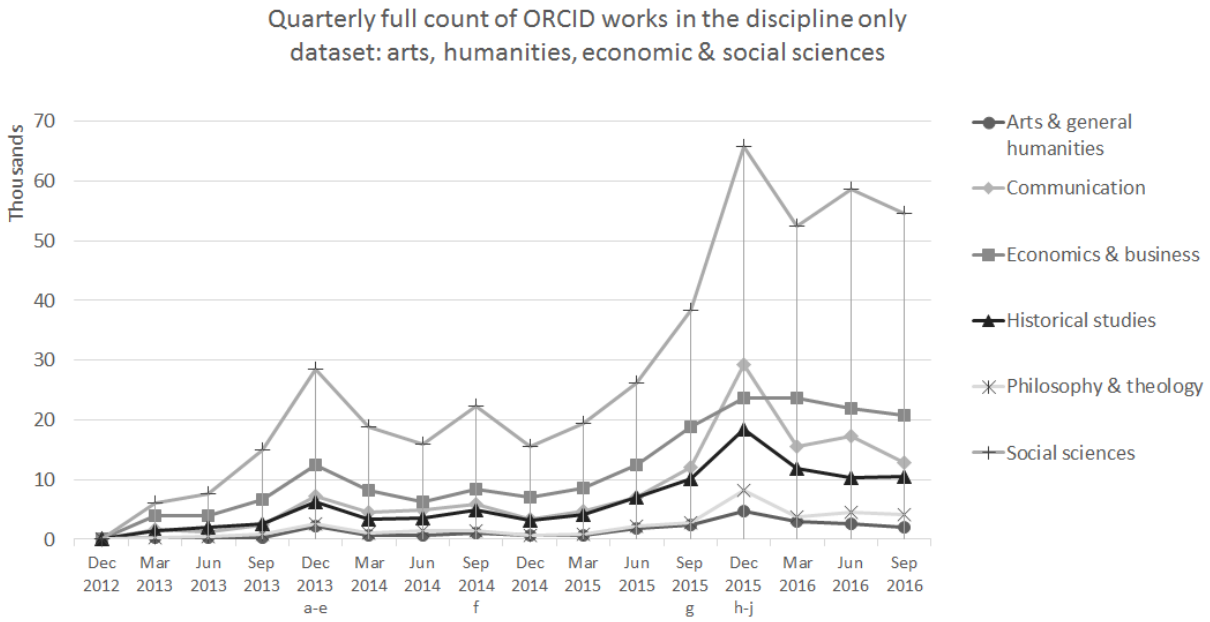

Figure 11b: Trend analysis of the full count of ORCID works within the category of arts, humanities, economic & social sciences from the *discipline only* dataset by quarter, according to the date they were added to ORCID records.

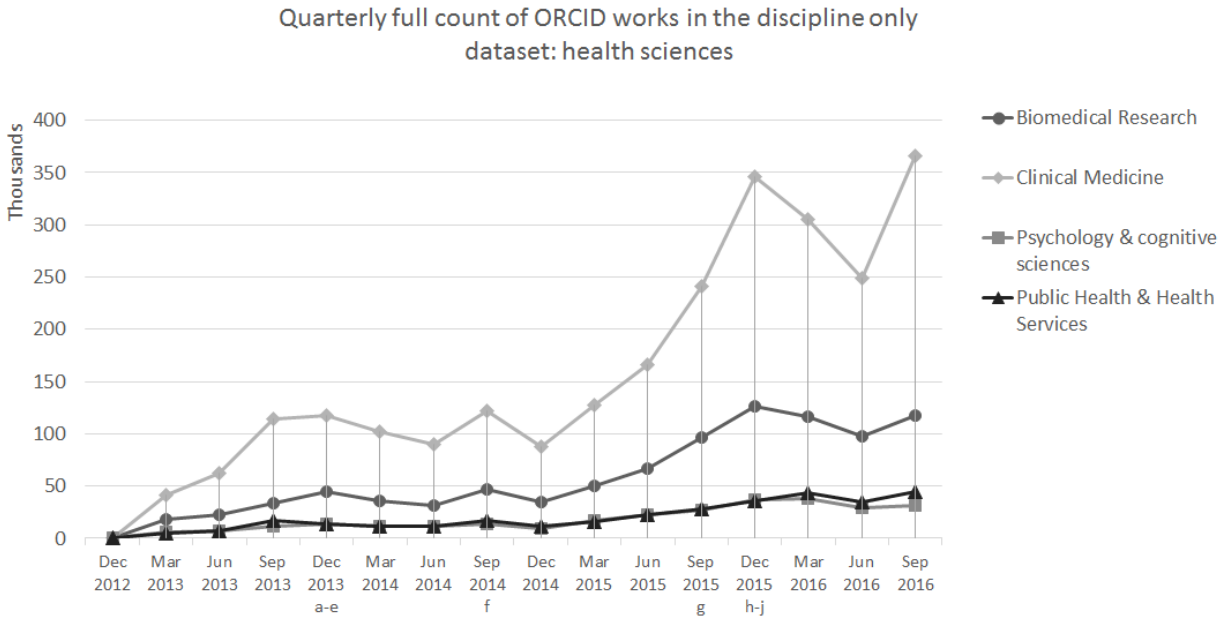

Figure 11c: Trend analysis of the full count of ORCID works within the category of health sciences from the *discipline only* dataset by quarter, according to the date they were added to ORCID records.

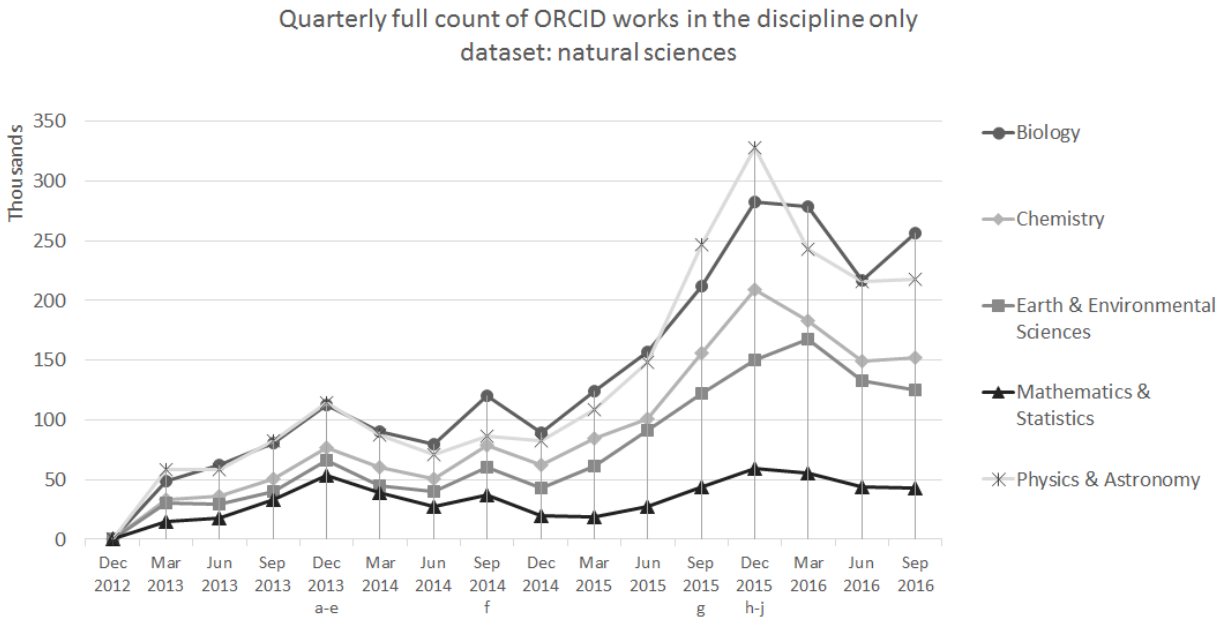

Figure 11d: Trend analysis of the full count of ORCID works within the category of natural sciences from the *discipline only* dataset by quarter, according to the date they were added to ORCID records.

2.3.3 Location

The ORCID dataset is more likely to be associated with European affiliations and less likely to be associated with Asia Pacific compared to the reference data.

Results from the location analysis show that Europe is by far the most active region for both ORCID iDs and works (Figure 12a and Figure 12b). The 0.33m iDs in the *location only* dataset make up 42% of the total, and the effect is even more pronounced for works, at 58%. These data are consistent with the fact that more than half (56%) of ORCID member integrations are in Europe (Meadows, 2016). Member integrations allow organisations such as publishers, research institutions and funders to connect with individual ORCID records via an application programming interface (API).

However, while the USA/Canada is ahead of Asia Pacific in terms of the number of integrations (27% and 14% respectively), the picture is almost reversed for ORCID iDs in this dataset, with 26% in the Asia Pacific region and 17% in the USA/Canada. The number of works attached to records in this dataset is similar for both regions (18% Asia Pacific; 17% USA/Canada). ORCID uptake is lowest in the Middle East & Africa and in Latin America, which currently have the fewest member integrations (2% and 1%, respectively). Nevertheless, Latin America accounts for 10% of ORCID iDs in the *location only* data.

For comparison, we also looked at the reference dataset by world region. Figure 13b shows a fairly similar picture to the ORCID data, with Europe dominating the number of publications. However, the estimated number of authors is greatest in the Asia Pacific region, rather than in Europe (see Figure 13a).

ORCID adoption is highest in countries with a national framework for ORCID (see 2.3.4). To see a more consistent adoption across Europe, a European-level framework would be needed in order to drive national integrations.

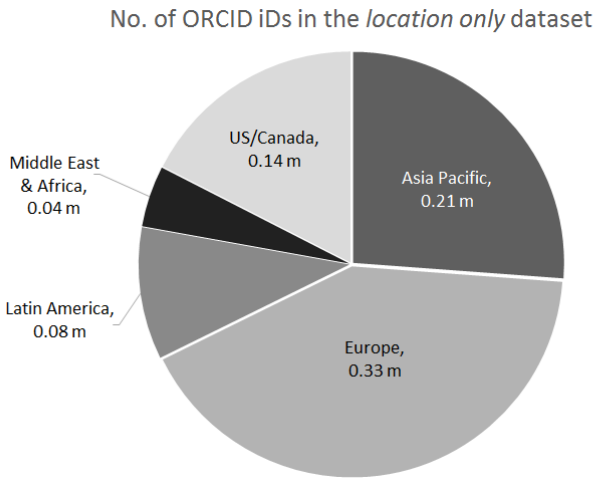

Figure 12a: The total number of ORCID iDs in the *location only* dataset by world region.

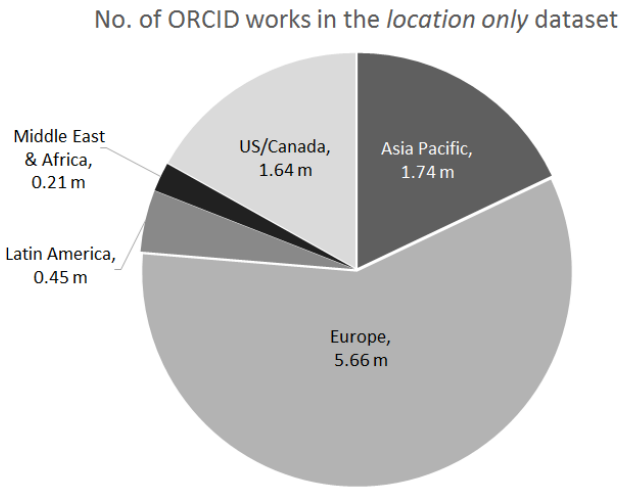

Figure 12b: The total number of ORCID iDs works in the *location only* dataset by world region.

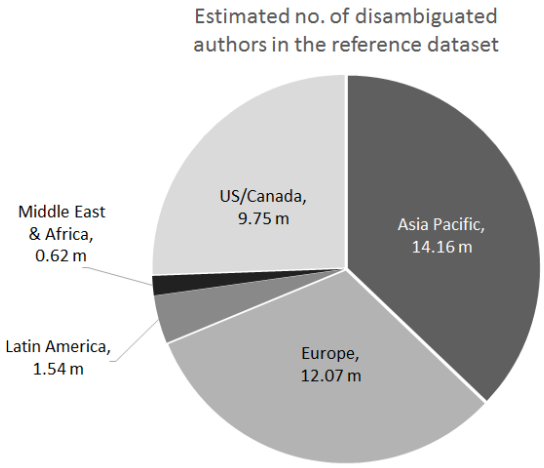

Figure 13a: The estimated number of disambiguated authors in the CWTS reference dataset by region

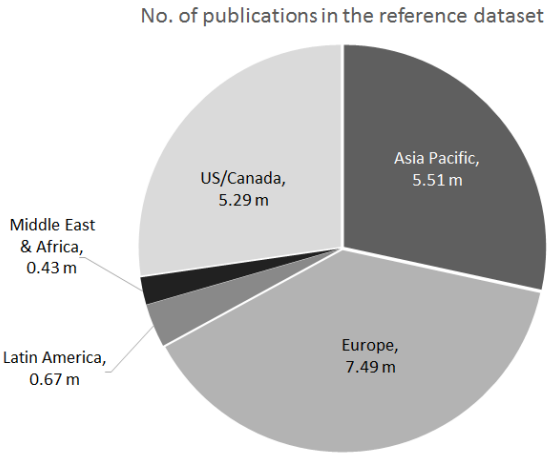

Figure 13b: The estimated number of publications in the CWTS reference dataset by region.

2.3.4 Trends in Geographical Distribution

ORCID region data shows most growth in Europe.

The level of ORCID uptake is increasing across all world regions, both in terms of iDs created and works added to records. However it is clear that the increase is happening at different rates between regions, and is fastest in Europe (Figure 14a and Figure 14b). Growth is slowest in Latin America and in the Middle East & Africa, especially for works (as opposed to iDs) added to the registry.

Peaks in ORCID data by world region are most apparent in Europe.

As was the case in the analysis of the *discipline only* dataset, we can identify a number of factors that correlate with changes in the rate of ORCID uptake by world region that are apparent in the *location only* dataset. Figures 15a and Figure 15b show the trend analysis broken down by world region and annotated with the events in Table 5, above. In Europe especially, multiple peaks in the overall upward trend suggest that specific initiatives have had an impact on uptake. This is most clearly apparent in the quarters ending September and December 2015 (when the JISC ORCID consortium and auto-update feature launched) when the number of ORCID works peaked here, but not in other regions.

National mandates correlate with increased ORCID uptake.

A more specific example of regional initiatives having a clear and beneficial effect on ORCID uptake can be seen in Figure 16a and Figure 16b. Peaks in both the number of new ORCID iDs and works are visible in the data from Portugal and Italy, where national mandates for ORCID use were implemented in the quarters ending December 2013 and September/December 2015, respectively.

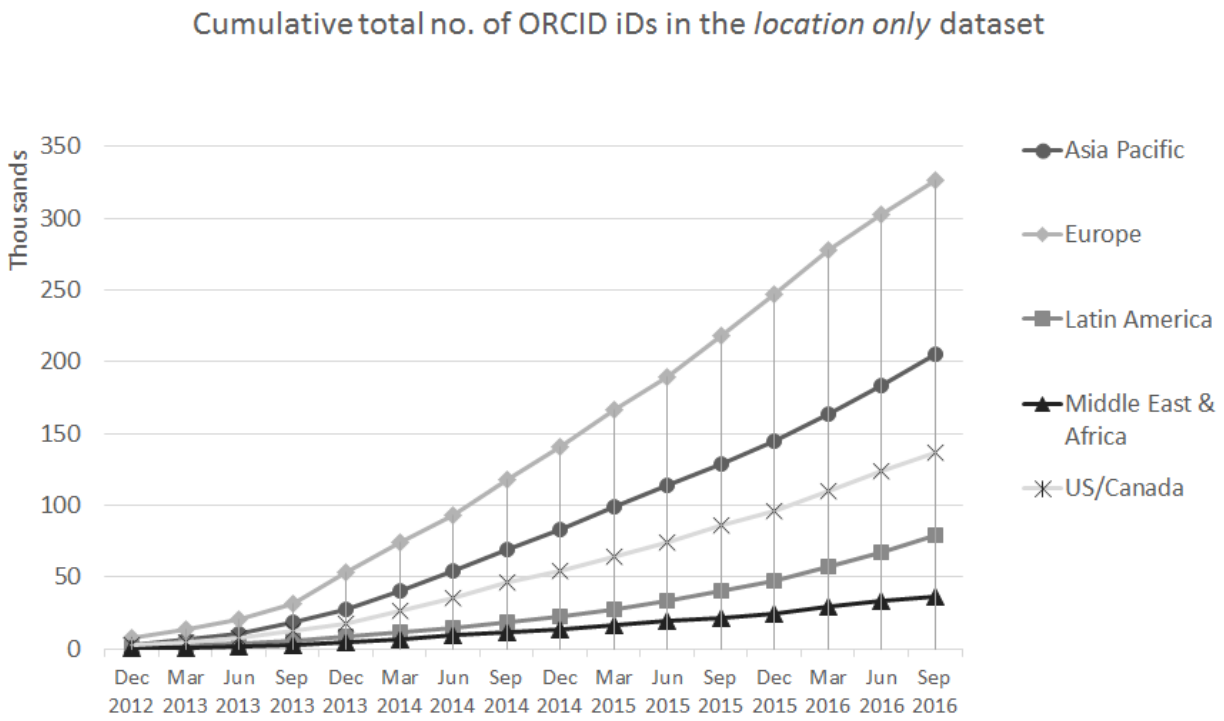

Figure 14a: Trend in the creation of ORCID iDs by region.

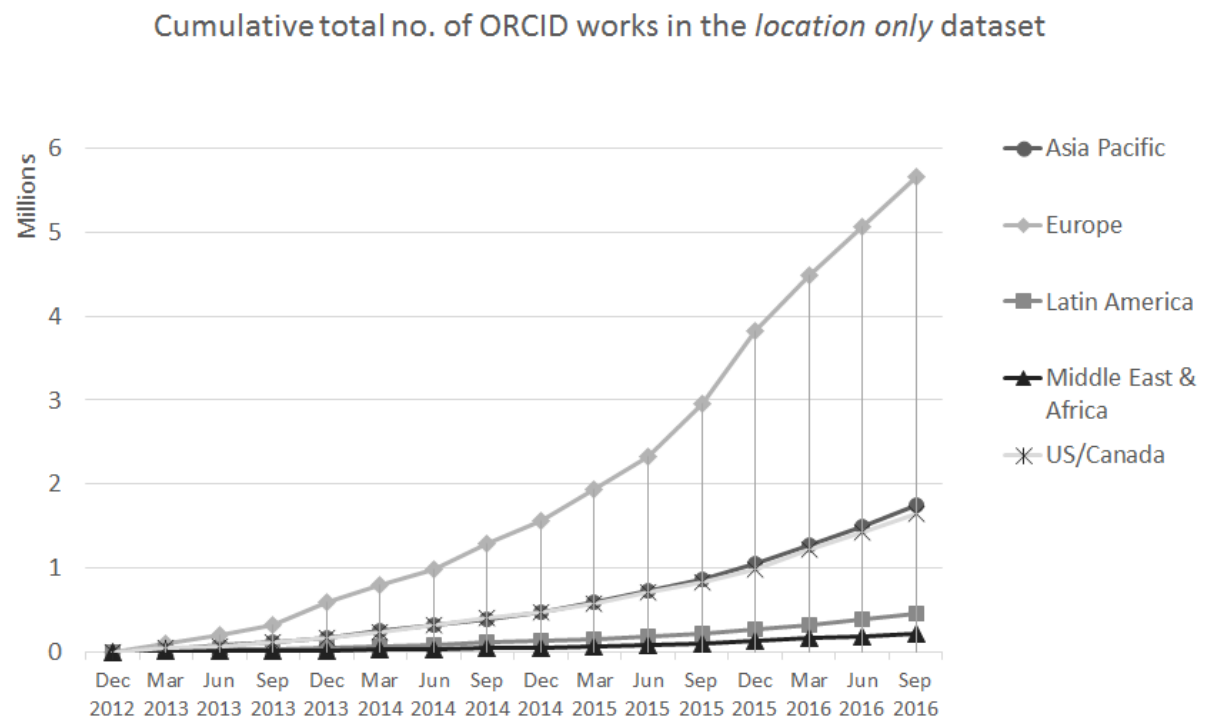

Figure 14b: Trend in the addition of works to ORCID records by region.

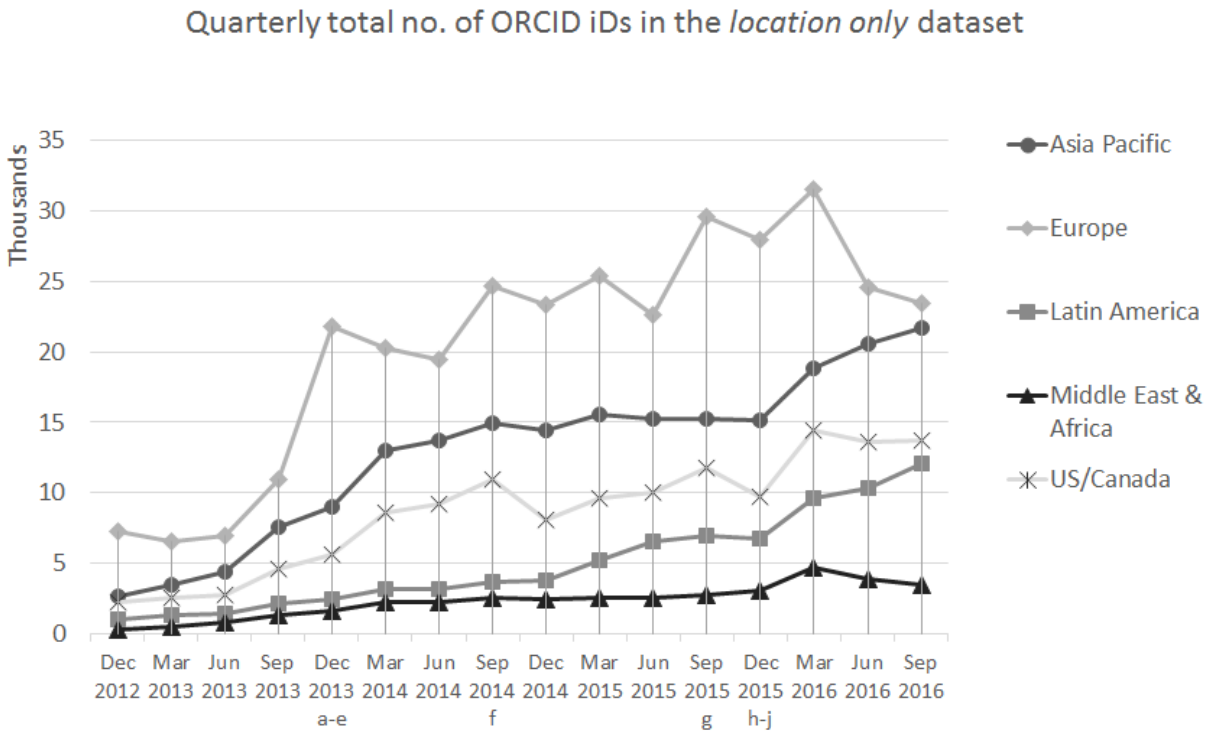

Figure 15a: Trend in the number of ORCID iDs added to the registry, quarter by quarter, by world region. See Table 5 for a key to the annotations a-j.

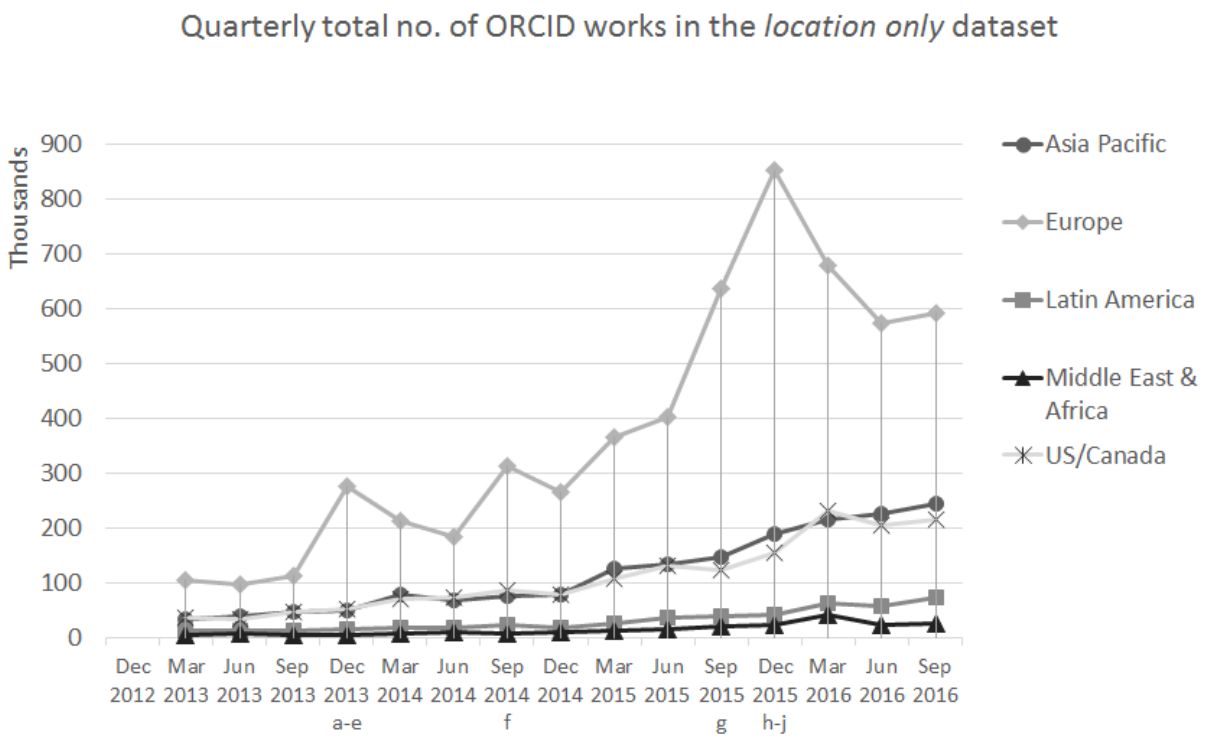

Figure 15b: Trend in the number of works added to the ORCID registry, quarter by quarter, by world region. See Table 5 for a key to the annotations a-j.

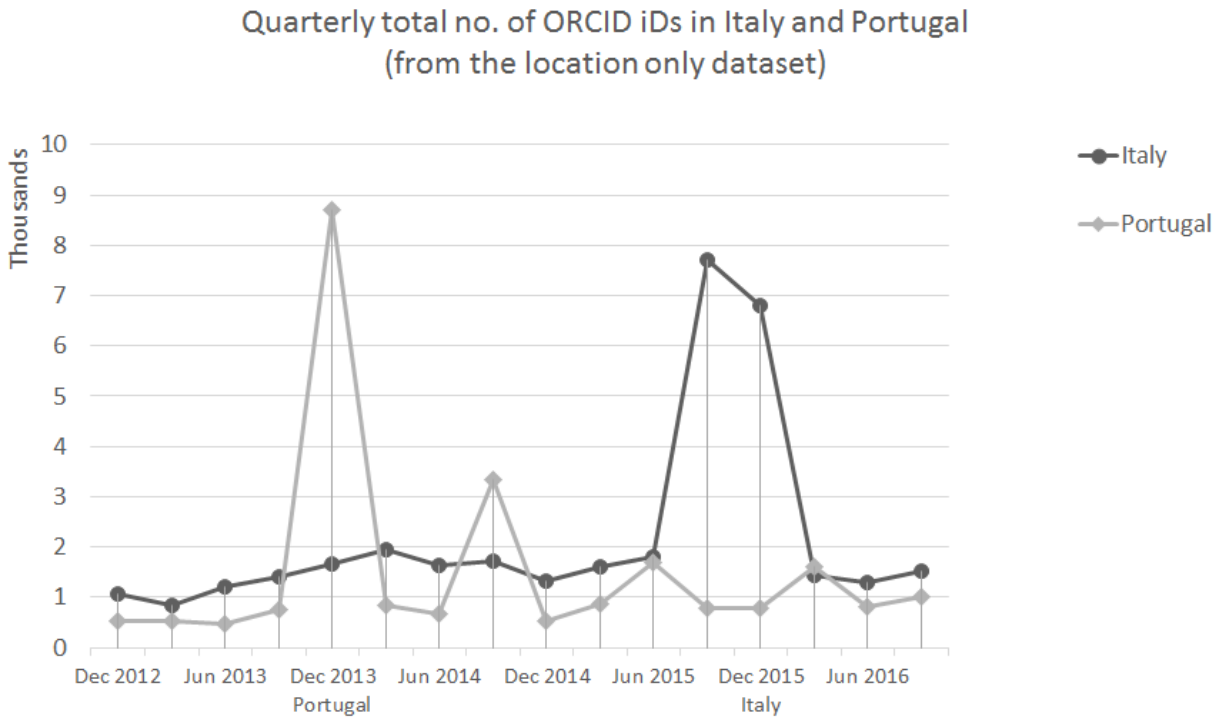

Figure 16a: Peaks in the number of ORCID iDs added to the registry by quarter in Italy and Portugal, following national mandates for ORCID use.

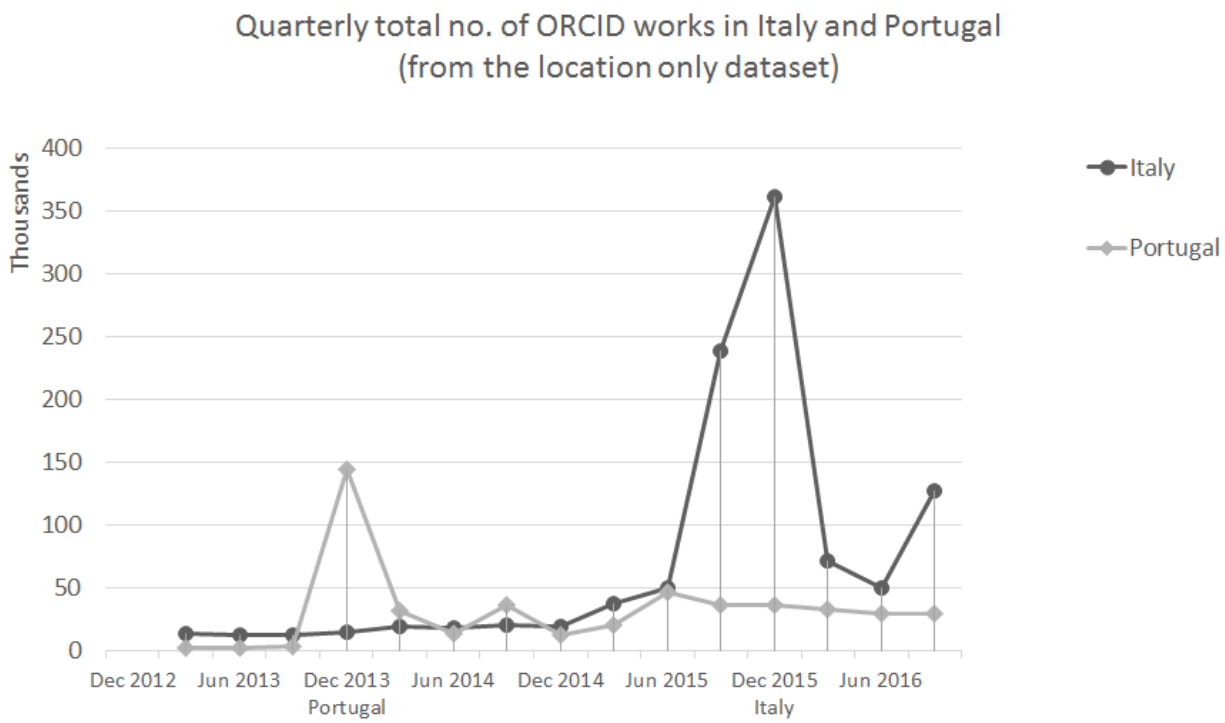

Figure 16b: Peaks in the number of works added to the ORCID registry by quarter in Italy and Portugal, following national mandates for ORCID use.

2.3.5 Discipline and Location Dataset

In addition to finding out about ORCID usage by discipline and region, we also wanted to examine the interaction of these two factors, using the *discipline + location* dataset. As we saw from looking at discipline on its own, there is considerable variation between the top-level categories. However, in the *location only* analysis, Europe clearly dominates, to the extent that the *discipline + location* data is very similar to location alone (not shown). Therefore, in this part of the analysis we chose to focus on *discipline + location* excluding Europe from the picture.

Figure 17 shows the full count of ORCID iDs created and works added to records, quarter-by-quarter for the four top-level subject areas and four of the five world regions in the *discipline + location* dataset. In the Asia Pacific region and in USA/Canada ORCID iDs are being created at a fairly steady rate over time, punctuated by the same sorts of peaks that we saw in the above analyses. The number of works being added to the registry is increasing in these regions overall, and also shows peaks of activity.

For most disciplinary areas, significant increases in the number of works added to the ORCID registry do not typically correspond with significant increases in the number of registered ORCID iDs. However, this is not true for the humanities, where peaks in the number of ORCID works added to the registry closely track increases in the number of ORCID iDs. This is likely because ORCID has fewer platform integrations in systems relevant to the humanities and social sciences, meaning there are fewer tools for researchers in those fields to add works to their ORCID records. Additionally, sources like Scopus and Web of Science have a well-documented STEM slant, so more of the works added via those tools will be from STEM subjects. Reducing friction for adding humanities and social sciences works to the ORCID registry should be a priority.

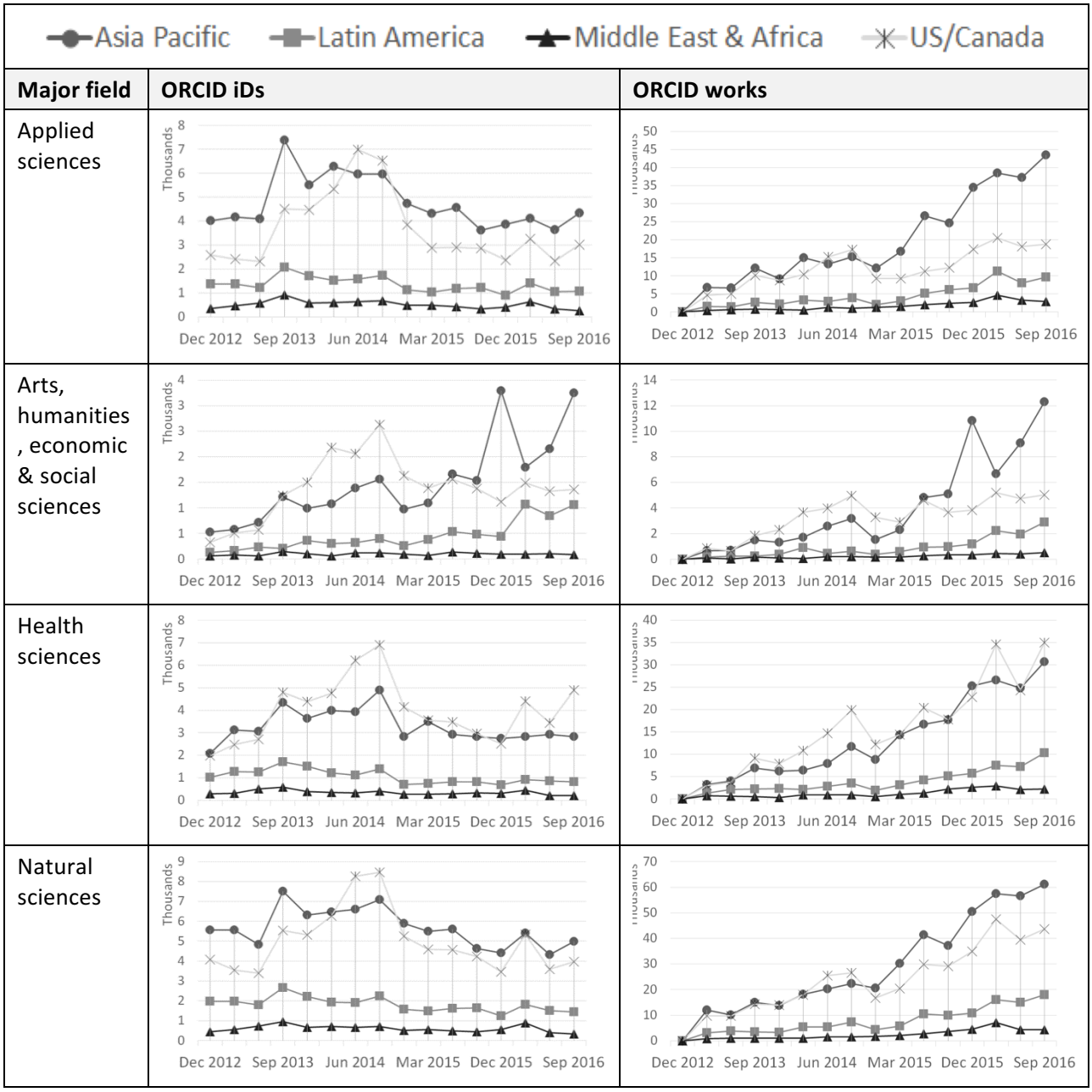

Figure 17: Quarterly full count of ORCID uptake by discipline and region, excluding Europe.

2.4 Connections Between DataCite and ORCID

To form a more complete picture of the connections between PIDs for authors and data, in keeping with the primary goal of THOR, we explored the inclusion of ORCID iDs in DataCite records through the DataCite API.

The inclusion of ORCID iDs in DataCite records is quite rare. As of 30 May, 2017, when this part of the analysis was carried out, only 3% of DataCite records contained at least one ORCID iD, meaning that at

least one of the authors or contributors of these datasets has an ORCID iD attached in the metadata of the DataCite DOI. In October 2017, this percentage was only marginally higher (almost 4%).

ORCID reporting by members is nearly uniformly low. The inclusion of ORCID iDs in DataCite records is almost negligible, with a few notable exceptions. A small number of DataCite members have actively incorporated ORCIDs. Interestingly, these are large European members: The British Library (BL)[5] (16%), Figshare[6] (9%), and the Technische Informationsbibliothek (TIB)[7] (6%). In the case of the BL and TIB, this is driven by a small number of data centres led by Imperial College and Pangaea respectively (see Figure 18).

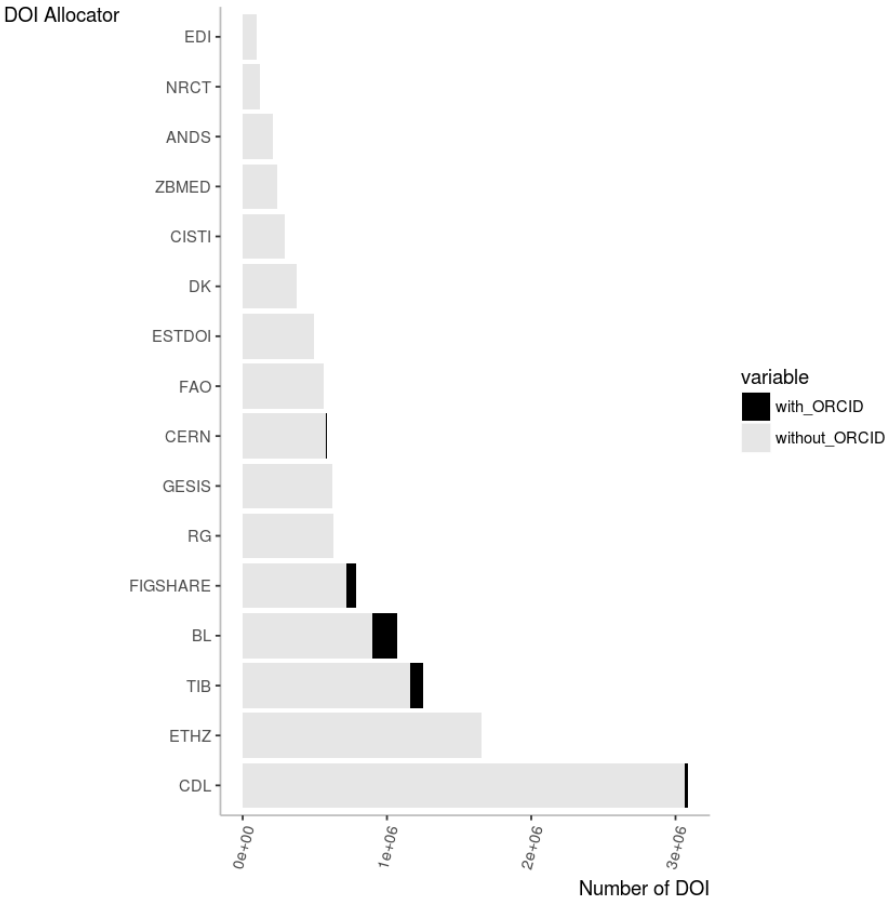

Figure 18: Proportion of DataCite DOIs that are associated with at least one ORCID iD created by member allocators with the highest number of DOIs

ORCID assignment is not evenly distributed across members of data centres. On further inspection, DataCite's members that show a high ORCID iD inclusion rate do not owe this overall high rate to uniformly high ORCID iD inclusion across all their constituent data centres. In every case the high ORCID iD inclusion rate for a member is due instead to a single constituent data centre with an exceptionally high ORCID inclusion rate (see Figure 19). Three particular data centres have strong integrations: Imperial College, Figshare and Pangaea. This is due to the number of datasets in their repository

[5] https://www.bl.uk

[6] https://figshare.com

[7] https://www.tib.eu

(Imperial), the level of adoption of ORCID iDs among their communities (Figshare) and their engagement (Pangaea as a THOR partner).

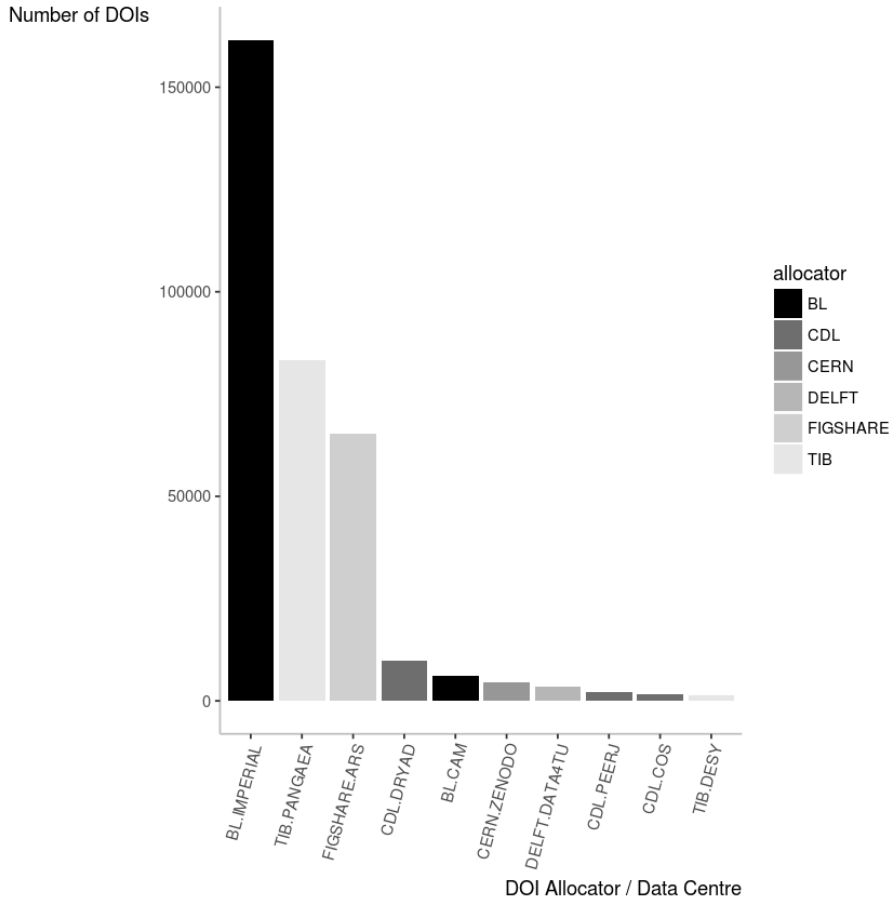

Figure 19: Data centres with the highest number of DOIs with at least one ORCID associated.

The experience of leading ORCID adopters for data sets should serve as a basis for articulating guidelines for the development of new integrations. In the upcoming, years, more and more repositories will take this step, but additional effort will be needed to understand how to best support them through the interoperable services developed by THOR and in other ways.

2.5 Summary

The aim of this study is to understand where and how ORCID iDs are being used. We studied those records from the 4-year period from October 2012 to September 2016 for which discipline and region could be determined.

Results from these datasets show that the representation of ORCID iDs within the natural, health and applied sciences is higher than in the arts, humanities, economic and social sciences. This is in keeping with results from the ORCID survey on community perceptions and understanding of ORCID published at the end of 2015 (Armstrong, Haak, Meadows, & Stone 2015). Of the 6000 respondents, 14% were from

the social sciences and humanities (compared to 41% in the physical sciences). The disciplinary distribution of these responses reflects a greater skew towards social sciences and humanities than the CWTS reference data, though the representation of the physical sciences in the ORCID survey is roughly on par with the reference data (see Figure 6). Approximately half (~51%) of humanities respondents were either "unfamiliar" with ORCID or "didn't know [ORCID] at all", as opposed to ~42% of the respondents in the physical sciences (Meadows, 2015).

The proportion of ORCID users in the humanities community has more than doubled from 4.1% in December 2012 to 9.6% in September 2016, however the same is not true for the works attached to their ORCID record. They accounted for 3.8% of ORCID works in March 2012 (the first quarter for which there are any data on works) and 5.5% in September 2016. This discrepancy might be explained by the fact that while ongoing outreach activities by THOR and ORCID to raise awareness amongst researchers in these disciplines is proving successful, there is still a relative lack of services for this field.

The Modern Language Association International Bibliography[8] became the first humanities-focused ORCID integration in June 2015. THOR has begun a series of workshops and interviews to better understand the requirements for persistent identifier services in the humanities, the first of which took place at the British Library in December 2016. Insights from these events will not only help to identify what kind of services this community needs but also help identify further training requirements.

Interestingly, while the distribution across disciplines in our data is broadly similar to the pattern of research activity in the CWTS reference dataset, the proportion of arts, humanities, economic & social science activity in the ORCID data is relatively high – 9.6% vs 4.1% in the reference dataset. This may be explained in part by the decision to match ORCID works with an ISBN identifier to academic field using Classify as an intermediate step, instead of searching for the ISBNs directly in WoS. Classify enabled us to access library collections beyond the remit of scientific literature databases and therefore likely resulted in a higher proportion of matches than would have been possible using WoS.

In the location analysis, we observed that ORCID records listing Europe as a location far outstrip the rest of the world, and listings are fewest in the southern hemisphere in this dataset. Work has already begun at ORCID to understand particular challenges and barriers to adoption in these regions and to address them. For example, in the Middle East & Africa the capacity for building ORCID integrations remains a challenge, despite interest in and awareness of the benefits to research and researchers. As mentioned previously, the adoption rate in Europe is highest in countries that have a national framework for ORCID.

Moves towards a national consortium approach are underway in several countries and ORCID's outreach activities are continuing to raise awareness. In 2016, Latin America accounted for 8% of traffic to the ORCID registry, an increase over the previous year likely due to ORCID's presence in the area.

Our trend analyses for both region and discipline show not only that ORCID uptake is on the rise overall, but also that specific interventions such as national mandates for ORCID use and the launch of new connections between services and providers, increase this trend effectively. It therefore makes sense that ORCID should continue to direct attention to these effective endeavours. Research is already underway to understand more about the provenance of useful connections between ORCID and other research information systems. This information together with the data on discipline and region from this study should inform future activity.

[8] https://www.mla.org/Publications/MLA-International-Bibliography

3 Discussion

The analysis of trends in ORCID data helps provide an understanding of the use and exploitation of ORCID iDs beyond the general aggregated trends available via the THOR dashboard (Dallmeier-Tiessen & Dasler, 2016). While the number of ORCID iDs overall has continued to rise steadily, this rise is not consistent across disciplines and regions. Whereas some disciplines (e.g. the natural sciences) seem to be at the forefront of ORCID adoption and can be considered "drivers" for the integration and usage of ORCID iDs, others might be in need of more support, further awareness-building or specific solutions. The specific causes for lagging adoption should be investigated; this may involve additional qualitative effort in order to develop a more nuanced picture.

The ORCID analysis underlined that those community activities jointly organised by ORCID and publishers or consortia have a considerable impact on the uptake of ORCID iDs. One lesson learnt from this analysis could be to better leverage such endeavours in order to help underrepresented communities to profit from ORCID iDs.

Institutional and funder data are emerging areas of interest that merit further exploration. In the future, it will be crucial to complement our existing views of the PID landscape with this missing institutional or funder data, but currently there are too many unknowns to conduct a reliable analysis with the existing datasets. Both ORCID and DataCite are currently exploring issues in this area, and both are founding members of the initiative to develop IDs for organisations (Fenner, 2016 and Meadows, 2017).

4 Conclusion

Conducting this study of ORCID adoption has provided the THOR project with valuable insights for ORCID and for the project as a whole. The identification of gaps in ORCID's coverage of disciplines and geographic regions highlights needs for targeted outreach efforts, or perhaps even specialised services, to reach areas currently underserved by ORCID. Efforts like these were realised in the outreach and service development activities of THOR, and this analysis emphasises the need for these types of efforts to continue. Additionally, through the process of conducting this analysis, the gaps in readily available data, such as institutional or funder data, have been further highlighted, showing the need for future work to close these gaps.

5 Acknowledgements

We would like to express our gratitude to a number of organisations external to THOR for their invaluable contributions to the data analysis in this study. In particular, our appreciation goes to Joseph Brightbill, for his work on data pre-processing and discipline assignment, and Patricia Brennan at Clarivate Analytics; and to Rodrigo Costas at CWTS for providing the reference data and advice on methodology for the ORCID analyses.

6 References

Armstrong, D., Haak, L., Meadows, A., & Stone, A. (2015). ORCID 2015 Survey Report (final). Figshare. https://doi.org/10.6084/m9.figshare.2008206.v1

Bohannon, J. (2017). Vast set of public CVs reveals the world's most migratory scientists. *Science*. https://doi.org/10.1126/science.aal1189

Caron, E., & van Eck, N.-J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. In *Proceedings of the Science and Technology Indicators Conference 2014* (pp. 79–86). Leiden.

Dallmeier-Tiessen, S., & Dasler, R. (2016). Analysis and comparison of persistent identifier use and integration across disciplines and sectors. https://doi.org/10.5281/zenodo.154592

Dasler, R. (2016). THOR: Metrics and tools. THOR. https://doi.org/10.5281/zenodo.46761

Deane-Pratt, A. (2017). ORCID data to accompany study of PID service adoption across disciplines and locations [Data file]. https://doi.org/10.5281/zenodo.841761

Fenner, M. (2016, November 1). Announcing the organization identifier project: A way forward. Retrieved from https://blog.datacite.org/announcing-organization-identifier-project/

Haak, L. (2015, August 3). 1.5 Million ORCID iDs served. Retrieved from https://orcid.org/blog/2015/07/31/15-million-orcid-ids-served

Haak, L., Brown, J., Buys, M., Cardoso, A. P., Demain, P., Demeranville, T., … Wright, D. (2016). ORCID public data file 2016 [Data file]. https://doi.org/10.6084/m9.figshare.4134027.v1

Klein, M., & Van de Sompel, H. (2017). Discovering scholarly orphans using ORCID. *arXiv:1703.09343 [cs]*. Retrieved from http://arxiv.org/abs/1703.09343

Larivière, V., & Costas, R. (2016). How many is too many? On the relationship between research productivity and impact. *PLOS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0162709

Meadows, A., Armstrong, D. (2015). ORCID 2015 survey. figshare. https://doi.org/10.6084/m9.figshare.2007522.v1

Meadows, A. (2017, March 15). Organization identifier working group. Retrieved from https://orcid.org/content/organization-identifier-working-group

Meadows, A., Brown, J., Haak, L., Paglione, L., Peters, R., & Wright, D. (2017). 2016 Annual report: Reflecting on the journey toward ORCID's vision. https://doi.org/10.6084/m9.figshare.4810213.v2

Ruiz-Castillo, J., & Costas, R. (2014). The skewness of scientific productivity. *Journal of Informetrics*, *8*(4), 917–934. https://doi.org/10.1016/j.joi.2014.09.006

Appendix A: Glossary

Term	Definition
Creation date for ORCID iDs	The date the record was added to the ORCID registry.
Creation date for ORCID works	The date the work was added to an ORCID record.
Cumulative full count	Running total of the full count of ORCID works or iDs according to their creation date.
Disambiguated authors	The estimated number of individual authors following computational analysis that clusters publications by their author names in order to try to uniquely identify them. (Ruiz-Castillo & Costas, 2014)
Full count of ORCID iDs	Each person is counted once for every discipline in which they have at least one work.
Full count of ORCID works	Each publication is counted once for each discipline it is classified as covering.
Quarterly full count	The number of ORCID iDs or works created during each quarter only, following full counting for each discipline they belong to.

Appendix B: External identifiers supported by ORCID

The full list of external identifier types supported for ORCID works:

- other-id: A unique identifier, used identifiers types not included in this list
- agr: Agricola identifier
- arxiv: ArXiv.
- asin: Amazon Standard Identification Number.
- asin-tld: ASIN top-level domain for Amazon sites other than the US; valid values: co.jp co.uk ca cn fr de it es
- bibcode: Bibcode; used by a number of astronomical data systems; example: 1924MNRAS..84..308E.
- cba: Chinese biological abstracts identifier
- cit: CiteSeer
- ctx: CiteXplore submission
- doi: Digital object identifier; example: 10.1038/news070508-7.
- eid: Identifier used by Scopus.
- ethos: EThOS Persistent identifier
- handle: Handle system
- hir: NHS Evidence identifier
- isbn: International Standard Book Number such as 978-0812695939.
- issn: International Standard Serial Number. [ISSN is not recommended to be included with journal articles]
- jfm: Jahrbuch über die Fortschritte der Mathematik
- jstor: JSTOR abstract.
- lccn: Library of Congress Control Number
- mr: Mathematical Reviews
- oclc: Online Computer Library centre
- ol: Open Library
- osti: Office of Scientific and Technical Information
- pat: patent number
- pmc: PubMed Central article number for full-text free repository of an article
- pmid: PubMed Unique Identifier
- rfc: Request for Comments
- source-work-id: local identifier. This field should be used when no standard identifiers exist for the work
- ssrn: Social Science Research Network
- uri: Uniform resource identifier
- urn: Uniform resource name
- wosuid: Identifier used by Web of Science™
- zbl: Zentralblatt MATH.

Appendix C: Discipline map used in the ORCID analyses

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
Applied Sciences	Agriculture, Fisheries & Forestry	Agricultural economics & policy	-	-
		Agricultural experiment station reports	-	-
		Agriculture - dairy & animal science	SF	Animal husbandry. Animal science
		Agriculture - multidisciplinary	S	Agriculture (general)
		Agronomy	-	-
		Biodiversity conservation	-	-
		Fisheries	SH	Aquaculture. Fisheries. Angling
		Forestry	SD	Forestry. Arboriculture. Silviculture
		Horticulture	SB	Horticulture. Plant propagation. Plant breeding
		Soil science	-	-
		Veterinary sciences	-	-
	Built Environ-ment & Design	Architecture	NA	Architecture
		Planning & development	-	-
	Engineering	Agricultural engineering	-	-
		Automation & control systems	-	-
		Construction & building technology	TG	Bridges
			TH	Building construction
		Engineering - aerospace	UG	Military engineering. Air forces
		Engineering - chemical	-	-
		Engineering - civil	TA	Engineering civil engineering (general).
		Engineering - electrical & electronic	TK	Electrical engineering. Electronics. Nuclear engineering
		Engineering - environmental	TD	Environmental technology. Sanitary engineering
		Engineering - geological	-	-
		Engineering - industrial	-	-
		Engineering - manufacturing	TS	Manufacturing engineering. Mass production
		Engineering - marine	-	-
		Engineering - mechanical	TJ	Mechanical engineering and machinery

[9] Level 1 and Level 2 are taken from Science Metrix. (See Section 2.1.2)

[10] Library of Congress classification codes

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			TL	Motor vehicles. Aeronautics. Astronautics
		Engineering - multidisciplinary	-	-
		Engineering - ocean	TC	Hydraulic engineering. Ocean engineering
		Engineering - petroleum	-	-
		Metallurgy & metallurgical engineering	TN	Mining engineering. Metallurgy
		Mining & mineral processing	-	-
		Transportation	HE	Transportation and communications
		Transportation science & technology	TE	Highway engineering. Roads and pavements
			TF	Railroad engineering and operation
	Information & Communication Technologies	Computer applications & cybernetics	-	-
		Computer science - artificial intelligence	-	-
		Computer science - hardware & architecture	-	-
		Computer science - information systems	-	-
		Computer science - interdisciplinary applications	-	-
		Computer science - software engineering	-	-
		Computer science - theory & methods	-	-
		Information science & library science	Z	Books (general). Writing. Paleography. Book industries and trade. Libraries. Bibliography
			ZA	Information resources/materials
		Telecommunications	-	-
	Technology & other Applied	Biotechnology & applied microbiology	-	-
		Energy & fuels	-	-
		Food science & technology	-	-
		Imaging science & photographic technology	TR	Photography
		Instruments & instrumentation	-	-
		Materials science - biomaterials	-	-

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
		Materials science - ceramics	-	-
		Materials science - characterization & testing	-	-
		Materials science - coatings & films	-	-
		Materials science - composites	-	-
		Materials science - multidisciplinary	-	-
		Materials science - paper & wood	-	-
		Materials science - textiles	-	-
		Medical laboratory technology	-	-
		Microscopy	-	-
		Multidisciplinary sciences	-	-
		Nanoscience & nanotechnology	-	-
		Nuclear science & technology	-	-
		Remote sensing	-	-
		Robotics	-	-
		Spectroscopy	-	-
Arts, humanities, economic & social	Arts & General Humanities	Art	N	Visual arts
			NB	Sculpture
			NC	Drawing. Design. Illustration
			ND	Painting
			NE	Print media
			NH	Fine arts
			NK	Decorative arts
			NX	Arts in general
			TT	Handicrafts. Arts and crafts
		Dance	-	-
		Film - radio - television	-	-
		Folklore	-	-
		Humanities - multidisciplinary	AC	Collections. Series. Collected works
			AE	Encyclopedias
			AG	Dictionaries and other general reference works
			AI	Indexes
			AM	Museums. Collectors and collecting
			AN	Newspapers
			AP	Periodicals
			AS	Academies and learned societies
			AW	General works
			AY	Yearbooks. Almanacs. Directories

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			AZ	History of scholarship and learning. The humanities
		Music	M	Music
			ML	Literature on music
			MT	Instruction and study
		Poetry	-	-
		Theater	-	-
	Communi-cation	Communication	-	-
		Language & linguistics theory	-	-
		Linguistics	P	Philology. Linguistics
			PC	Romanic languages
			PD	Germanic languages. Scandinavian languages
			PE	English language
			PF	West germanic languages
			PG	Slavic languages and literatures. Baltic languages. Albanian language
			PH	Uralic languages. Basque language
			PM	Hyperborean; native american; and artificial languages
		Literary reviews	-	-
		Literary theory & criticism	-	-
		Literature	PA	Greek language and literature. Latin language and literature
			PB	Modern languages. Celtic languages and literature
			PJ	Oriental languages and literatures
			PK	Indo-iranian languages and literatures
			PL	Languages and literatures of eastern asia; africa; oceania
			PN	Literature (general)
			PZ	Fiction and juvenile belles lettres
		Literature - african - australian - canadian	-	-
		Literature - american	PS	American literature
		Literature - british isles	PR	English literature
		Literature - german - dutch - scandinavian	PT	German literature. Dutch literature. Flemish literature since 1830. Afrikaans literature -scandinavian literature. Old norse literature: old icelandic and old norwegian. Modern icelandic literature. Faroese literature. Danish literature. Norwegian literature. Swedish literature

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
		Literature - romance	PQ	French literature. Italian literature. Spanish literature. Portuguese literature
		Literature - slavic	-	-
	Economics & Business	Business	HF	Commerce
		Business - finance	HG	Finance
		Economics	HJ	Public finance
		Hospitality - leisure - sport & tourism	SK	Hunting
		Management	-	-
		Operations research & management science	-	-
	Historical Studies	Anthropology	GF	Human ecology. Anthropogeography
			GN	Anthropology
			GR	Folklore
			GT	Manners and customs (general)
			GV	Recreation. Leisure
		Archaeology	CC	Archaeology
		Area studies	-	-
		Classics	-	-
		History	C	Auxiliary sciences of history (general)
			CB	History of civilization
			CD	Diplomatics. Archives. Seals
			CE	Technical chronology. Calendar
			CJ	Numismatics
			CN	Inscriptions. Epigraphy
			CR	Heraldry
			CS	Genealogy
			CT	Biography
			D	History (general)
			DA	Great britain
			DAW	Central europe
			DB	Austria; liechtenstein; hungary; czechoslovakia
			DC	France; andorra; monaco
			DD	Germany
			DE	Greco-roman world
			DF	Greece
			DG	Italy; malta
			DH	Low countries; benelux countries
			DJ	Netherlands (holland)
			DJK	Eastern europe (general)
			DK	Russia. Soviet union. Former soviet republics; poland

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			DL	Northern europe. Scandinavia
			DP	Spain; portugal
			DPC	History of spain & portugal
			DPV	History of spain & portugal
			DQ	Switzerland
			DR	Balkan peninsula
			DS	Asia
			DT	Africa
			DU	Oceania (south seas)
			DX	History of the romani people
			E	History of america
			F	Local history of the united states and british; dutch; french; and latin america
			KL	History of law. The ancient orient
		History & philosophy of science	-	-
		History of social sciences	-	-
		Medieval & renaissance studies	-	-
	Philosophy & Theology	Ethics	BJ	Ethics
		Philosophy	B	Philosophy (general)
			BD	Speculative philosophy
			BH	Aesthetics
		Religion	BL	Religions. Mythology. Rationalism
			BM	Judaism
			BP	Islam. Bahaism. Theosophy etc
			BQ	Buddhism
			BR	Christianity
			BS	The bible
			BT	Doctrinal theology
			BV	Practical theology
			BX	Christian denominations
			KB	Religious law in general. Comparative religious law. Jurisprudence
			KBM	Jewish law
			KBP	Islamic law
			KBR	History of canon law
			KBS	Canon law of eastern churches
			KBT	Canon law of eastern rite churches in communion with the holy see of rome
			KBU	Law of the roman catholic church. The holy see

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
	Social Sciences	Asian studies	-	-
		Criminology & penology	-	-
		Cultural studies	-	-
		Demography	HB	Economic theory. Demography
		Education - scientific disciplines	-	-
		Education - special	LC	Special aspects of education
		Education & educational research	L	Education (general)
			LA	History of education
			LB	Theory and practice of education
			LD	Individual institutions - united states
			LE	Individual institutions - america (except united states)
			LF	Individual institutions - europe
			LG	Individual institutions - asia; africa; indian ocean islands; australia; new zealand; pacific islands
			LGF	Education - asia, africa, pacific
			LH	College and school magazines and papers
			LJ	Student fraternities and societies; united states
			LN	Education - asia, africa, pacific
			LT	Textbooks
		Ergonomics	-	-
		Ethnic studies	-	-
		Family studies	-	-
		Industrial relations & labor	HD	Industries. Land use. Labor
		International relations	JZ	International relations
		Law	JX	International law; see jz and kz (obsolete)
			K	Law in general. Comparative and uniform law. Jurisprudence
			KC	Law
			KD	United kingdom
			KDC	Law - scotland
			KDK	Ireland
			KDZ	America. North america
			KE	Canada
			KF	United states
			KFM	Law - usa
			KFN	Law - usa
			KFO	Law - usa

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			KG	Latin america; mexico and central america; west indies. Caribbean area
			KGF	Law - mexico
			KGQ	Law - dominican republic
			KGV	Law - puerto rico
			KH	South america
			KHA	Law - argentina
			KHD	Law - brazil
			KHF	Law - chile
			KHH	Law - colombia
			KHK	Law - ecuador
			KHQ	Law - peru
			KHU	Law - uruguay
			KHW	Law - venezuela
			KJ	Europe
			KJA	Roman law
			KJC	Regional comparative and uniform law
			KJE	Regional organization and integration. Comparative law
			KJG	Law - albania
			KJJ	Law - austria
			KJK	Law - belgium
			KJN	Law - cyprus
			KJP	Law - czechosolovakia
			KJQ	Law - european
			KJR	Law - denmark
			KJV	European national laws
			KK	Europe
			KKF	Law - hungary
			KKH	Law - italy
			KKJ	Law - lichtenstein
			KKM	Law - netherlands
			KKQ	Law - portugal
			KKR	Law - romania
			KKS	Law - san marino
			KKT	Law - spain
			KKV	Law - sweden
			KKW	Law - switzerland
			KKX	Law - turkey
			KKY	Law - ukraine
			KLA	Law - russia
			KLB	Law - russia
			KLR	Law - kazakhstan
			KM	General. Middle east. Southwest asia

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			KMC	Regional comparative and uniform law
			KN	South asia. Southeast asia. East asia
			KNC	Regional comparative and uniform law
			KNP	Law - china
			KNQ	Law - china
			KNR	Law - hong kong
			KNS	Law - india
			KNW	Law - indonesia
			KNX	Law - japan
			KP	South asia. Southeast asia. East asia
			KPG	Law - malaysia
			KPP	Law - singapore
			KQ	Africa
			KQC	Regional comparative and uniform law
			KQW	Law - cameroon
			KR	Africa
			KS	Africa
			KSW	Law - morocco
			KSX	Law - mozambique
			KT	Africa
			KTA	Law - nigeria
			KU	Australia and new zealand
			KUA	Law - australia
			KUC	Law - australia
			KUD	Law - australia
			KUG	Law - australia
			KUQ	Law - new zealand
			KV	Pacific area including antarctica
			KVC	Law - other pcfic area jurisdictions
			KW	Pacific area including antarctica
			KWX	Law - antarctica
			KZ	Law of nations
			KZA	Law of the sea
		Medical ethics	-	-
		Political science	HX	Socialism Communism Anarchism
			J	General legislative and executive papers
			JA	Political science (general)
			JC	Political theory
			JF	Political institutions and public administration

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			JJ	Political institutions and public administration (north america)
			JK	Political institutions and public administration (united states)
			JL	Political institutions and public administration (canada; latin america; etc.)
			JN	Political institutions and public administration (europe)
			JQ	Political institutions and public administration (asia; africa; australia; pacific area; etc.)
			JS	Local government. Municipal government
			JV	Colonies and colonisation. Emigration and immigration. International migration
		Public administration	U	Military science (general)
			UA	Armies: organization; distribution; military situation
			UB	Military administration
			UC	Military maintenance and transportation
			UD	Infantry
			UE	Cavalry. Armor
			UF	Artillery
			UH	Other military services
			V	Naval science (general)
			VA	Navies: organization; distribution; naval situation
			VB	Naval administration
			VC	Naval maintenance
			VD	Naval seamen
			VE	Marines
			VF	Naval ordnance
			VG	Minor services of navies
			VK	Navigation. Merchant marine
			VM	Naval architecture. Shipbuilding. Marine engineering
		Rehabilitation	-	-
		Social issues	HN	Social history and conditions. Social problems. Social reform
			HV	Social pathology. Social and public welfare. Criminology
		Social sciences - biomedical	-	-
		Social sciences - interdisciplinary	H	Social sciences (general)

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			HC	Economic history and conditions
			TX	Home economics
		Social sciences - mathematical methods	HA	Statistics
		Social work	-	-
		Sociology	HM	Sociology (general)
			HQ	The family. Marriage; women and sexuality
			HS	Societies: secret; benevolent; etc.
			HT	Communities. Classes. Races
		Sport sciences	-	-
		Urban studies	-	-
		Women's studies	-	-
Health Sciences	Biomedical Research	Biochemical research methods	-	-
		Cell & tissue engineering	-	-
		Chemistry - medicinal	-	-
		Cytology & histology	-	-
		Engineering - biomedical	-	-
		Immunology	-	-
		Infectious diseases	-	-
		Medical informatics	-	-
		Medicine - research & experimental	-	-
		Nutrition & dietetics	-	-
		Pathology	RB	Pathology
		Toxicology	-	-
	Clinical Medicine	Allergy	-	-
		Andrology	-	-
		Anesthesiology	-	-
		Cardiac & cardiovascular systems	-	-
		Clinical neurology	-	-
		Critical care medicine	-	-
		Dentistry/oral surgery & medicine	RK	Dentistry
		Dermatology	RL	Dermatology
		Emergency medicine	-	-
		Endocrinology & metabolism	-	-
		Gastroenterology & hepatology	-	-
		Geriatrics & gerontology	-	-
		Hematology	-	-
		Integrative & complementary medicine	RV	Botanic; thomsonian; and eclectic medicine
			RX	Homeopathy

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
			RZ	Other systems of medicine
		Medicine - general & internal	R	Medicine (general)
			RC	Internal medicine
		Neuroimaging	-	-
		Obstetrics & gynecology	RG	Gynecology and obstetrics
		Oncology	-	-
		Ophthalmology	RE	Ophthalmology
		Orthopedics	-	-
		Otorhinolaryngology	RF	Otorhinolaryngology
		Pediatrics	RJ	Pediatrics
		Peripheral vascular disease	-	-
		Pharmacology & pharmacy	RM	Therapeutics. Pharmacology
			RS	Pharmacy and materia medica
		Psychiatry	-	-
		Radiology & nuclear medicine	-	-
		Respiratory system	-	-
		Rheumatology	-	-
		Surgery	RD	Surgery
		Transplantation	-	-
		Tropical medicine	-	-
		Urology & nephrology	-	-
	Psychology & Cognitive Sciences	Behavioral sciences	-	-
		Psychology	-	-
		Psychology - applied	-	-
		Psychology - biological	-	-
		Psychology - clinical	-	-
		Psychology - developmental	-	-
		Psychology - educational	-	-
		Psychology - experimental	-	-
		Psychology - mathematical	-	-
		Psychology - multidisciplinary	BF	Psychology
		Psychology - psychoanalysis	-	-
		Psychology - social	-	-
	Public Health & Health Services	Audiology & speech-language pathology	-	-
		Gerontology	-	-
		Health care sciences & services	-	-
		Health policy & services	-	-
		Medicine - legal	-	-
		Nursing	RT	Nursing
		Primary health care	-	-
		Public - environmental & occupational health	RA	Public aspects of medicine

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
		Substance abuse	-	-
Natural Sciences	Biology	Anatomy & morphology	QM	Human anatomy
		Biochemistry & molecular biology	-	-
		Biology	-	-
		Biophysics	-	-
		Cell biology	-	-
		Developmental biology	-	-
		Entomology	-	-
		Evolutionary biology	-	-
		Genetics & heredity	-	-
		Microbiology	QR	Microbiology
		Mycology	-	-
		Neurosciences	-	-
		Ornithology	-	-
		Parasitology	-	-
		Physiology	QP	Physiology
		Reproductive biology	-	-
		Virology	-	-
		Zoology	QL	Zoology
	Chemistry	Chemistry - analytical	-	-
		Chemistry - applied	TP	Chemical technology
		Chemistry - inorganic & nuclear	-	-
		Chemistry - multidisciplinary	QD	Chemistry
		Chemistry - organic	-	-
		Chemistry - physical	-	-
		Electrochemistry	-	-
		Polymer science	-	-
	Earth & Environ-mental Sciences	Ecology	-	-
		Environmental sciences	GE	Environmental sciences
		Environmental studies	-	-
		Geochemistry & geophysics	-	-
		Geography	G	Geography (general). Atlases. Maps
			GA	Mathematical geography. Cartography
		Geography - physical	GB	Physical geography
		Geology	QE	Geology
		Geosciences - multidisciplinary	-	-
		Green & sustainable science & technology	-	-
		Limnology	-	-
		Marine & freshwater biology	-	-

Level 1[9]	Level 2[9]	Level 3 (WoS subject)	LoC[10] code	LoC subject
		Mineralogy	-	-
		Oceanography	GC	Oceanography
		Paleontology	-	-
		Plant sciences	QK	Botany
		Water resources	-	-
	Mathe-matics & Statistics	Logic	BC	Logic
		Mathematical & computational biology	-	-
		Mathematics	QA	Mathematics
		Mathematics - applied	-	-
		Mathematics - interdisciplinary applications	-	-
		Physics - mathematical	-	-
		Statistics & probability	-	-
	Physics & Astronomy	Acoustics	-	-
		Astronomy & astrophysics	QB	Astronomy
		Crystallography	-	-
		Mechanics	-	-
		Meteorology & atmospheric sciences	-	-
		Optics	-	-
		Physics - applied	-	-
		Physics - atomic - molecular & chemical	-	-
		Physics - condensed matter	-	-
		Physics - fluids & plasmas	-	-
		Physics - multidisciplinary	QC	Physics
		Physics - nuclear	-	-
		Physics - particles & fields	-	-
		Thermodynamics	-	-