

NOVA University of Newcastle Research Online

nova.newcastle.edu.au

Gu, Xin; Blackmore, Karen. "Characterisation of academic journals in the digital age". Published in Scientometrics Vol. 110, Issue 3, p. 1333-1350 (2017)

Available from: <u>http://dx.doi.org/10.1007/s11192-016-2219-4</u>

This is a post-peer-review, pre-copyedit version of an article published in Scientometrics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s11192-016-2219-4

Accessed from: http://hdl.handle.net/1959.13/1349365

Characterisation of Academic Journals in the Digital Age

Xin Gu

xin.gu@uon.edu.au

Karen Blackmore karen.blackmore@newcastle.edu.au

School of Design, Communication and Information Technology

The University of Newcastle University Drive Callaghan NSW 2308 Australia

Abstract

Innovations in scholarly publishing have led to new possibilities for academic journals (e.g., open access), and provided scholars with a range of indicators that can be used to evaluate their characteristics and their impact. This study identifies and evaluates the journal characteristics reported in five databases: Ulrich's Periodicals Directory (*Ulrichs*), Journal Citation Reports (*JCR*), SCImago Journal & Country Rank (*SJR*), Google Scholar Metrics (*GS*), and Cabell's Periodical Directory (*Cabells*). It describes the 13 indicators (variables) that are available through these databases—scholarly impact, subject category, age, total articles, distribution medium, open access, peer review, acceptance rate, pricing, language, country, status, and issue frequency—and highlights the similarities and differences in the ways these indicators are defined and reported. The study also addresses the ways in which this kind of information can be used to better understand particular journals as well as the scholarly publishing system.

Keywords

Academic journal, Attributes, Impact factor, Ranking, Journal age, Acceptance rate

MSC

94A99

JEL

I250

Introduction

Academic journals serve as forums for the introduction of new research, as well as the evaluation and critique of established theories, methods, and findings (Blake & Bly, 1993). The scholarly communication system certifies additions to the body of accepted knowledge. At the same time, the system provides a mechanism by which individual scholars compete for priority and recognition (Hargens, 1988; Rice & Stankus, 1983).

Since the publication of the first modern academic journal 350 years ago (Solomon, 2007), the number of active academic titles has grown by an average of 3.3% per year (Mabe & Amin, 2001). More recently, new technologies have led to major changes in the scholarly publishing system (Nisonger, 1998): online journals (Björk & Hedlund, 2004; Statzner & Resh, 2010), large full-text databases (Bar-Ilan, 2008), digital submission systems, e-repositories (Statzner & Resh, 2010), open access journals and archives (Guerrero & Piqueras, 2010), and research-focused search engines such as Google Scholar (Larsen & von Ins, 2010). The increased connectivity of the scholarly community has also brought greater international collaboration among authors and journal editors (Lee & Bozeman, 2005; Statzner & Resh, 2010).

These changes in the scholarly publication system have been accompanied by changes in the external environment—in particular, a greater emphasis on accountability and a culture of assessment that focuses on the performance of individuals, institutions, and publication outlets. For instance, university ranking schemes and performance comparisons have had a significant impact on the international publishing system (Adler & Harzing, 2009; Hazelkorn, 2013). These same trends inform institutional decision-making, especially with regard to the allocation of resources, and influence the career paths of individual scholars (Statzner & Resh, 2010). At the same time, a second boom in journal publication has commenced, with the number of academic journals growing at an annual rate of 3.7% over the past 30 years (Gu & Blackmore, 2016).

The information available to authors, editors, librarians, and evaluators of research has also increased over time. Well-established sources of information about journal characteristics and impact—Ulrich's Periodicals Directory (*Ulrichs*) and Journal Citation Reports (*JCR*), for example—coexist with new databases such as Cabell's Periodical Directory (*Cabells*), Google Scholar Metrics (*GS*), and Elsevier's SCImago Journal & Country Rank (*SJR*). Moreover, the impact factor, introduced in 1955 (Garfield, 2006; Starbuck, 2005), no longer stands unrivalled as the foremost indicator of scholarly influence. New metrics such as the Eigenfactor, SNIP, and SJR permit the comparison of citation impact across disciplines (Colledge et al., 2010; Elsevier, 2016; MacRoberts & MacRoberts, 1989; Thomson Reuters, 2016). Perhaps more importantly, many of these indicators are freely available online.

Changes in the journal publishing landscape have created a dynamic system in which individual actors (agents) seek to maximise their performance, usually in terms of standard publication or citation metrics. Modelling approaches such as agent-based modelling may help clarify and even predict the behaviour of actors as well as the performance of the system as a whole (Axelrod, 1997; Bonabeau, 2002; Gu, Blackmore, Cornforth, & Nesbitt, 2015; Macal & North, 2009; Mölders, Fink, & Weyer, 2011). Increasing specialization, the proliferation of journals, a culture of assessment, and greater awareness of journal characteristics and citation metrics have led to heightened competition for space in the top journals as well as competition among journals to capture the most important contributions to the literature (Statzner & Resh, 2010).

This study identifies and evaluates the journal characteristics (variables) reported in *Ulrichs, JCR, SJR, GS*, and *Cabells*. Specifically, it describes 13 key indicators that can be found within those databases: scholarly impact, subject category, age, total articles, distribution medium, open access, peer review, acceptance rate, pricing, language, country, status, and issue frequency. We highlight the similarities and differences in the ways these indicators are defined and reported, since each variable—subject category and scholarly impact, for instance—can vary significantly among the five data sources. Finally, we address the ways in which information of this type can be used to better understand the scholarly publishing system.

Methodology

Methodology section includes two parts, data collection and data consolidation.

Data collection

The data analysis approach is a practical method to understand the dimensions of academic journals that

involves. In this section, the detailed process of the data collection and analysis approach to capture the relevant attributes of academic journals are provided. First of all, we identified the major journal data sources and download relevant data from these data sources. Then we consolidated these data from several sources into one based on *ISSN* and *Journal Title*. Consequently, we conducted data analysis over the consolidated data source. Eventually, the relevant attributes of academic journals are identified.

Based on literature search and Internet search, we discovered five major journal data sources. Although there are many other data sources, only five data sources are adopted for this study due to their wide coverage. For example, Ulrich's Periodicals Directory (*Ulrichs*) has wide coverage than The Serial Directory, and they both serve the same purpose, therefore only *Ulrichs* is adopted for this study (Tuttle, 1987). The data sources, representative references, and URLs are listed in Table 1.

Data Source	Reference & URL
Ulrich's Periodicals Directory	Jinha (2010); Tenopir and King (2009a)
(Ulrichs)	http://ulrichsweb.serialssolutions.com
Journal Citation Report	Leydesdorff (2006); Solari and Magri (2000)
(JCR)	http://thomsonreuters.com/journal-citation-reports/
SCImago Journal & Country Rank	Guz and Rushchitsky (2009)
(SJC)	http://www.scimagojr.com/journalrank.php
Google Scholar Metrics	Delgado-López-Cózar and Cabezas-Clavijo (2012).
(<i>GS</i>)	https://scholar.google.com/citations?view_op=top_venues
Cabells	Kapelianis and Cabell (1999)
(Cabells)	https://ssl2.cabells.com

Table 1 List of academic journal data sources

In this study, we collected data on academic journals from five different data sources, these being *Ulrichs*, *JCR*, *SJR*, *GS*, and *Cabells*. Once downloaded, the data records are stored in a MS SQL database, with records from each data source stored in an independent table within the database. These five data sources are described in the following sections.

Ulrich's Periodicals Directory (Ulrichs)

Ulrich's Periodicals Directory (Ulrichs) is regarded as the most complete and consistent source for monitoring the number of periodical titles (Jinha, 2010; Tenopir & King, 2009a). Ulrichs contains all periodicals using descriptors that include 'Active or others', 'Journal or others', and 'Academic / Scholarly or others'. In Ulrichs, the definition of Active means that the periodical title is currently actively in publishing, rather than Announced Never Published, Ceased, Forthcoming, Merged / Incorporated, Researched / Unresolved, or Suspended. The definition of Journal means that the periodical title is published in the format of journal, rather than Abstract/Index, Database, Magazine, Bulletin, Catalog, Directory, Monographic Series, Newsletter, Proceedings, Yearbook, Report, or Handbook / Manual. The definition of Academic / Scholarly means that the periodical title is publishing with academic content, rather than Bibliography, Consumer, Trade, Corporate, or Government. For Journal periodicals, there are two types of journals: Refereed / Peer Reviewed, or Not Refereed / Peer Reviewed. In this paper, our study focuses only on refereed academic journals under these definitions.

On 13 January 2016, we downloaded data from *Ulrichs*, which is available on the Web via *http://ulrichsweb.serialssolutions.com/*. These *Ulrichs* records match two search criteria: *Academic / Scholarly*, and *Journal*. A total of 129,455 records are located. Within these records, there are 52,536 duplicate records with the same *Journal Title* and *Publisher*. This occurs when a single journal title has multiple *ISSNs*, as is the case when a journal is published in multiple formats. After removing the duplicates, there are 76,919 unique *Ulrichs* records in total. Then, the *Non-Refereed* journals are removed. This results in 41,787 *Refereed Academic Journals*, which is the final dataset. The data collection and cleaning process is described in Figure 1.



Fig. 1 Workflow of data collection and cleansing on Ulrichs data dated in January 2016

Journal Citation Report (JCR), SCImago Journal & Country Rank (SJC), & Google Scholar Metrics (GS)

Ulrichs does not cover all aspects of academic journals with important measures, such as, impact factor not included. Therefore, additional data sources are introduced to capture these additional measures. Web of Science, Scopus, and Google Scholar are recognised as the three main scholarly bibliographic data sources (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Meho & Yang, 2007). Each has developed its own metrics to measure the value or quality of academic journals. The journal lists from Web of Science, Scopus, or Google Scholar not only show the ranking of the academic journals, but also reveal the total number of the journal titles. Therefore, we adopted the Journal Citation Report (*JCR*) from Web of Science, SCImago Journal & Country Rank (*SJR*) from *Scopus*, and Google Scholar Metrics (*GS*) from Google, as additional data sources for this analysis. That means GS indexes include sources that may not be appropriately considered as academic work.

Each of these data sources are accessed independently, with varying levels of completeness.

JCR can be downloaded through subscription via http://thomsonreuters.com/journal-citation-reports/ or from a subscribed university library. For each calendar year, *JCR* has two editions: *Science* and *Social Sciences*. The latest *JCR* available in January 2016 is 2014 version, due to data latency. From this dataset, there are 11,813 records in total.

SJR is free to download and available via http://www.scimagojr.com/journalrank.php. Available search parameters include *Subject Area, Subject Category, Region / Country,* and *Year.* Except for the *Year* parameter, which needs to be specific, the other parameters do not need to be specified. Available *SJR* reports for download are dated from 1999 to 2013. Like *JCR*, the latest *SJR* report accessible in January 2016 is the 2014 version, due to data latency. There are 15,911 records in total in this dataset.

GS is free to download and available on the Google Scholar (*GS*) Metrics website at https://scholar.google.com/citations?view_op=top_venues. The latest *GS* data available in January 2016 is dated on June 2015 (Google, 2016b). There is no complete list of academic journals and there is no total number of journals available at the *GS* Metrics website. The only available search parameter is *Journal Title*. Therefore, we searched the list of *Ulrichs* journals one by one based on *Journal Title* on the *GS* Metrics website. Although GS is unreliable in providing academic related articles as a bibliographic database, it is a powerful search engine to access scholarly publications from all publishers with greater coverage (Larsen & von Ins, 2010). It is included in this study for the purpose of covering the maximum number of academic journals. We discovered 10354 *GS* records based on *Ulrichs* data. Surprisingly, it was less than the number of matching records discovered in *SJR*.

Cabell's Periodical Directory (Cabells)

Cabell's Periodical Directory (*Cabells*) is a large directory of journals that includes rich information on journals, such as style/formatting guidelines, review process details, and acceptance rates for each journal included in the dataset (Thomson Reuters, 2014) . *Cabells* is on paid subscription and available via https://ssl2.cabells.com/. We carried out a search on all parameters, and downloaded all the records as at December 2014 to be consistent with other available data sources. A total of 8,566 academic journal records are identified from *Cabells*.

Data consolidation

Following data collection, we consolidated all data records from *Ulrichs, JCR, SJR, GS,* and *Cabells* into one master dataset. *Ulrichs, JCR, SJR, GS,* and *Cabells* data are saved in five data tables in MS SQL Server. *Ulrichs* contains the most journal records, therefore it is used as the primary dataset. A match based on *ISSN* and *Journal Title* was conducted across the five datasets. The match process and results is shown in Figure 2.



Fig. 2 Journal data collection and consolidation process

The following points are noted with regard to the matching process:

- *GS* journals are downloaded based on *Ulrichs* journal dataset. Therefore, there is an exact match on *Journal Title* between *Ulrichs* data and *GS* data. That process links 10,354 *GS* records with *Ulrichs*.
- Two matches are conducted between *JCR* and *Ulrichs* journals. One is based on *Journal Title*, and the other one is based on *ISSN*. The matches discover 9,583 *JCR* records in *Ulrichs*.
- Two matches are also conducted between *SJR* and *Ulrichs* journals. One match term is *Journal Title*, and the other one is *ISSN*. Two matches reveal 15,911 *SJR* records in *Ulrichs*.
- The last match procedure is conducted between *Cabells* and *Ulrichs*. The match terms are still *Journal Title* and *ISSN*. In total 6,434 *Cabells* journals are discovered in *Ulrichs*.

After data consolidation, the final master dataset is based on the original Ulrichs set and contains 41,787

records, 10,354 of which are supplemented *GS* data, 9,583 of which are supplemented *JCR* data, 15,911 of which are supplemented *SJR* data, and 6,434 of which are supplemented *Cabells* data.

Results

There are 70 data fields in total derived from the five data sources. Some data fields are considered not relevant to our identified journal attributes, such as URLs from *JCR* and *Cabells* data tables. Some data fields deliver the same information, such as *JournalTitle* in *Ulrichs*, *JCR*, *SJR*, and *Cabells* tables. We extracted the meaningful data fields, and grouped them together.

Relevant attributes

Through the data-driven approach, we analysed all data fields with their values, categorised them into meaningful groups, and developed 13 relevant attributes that differentiate academic journals. The 13 relevant attributes are listed in Table 2.

Attributes	Relevant Data Field(s)	Source	Data Analysis Summary
Scholarly Impact	JCRTotalCites JCRImpactFactor JCR5YearImpactFactor JCRImmediacyIndex Rank SJR Hindex TotalRefs TotalCites3Years CitableDocs3years CitesDoc2Years RefPerDoc H5Index H5Median	JCR JCR JCR SJR SJR SJR SJR SJR SJR SJR SJR SJR SJ	JCRImpactFactor value from JCR, SJR score from SJR, and H5Index value from GS are measures of journal impact from three different organisations. SJR has the most journal coverage, therefore SJR score will be used as the fundamental <i>Scholarly Impact</i> value. Some records in <i>JCR</i> and <i>GS</i> can match <i>SJR</i> records via ISSN or JournalTitle. Because the correlation among <i>JCR</i> , <i>SJR</i> , and <i>GS</i> is high, <i>SJR</i> regression value can be imputed based on JCRImpactFactor value or H5Index. Using this method, SJR regression values for <i>JCR</i> and <i>GS</i> records with missing SJR values are imputed.
Subject category	Section SubjectCategory CategoryName	JCR JCR Cabells	<i>JCR</i> only provides two major categories: Science and Social Science. <i>Cabells</i> provides detailed subject category for academic journals, such as management, education, nursing, and so on. Another main reason is that <i>Cabells</i> has AcceptanceRate. Therefore <i>Cabells</i> category will be used for analysis.
Age	StartYear Year	Ulrichs JCR	If <i>Year</i> from <i>JCR</i> can not be captured, <i>StartYear</i> from <i>Ulrichs</i> will be used for age calculation.
Total articles	JCRArticles TotalDocs2014 TotalDocs3Years	JCR SJR SJR	<i>SJR</i> has more coverage than <i>JCR</i> . <i>TotalDocs3Years</i> from <i>SJR</i> will be used as <i>ArticleNumber</i> for analysis.
Distribution medium	Format PrintedISSN ElectronicISSN	Ulrichs Cabells Cabells	Distribution medium value can be captured from <i>Ulrichs</i> and <i>Cabells</i> . In this study, two types of distribution medium - <i>Print</i> and <i>Online</i> – will be studied. If <i>Format</i> is not available in <i>Ulrichs</i> , <i>Cabells</i> value will be considered.
Open access	OpenAccess	Ulrichs	Open access value is captured through Ulrichs.

Peer review	Refereed ReviewType	Ulrichs Cabells	Peer review value is captured through <i>Ulrichs</i> . Only peer-reviewed journals are considered for this study.		
Acceptance rate	AcceptanceRate	Cabells	Acceptance rate value is only available via <i>Cabells</i> .		
Pricing	Price	Ulrichs	Pricing value is available on <i>Ulrichs</i> . The subscription prices are in different currencies. The currencies require conversion to USD for analysis.		
Language	Language	Ulrichs, JCR	Language value is available on both <i>Ulrichs</i> and <i>JCR</i> . If <i>Ulrichs</i> does not indicate language, or <i>Ulrichs</i> and <i>JCR</i> have different languages, <i>JCR</i> language will be used.		
Country	Country	Ulrichs, JCR, SJR	Country value is available on <i>Ulrichs</i> , <i>JCR</i> , and <i>SJR</i> . Our data collection rule specifies <i>JCR</i> to be the primary data source for Country value. If thi journal does not have a matching JCR record, country information will be extract from SJR match, country information will be extracted from SJR record. If there is no JCR or <i>SJR</i> matching records, <i>Ulrichs</i> will be used to provide country information for this journal.		
Status	Status	Ulrichs	Status value is available on <i>Ulrichs</i> . There are two types of <i>Status</i> : <i>Active</i> and <i>Inactive</i> .		
IssueFrequency	ListFrequency IssuePerYear FrequencyOfIssue	Ulrichs JCR Cabells	Issue frequency value can be captured through Ulrichs, JCR, and Cabells. It is called ListFrequency in Ulrichs, IssuePerYear in JCR, and FreqncyOfIssue in Cabells. If JCR provides IssuePerYear, IssueFrequency will be the same as JCR record. If Ulrichs does not have IssueFrequency, Cabells information will be used.		

Table 2 List of relevant attributes defined via data collection and systematic review

Discussion

A comprehensive data collection over five major bibliographic databases was conducted. By considering the available data, a total of 13 attributes were derived from data collection and consolidation, including: *Scholarly Impact, Subject Category, Age, Total Articles, Distribution Medium, Open Access, Peer Review, Acceptance Rate, Pricing, Language, Country, Status, and Issue Frequency.* These attributes are described in detail in the section below. The discussion is based on the consolidated dataset obtained from the five data sources, and outlines the process for creating a single, complete dataset to characterise journals.

Scholarly Impact

Scholarly Impact represents the quality, performance, and prestige of the journal. Scholarly Impact is determined by ranking, which is based on impact factor, age, and number of articles (Jacsó, 2001; Kalaitzidakis, Mamuneas, & Stengos, 2003). Both JCR and GS do not provide rankings, consequently the 5-year impact factor in JCR (data field: JCR5YearImpactFactor) and 5-year h-index value (data field: H5Index) from GS are used to represent journal impact in JCR and GS. SJR score, using a different algorithm from impact or h-index, is introduced by SJR to represent the impact of journals. There are other types of impact related values, such as total cites, citable docs, reference per doc, and so on. These impact related values are already integrated in calculations of impact. Consequently, these impact related values will not be discussed or taken into consideration for further analysis. Thus, only three values, 5-year impact factor in JCR, 5-year h-index in GS, and SJR score in SJR, should be used to calculate the Scholarly Impact

of academic journals Although they are not the same measure, they perform the same purpose and are used to describe the *Scholarly Impact* for journals in their respective datasets. An individual SJR score in *SJR* cannot be compared to an individual 5-year *h*-index value in *GS*. However, from statistical point of view, *SJR score* can be used as a measure to predict the regression values for missing *SJR* data in *GS*.

Over the consolidated dataset, we conducted a quantitative relationship analysis among *Ulrichs*, *JCR*, *SJR*, and *GS*. From *Ulrichs*, 17,818 records have a match to a *JCR*, *SJR*, or *GS* record, while 23,909 *Ulrichs* records have no matches to any *JCR*, *SJR*, or *GS* records. *JCR* and *SJR* have 9396 matching records. *JCR* and *GS* have 6116 matching records. *SJR* and *GS* have 8585 matching records. There are 6067 matching records among *JCR*, *SJR*, and *GS*. The quantitative relationships among *Ulrichs*, *JCR*, *SJR*, and *GS* are presented in Figure 3.



Fig. 3 Quantitative relationship among Ulrichs, JCR, SJR, and GS journals

From Figure 3, *SJR* has better journal coverage than *JCR* and *GS*. Therefore, *SJR* score will be used as the fundamental *Scholarly Impact* value. The data range of *SJR* score is from 0.1 to 37.384. The correlation value between *JCR* and *SJR* is 0.87, and the correlation value between *SJR* and *GS* is 0.71 (Table 3).

	JCR impact	SJR impact	GS impact
JCR impact	1		
SJR impact	0.869	1	
GS impact	0.772	0.719	1

Table 3 Correlation among JCR, SJR, and GS

The correlation shows a strong relationship between *JCR* and *SJR*, and between *GS* and *SJR*. Analysis of variance (ANOVA) is used to test the difference between *SJR* and *JCR*, and between *SJR* and *GS*. The statistical results show *p* values for both scenarios are 0.00, which explains that any of the differences between the variables are statistically significant. Therefore, theoretically, a *Scholarly Impact* value for the *JCR* or *GS* journals without a *SJR* match can be calculated from regression values. Therefore, *JCR* and *GS* journals have their regression values calculated to replace the missing *SJR* impacts. The calculation is conducted in IBM SPSS statistic software package version 24.

Many studies have done on the two types of impact measures, primarily using regression analysis, although the results are contradictive. Via our statistical regression analysis, we have confirmed the research results from Schubert and Glänzel (2007), that there is a definite relationship between the impact factor and h-index, and this result can be extended from the relationship between impact factor and h-index, to the relationship among impact factors from JCR and h-index from GS, to SJR score from Scopus.

In the consolidated dataset, the *Scholarly Impact* value is in the format of number. Any journals outside the scope of *JCR*, *SJR*, and *GS* have the value of null. There are 16310 records in total having *Scholarly Impact* values, which leaves 39% of the final dataset as null.

Subject Category

Subject category represents the category that a journal's main topics belong to. The subject categories are normally developed by bibliographic services, such as ISI. Examples of categories are mathematics, chemistry, biology, etc. Subject category has a strong influence on journal ranking (Solari & Magri, 2000). Evidence shows that different journal categories have been identified as having different acceptance rates (Zuckerman & Merton, 1971). This suggests that *Subject Category* has a close relationship to acceptance rates. Among the five data sources, acceptance rates data can only be captured through *Cabells*. Therefore, we adopt the *CategoryName* value from *Cabells* to represent *Subject Category*. In *Cabells* there are 11 *Subject Categories* with the matching *Cabells* journal counts and *Ulrichs* journal counts, shown in Table4.

Subject Category	Cabells Count	<i>Cabells</i> Percentage	<i>Ulrichs</i> Count	Ulrichs Percentage
Accounting	624	7%	337	1%
Economics & Finance	1,539	18%	908	2%
Management	2,337	27%	1,318	3%
Marketing	785	9%	469	1%
Computer Science – Business Information Systems	1,716	20%	824	2%
Psychology & Psychiatry	1,872	22%	873	2%
Educational Curriculum & Methods	937	11%	542	1%
Educational Technology & Library Science	498	6%	287	1%
Educational Psychology & Administration	746	9%	460	1%
Health Administration	732	9%	321	1%
Nursing	555	6%	268	1%

Table 4 List of Cabells subject category with its Cabells journal count and Ulrichs journal count

Age

Age represents the number of years that an academic journal has been publishing for. It is one of the minor considerations to measure journals' ranking (Kalaitzidakis et al., 2003). The Age value can be retrieved from Ulrichs and JCR. JCR data (9583 records) only covers 22% of the final dataset (41787 records), but Ulrichs data, as the primary data source, cover 100% of the final dataset. Therefore StartYear value from Ulrichs is used to calculate the Age value for journals. However Ulrichs does not provide StartYear value for all journals. When StartYear value from Ulrichs for particular journal is not provided, Year value (equivalent to StartYear) from JCR will be used for calculation. The Age is calculated based on the year difference between year of 2016 and the start year of the journal. If a journal is Inactive, Age of the journal is calculated based on its EndYear and StartYear. If EndYear cannot be found, the StartYear value will be used as the EndYear value. The calculation equation for Active journals is Age = 2016 – StartYear; the calculation equation for Inactive journals is Age = EndYear – StartYear. Both Status and EndYear values are provided by Ulrichs. For the consolidated dataset, the Age value is stored in the format of numbers or null.

Total Articles

Total Articles stands for the total number of articles each journal publishes per year, irrespective of the number of issues. The number of articles is taken into consideration to measure journals' ranking (Kalaitzidakis et al., 2003). This value can be only captured from *JCR* and *SJR*. In *SJR*, *Total Articles* is represented by *TotalDocs3Years*, which is a count of the total number of articles in the previous 3 years. In *JCR*, *Total Articles* is represented by *JCRArticles*, which relates to the total number of articles in the current year. *SJR* (15911 records) covers more journals than *JCR* (9583 records). Consequently, *TotalDocs3Years* from *SJR* is used as the primary value for the total number of articles. *TotalDocs3Years* from *SJR* and *JCRArticles* from *JCR* have a very strong relationship as the correlation value between them is 0.883 and ANOVA shows the differences of these two variables are significant. Therefore, *JCR* journals without a *SJR* match can use a regression model to approximate its size. In total 16098 journals contain the value of article number. On average, 304 articles are published for each journal per year. Of these, 4,010 journals publish 304 articles or more each year, while 12088 journals publish less than 304 articles each year. In the consolidated dataset, the value of *Total Articles* is represented in a numeric format or null.

Distribution Medium

The Distribution Medium value can be captured from both Ulrichs and Cabells. Cabells cover only 10% of Ulrichs journals, therefore the Distribution Medium value will be extracted from Ulrichs. There are two major types of Distribution Medium: Print and Online. Empirical study shows that traditional subscription is still valid, while the number of online journal is growing rapidly (Gu & Blackmore, 2016). Based on Ulrichs data collection, there are 15.9% (6661 out of 41787) Print-Only journals, 18% (7529 out of 41787) Online-Only journals, 66% (27535 out of 41787) journals providing both Print and Online access, and 0.15% (59 out of 41787) journals without any Distribution Medium information. In the consolidated dataset, the Distribution Medium is appropriately represented as a categorical value, with Print-Only is represented by 1, Online-Only is represented by 2, while Print and Online is represented by 3. The journals without Distribution Medium information are noted with null for their distribution medium values.

Open Access

Open Access indicates whether a journal is accessible free, without subscription, for academics and interested readers. Open access is not only about public access rights or the general dissemination of knowledge, but also about increasing the impact and thereby the progress of research itself (Gargouri et al., 2010). The relationship between impact and open access has attracted attention from scholars. Some studies show that open access articles are cited significantly more than articles in the same journal and year that have not been made open access (Antelman, 2004; Harnad & Brody, 2004) . However open access does not make an unusable (un-citable) article usable; it simply makes a useful paper more visible. Some studies indicate that open access journals indexed in Web of Science and/or Scopus are approaching the same scholarly impact and quality as subscription journals (Björk & Solomon, 2012). Although these empirical studies are contradictive, it clearly demonstrations that *Open Access* is an important attribute of academic journals. Data collection for *Open Access* is straightforward. The relevant information is only available via *Ulrichs*. Based on the collected data, there are 6407 (15%) of *Ulrichs* journals providing *Open Access*, while majority 35380 (85%) of Ulrichs Journals do not provide *Open Access*. In the consolidated dataset, journals with open access are identified using a categorical value as 1, otherwise marked as 0, for the *Open Access* value.

Peer Review

Peer review is the 'gate watcher' for all journals so the best work can be selected to suit the journals' own characteristics (Relman, 1990). Therefore, only peer-reviewed journals are considered for this study. Non-peer-reviewed journals are excluded for data analysis. All of 41787 Ulrichs journals downloaded in Figure 1 are peer-reviewed.

Acceptance Rate

Acceptance rates of scholarly journals show substantial variation between disciplines, with the study of Zuckerman and Merton (1971) appearing as an important work in this area. They found disciplinary variation in rejection rates, with 20 to 40 percent in the physical science compared to 70 to 90 percent in the social sciences and humanities. The Acceptance Rate value can only be retrieved through Cabells. Only 6434 out of 41787 (15.4%) Ulrichs journals contain Acceptance Rate values. Some journals provide a single value for Acceptance Rate, and some journals provide a range for Acceptance Rate. Such as Accounting and Finance has 15-20% Acceptance Rate. For this study, the median value is adopted to calculate the

specific Acceptance Rate for the journals with a range of *Acceptance Rate*. Among 6434 *Ulrichs* journals, 416 journals have 20% acceptance rate, 299 journals have 50% acceptance rate, and 295 journals have 30% acceptance rate. These three groups of journals are top three types of *Acceptance Rate*, representing almost 25% of these 6434 *Ulrichs* journals. The *Acceptance Rate* value is recorded as a percentage in the consolidated dataset. Generally, different journal categories have different acceptance rates (Zuckerman & Merton, 1971). Therefore, the subject category will be used as a primary selection filter on the consolidated dataset for further analysis.

Pricing

Pricing represents the subscription cost of an academic journal. Subscription cost is one of the key measures for journal impact (Anderson, 1997). Among the five data sources identified in this study, the *Pricing* value can only be captured from *Ulrichs*. In *Ulrichs*, there are 62 types of currencies, including, for example, USD, AUD, CAD, EUR, GBP, NOK, and NZD. In the consolidated dataset, all *Pricing* values in *Ulrichs* are converted to USD using Google Finance Currency Converter (Google, 2016a) and XE Currency Exchange Website ("XE Currency Exchange Website," 2016) on the 15th of June 2016. Some journals provide annual subscription in multiple currencies. For example, subscription for *Building Acoustics* was 387.00 GBP or 716.00 USD per year (effective 2016) dated in Jan 2016. In this scenario, USD value will be noted as its subscription pricing in the consolidated dataset. Some journals have different charges based on geography. For example, *Journal of Family and Consumer Sciences* has five types of *Pricing*. They are:

- USD 175.00 subscription per year domestic to individuals (effective 2015)
- USD 200.00 subscription per year foreign to individuals (effective 2015)
- USD 350.00 subscription per year domestic to institutions (effective 2015)
- USD 375.00 subscription per year foreign to institutions
- USD 225.00 combined subscription per year domestic to individuals (print & online eds.) (effective 2015)

In this scenario, the highest annual subscription value is captured as the subscription pricing in the consolidated dataset. In the consolidated dataset, the annual subscription value is recorded as *Pricing* in USD value and shown in a number format.

Language

Language values are provided by two data sources: *Ulrichs* and *JCR*. Eighty percent (80%) of *Ulrichs* journals are published in English and twenty percent (20%) in Non-English. However, *JCR* journals are published in eighty percent (87%) in English, eight percent (8%) in multi-languages, and only five percent (5%) in non-English. Evidence has shown that *JCR* journals have higher impact than other journals (Gu & Blackmore, 2016). This comparison shows that higher-impact journals are published in English. This result validates the empirical study by Svensson, Rosenstreich, and Wooliscroft (2006) that scholars emulate US research approach and writing style to improve publishing success under the current status quo. In the consolidated dataset, there are only two types of languages: English and non-English. English is represented by a categorical value (1), while non-English is represented by digit 0 in the consolidated dataset.

Country

Country represents the geographical location where an academic journal administration is located. The *Country* values can be captured through three data sources: *Ulrichs, JCR*, and *SJR*. There are 152 distinct countries in *Ulrichs*, 82 distinct countries in *JCR*, and 102 distinct countries in *SJR*. All three data sources show that 11360 out of 41787 journals are published from United States (US) (27%) and 7201 out of 41787 journals are published from United States (US) (27%) and 7201 out of 41787 journals are published from United Kingdom (UK) (17%), with US propagating the highest number of academic journals. Forty-four percent (44%) of *Ulrichs* journals are oriented from US and UK, while fifty percent (50%) *SJR* journal and fifty-nine percent (59%) *JCR* journals are published within the US and UK. This result validates the conclusion from Zitt and Bassecoulard (1998), which most of academic research results are submitted to UK and US journals. In the consolidated dataset, the full name of *Country* is recorded in the consolidated file as *Country* value.

Status

Status is an indicator of the current status of an academic journal. There are quite a few statuses of journals: *Active, Announced Never Published, Ceased, Forthcoming, Merged / Incorporated, Researched / Unresolved,* and *Suspended.* The *Status* values of journals are only available from *Ulrichs.* In *Ulrichs,* the

definition of Active means that the periodical title is currently active in publishing. The Status of Announced Never Published, Ceased, Forthcoming, Merged / Incorporated, Researched / Unresolved, or Suspended stands for journals paused or stopped publishing. Therefore, in this research, there are two types of Status are considered: Active and Inactive. In Ulrichs 90% journals are Active and 10% journals are Inactive. In the consolidated dataset, Active journals are represented using a categorical value in 1 while Inactive journals are represented in 0.

Issue Frequency

Issue Frequency represents how often an academic journal publishes its accepted research articles. The value of *Issue Frequency* can be captured in *Ulrichs* and *JCR*. In *Ulrichs*, the *Issue Frequency* is grouped in specific formats. There are 14 types of *Issue Frequency*: Quarterly (29.15%), Semi-annually (18.35%), Other (15.30%), Bi-monthly (11.42%), Monthly (9.12%), Annual (8.39%), Irregular (4.21%), Continuously (3.19%), Semi-monthly (0.47%), Weekly (0.19%), Biennial (0.12%), Fortnightly (0.04%), Daily (0.03%), and Triennial (0.01%). The most common *Issue Frequency* is *Quarterly* which is adopted by 29% *Ulrichs* journals. The second most common *Issue Frequency* is *Semi-annually* which is adopted by 18% *Ulrichs* journals. In *JCR*, the *Issue Frequency* is rated in numbers, ranked from 0 to 60. The top 3 types of *Issue Frequency* are 4 (31%), 6 (23%), and 12 (20%) times per annum. In *JCR*, 3661 out of 11813 *JCR* journals issue 4 times per year, which means the most common issue frequency in *JCR* is the same as *Ulrichs*.

In *Ulrichs*, three *Issue Frequency* types (Irregular, Continuously, and Other) cannot be converted to number format. These three types are analysed statistically in SPSS in Table 5.

Туре	Count	Average	Range	Standard Deviation	Mode	Median
Irregular	179	1	12	2.53	0	0
Other	2437	10	60	7.98	8	8
Continuously	89	3	12	4.49	0	0

Table 5 Statistic descriptive of three special Issue Frequency Types in Ulrichs

The *Ulrichs* journals with three *Issue Frequency* types are matched with *JCR*. Average values are adopted as the number replacement for these three *Issue Frequency* types. Therefore, in the consolidated dataset, the *Issue Frequency* is mainly abstracted from *Ulrichs* and *JCR*. *Issue Frequency* is stored in the data file in the format of number.

Conclusion

To develop attributes for a consolidated dataset of academic journals in digital age, we downloaded and consolidated data from the five major bibliographic data sources of *Ulrichs*, *JCR*, *SJR*, *GS*, and *Cabells*. The data analysis approach collated 70 data fields and resulted in a set of attributes: *Scholarly Impact*, *Subject Category*, *Age*, *Distribution Medium*, *Open Access*, *Peer Review*, *Acceptance Rate*, *Pricing*, *Language*, *Country*, *Status*, *and Issue Frequency*, which are used for characterising and understanding the dimensions of academic journals. Through this data work, a consolidated dataset from the five major bibliographic data sources is created for the purpose of discussion, and to form the basis for future analysis. Each attributed was discussed, and the rationale for how they should be presented in the consolidated dataset is provided.

This study identifies relevant attributes for characterising journals from the existing set of variable available across all existing bibliographic databases. It does not, however, extend to consider what additional variables, beyond those currently captured by these bibliographic databases, might be of use. As such, the work presented in this study provides a foundation from which to build and expand the collection of data on academic journals. It also does not extend to variables that could be calculated from existing data. For example, it might be possible to infer the degree of disciplinarity from the topics covered by a particular journal, but as this is not an existing attribute captured explicitly in bibliographic databases, it is not considered in this work. This, however, does not suggest that these additional variables might not be useful.

It is well recognised that there are inherent difficulties in covering all journal titles (Hicks, 1999). Only five data sources were adopted for this study, although *Ulrichs* is being regarded as the most complete journal

library (Jinha, 2010; Tenopir & King, 2009b). Therefore, the journal titles we collected through data work may be limited. Additionally, we only considered peer-reviewed journals for this study. In reality, some non-peer-reviewed journals are popular for scholars and scholar-practitioners, and contain valuable knowledge. For example, many information sources for library information system (LIS) practitioners are excluded from this study. However, restricting this work to peer-reviewed publications creates a consistent data source across disciplines, particularly where non-peer review journals are not common repositories for scholarly output.

Despite these limitations, the analysis presented in this paper makes a valuable contribution to research on academic journals. The consolidated dataset produced identifies some key attributes that are not well captured in existing bibliographic sources, such as, *Status* and *Issue Frequency*, and thus highlights where improved data collection would be useful due to the limited system literature collection approach. Additionally, the lack of a formal data dictionary for key attributes to characterise and describe journals leads to inconsistencies across data collection agencies and organisations. This is also relevant when comparing research on academic journals, where differences may occur in how attributes are considered. This work goes some way to defining this data dictionary, and represents an important first step in this direction.

Lastly, the analysis presented here establishes a data framework for future research in the evaluation and analysis of the dimensions of academic journal publishing. This characterisation of academic journals and the consolidated dataset open up plenty of future research opportunities. One key area for future work is aimed at the development of an academic journal typology, which can inform the development and validation of system models, such as that used in agent-based modelling and simulation approaches. This future work would consider similarities or groupings in variables from the consolidated dataset, and analyse patterns in these attributes to allow the differentiation of journal subgroups or types.

Another area for future work is aimed at quantitative studies of the characteristics of academic journals, which involves investigating the correlation among the characteristics and the quantitative relationships between different journal types. Extended future work could focus on changes in the eight identified research streams that occur due to real-world developments and pressures, changes in authorship order within and across disciplines, and changes in research streams by geographical areas, and so on.

Lastly, results of quantitative studies can also be used for validating system models such as agent-based models and simulations, and are also valuable for more general evaluation of the academic publishing system over time. Academic journals in the publishing system are under undergoing dramatic changes in the current digital age. Journal types based on the relevant attributes will assist us to better understand the development pattern of academic journals in publishing system, and explore the impact of patterns over time.

References

- Adler, N. J., & Harzing, A.-W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. Academy of Management Learning & Education, 8(1), 72-95.
- Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3(4), 273-289. doi:<u>http://dx.doi.org/10.1016/j.joi.2009.04.001</u>
- Anderson, P. (1997). 'Gatekeepers' and the quality of the journal literature: Findings from a survey of journal editors into the issue of alleged excessive publication in scholarly and scientific journals. *Serials Review*, 23(2), 45-57.
- Antelman, K. (2004). Do open-access articles have a greater research impact? *College & research libraries*, 65(5), 372-382.
- Axelrod, R. M. (1997). The complexity of cooperation: Agent-based models of competition and collaboration: Princeton University Press.
- Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. Scientometrics, 74(2), 257-271.
- Björk, B.-C., & Hedlund, T. (2004). A formalised model of the scientific publication process. Online Information Review, 28(1), 8-21.
- Björk, B.-C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. BMC medicine, 10(1), 73.
- Blake, G., & Bly, R. W. (1993). The elements of technical writing: Longman.
- Bonabeau, E. (2002). Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. Proceedings of the National Academy of Sciences of the United States of America, 99(Suppl 3), 7280-7287.
- Colledge, L., de Moya-Anegón, F., Guerrero-Bote, V., López-Illescas, C., El Aisati, M., & Moed, H. F. (2010). SJR and SNIP: two new journal metrics in Elsevier's Scopus. *Serials*, 23(3), 215-221.
- Delgado-López-Cózar, E., & Cabezas-Clavijo, Á. (2012). Google Scholar Metrics: an unreliable tool for assessing scientific journals. *El profesional de la información*, 21(4), 419-427.
- Elsevier. (2016). Scopus. from Elsevier http://www.scopus.com/
- Garfield, E. (2006). The history and meaning of the journal impact factor. JAMA: the journal of the American Medical Association, 295(1), 90-93.
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PloS one*, 5(10), e13636.
- Google. (2016a). Google Finance Currency Converter. Retrieved from https://www.google.com/finance/converter
- Google. (2016b). Google Scholar Metrics. Retrieved January 2016 https://scholar.google.com/intl/en/scholar/metrics.html
- Gu, X., Blackmore, K., Cornforth, D., & Nesbitt, K. (2015). Modelling Academics as Agents: An Implementation of an Agent-Based Strategic Publication Model. *Journal of Artificial Societies* and Social Simulation.
- Gu, X., & Blackmore, K. L. (2016). Recent trends in academic journal growth. Scientometrics, 1-24.
- Guerrero, R., & Piqueras, M. (2010). Open access. A turning point in scientific publication. *International Microbiology*, 7(3), 157-161.
- Guz, A., & Rushchitsky, J. (2009). Scopus: A system for the evaluation of scientific journals. *International Applied Mechanics*, 45(4), 351-362.
- Hargens, L. L. (1988). Scholarly consensus and journal rejection rates. American Sociological Review, 139-151.

- Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-lib Magazine*, *10*(6).
- Hazelkorn, E. (2013). *How Rankings are Reshaping Higher Education?* Paper presented at the IAU 13th General Conference, Utrecht, The Netherlands.
- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44(2), 193-215.
- Jacsó, P. (2001). A deficiency in the algorithm for calculating the impact factor of scholarly journals: the journal impact factor. *Cortex*, 37(4), 590-594.
- Jinha, A. E. (2010). Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing*, 23(3), 258-263.
- Kalaitzidakis, P., Mamuneas, T. P., & Stengos, T. (2003). Rankings of academic journals and institutions in economics. *Journal of the European Economic Association*, 1(6), 1346-1366.
- Kapelianis, D., & Cabell, D. W. (1999). Cabell's directory of publishing opportunities in management and marketing. 440-442.
- Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575-603.
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social studies of science*, *35*(5), 673-702.
- Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*, 57(5), 601-613.
- Mabe, M., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51(1), 147-162.
- Macal, C. M., & North, M. J. (2009). Agent-Based Modeling and Simulation. Paper presented at the 2009 Winter Simulation Conference, Austin, TX, USA.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal* of the American Society for Information Science, 40(5), 342.
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.
- Mölders, M., Fink, R. D., & Weyer, J. (2011). Modeling Scientists as Agents. How Scientists Cope with the Challenges of the New Public Management of Science. *Journal of Artificial Societies and Social Simulation*, 14(4), 6.
- Nisonger, T. E. (1998). Management of Serials in Libraries: ERIC.
- Relman, A. S. (1990). Peer review in scientific journals--what good is it? Western Journal of Medicine, 153(5), 520.
- Rice, B. A., & Stankus, T. (1983). Publication quality indicators for tenure or promotion decisions: what can the librarian ethically report? *College and Research Libraries*, 44(2), 173-178.
- Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, 1(3), 179-184.
- Solari, A., & Magri, M.-H. (2000). A new approach to the SCI Journal Citation Reports, a system for evaluating scientific journals. *Scientometrics*, 47(3), 605-625.
- Solomon, D. J. (2007). The role of peer review for scholarly journals in the information age. *Journal of Electronic Publishing*, 10(1).
- Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science*, *16*(2), 180-200.
- Statzner, B., & Resh, V. H. (2010). Negative changes in the scientific publication process in ecology: potential causes and consequences. *Freshwater Biology*, 55(12), 2639-2653.

- Svensson, G., Rosenstreich, D., & Wooliscroft, B. (2006). How international are the top academic journals? The case of marketing. *European Business Review*, 18(6), 422-436.
- Tenopir, C., & King, D. W. (2009a). The growth of journals publishing. *The future of the academic journal, 1*(84334), 416.

Tenopir, C., & King, D. W. (2009b). The growth of journals publishing: Chandos Oxford.

Thomson Reuters. (2014). Cabell's International. Retrieved from http://cabells.com/index.aspx

- Thomson Reuters. (2016). Journal Citation Reports®. Retrieved from <u>http://thomsonreuters.com/journal-citation-reports/</u>
- Tuttle, M. (1987). The serials directory: An international reference book. Serials Review, 13(2), 5-14.

XE Currency Exchange Website. (2016). Retrieved from http://www.xe.com

- Zitt, M., & Bassecoulard, E. (1998). Internationalization of scientific journals: a measurement based on publication and citation scope. *Scientometrics*, 41(1), 255-271.
- Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, 9(1), 66-100.