

# Not Scotch, but Rum: The Scope and Diffusion of the Scottish Presence in the Published Record

Brian Lavoie

Research Scientist  
OCLC Research



Not Scotch, but Rum: The Scope and Diffusion of the Scottish Presence in the Published Record

Brian Lavoie, for OCLC Research

© 2013 OCLC Online Computer Library Center, Inc.

This work is licensed under a Creative Commons Attribution 3.0 Unported License.

<http://creativecommons.org/licenses/by/3.0/>



September 2013

OCLC Research

Dublin, Ohio 43017 USA

[www.oclc.org](http://www.oclc.org)

ISBN: 1-55653-463-9 (978-1-55653-463-8)

OCLC (WorldCat): 856661737

Please direct correspondence to:

Brian Lavoie

Research Scientist

[lavoie@oclc.org](mailto:lavoie@oclc.org)

Suggested citation:

Brian Lavoie. 2013. *Not Scotch, but Rum: The Scope and Diffusion of the Scottish Presence in the Published Record*. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/content/dam/research/publications/library/2013/2013-07.pdf>.

## Contents

Acknowledgements.....	6
Introduction .....	7
A National Presence in the Published Record .....	8
Identifying the Scottish National Presence in the Published Record.....	9
Materials Published in Scotland.....	11
Materials Published by Scottish People.....	12
Materials Published About Scotland.....	14
The Scottish National Presence in the Published Record .....	15
Global Diffusion of the Scottish National Presence in the Published Record .....	21
International Patterns of Diffusion of the Scottish National Presence.....	21
Core Works in the Scottish National Presence.....	24
Conclusion.....	34

## Tables

Table 1. Overlap across components of the Scottish national presence* .....	16
Table 2. Five most frequently-occurring languages other than English .....	18
Table 3. Worldwide holdings of the Scottish national presence.....	21
Table 4. Largest concentrations of materials in the Scottish nat'l presence, worldwide.....	23
Table 5. Top 10 largest work clusters in the Scottish national presence .....	25
Table 6. Top 10 works in Scottish national presence most widely held in library collections. 26	
Table 7. Works in Scottish nat'l presence most widely held in lib. collections (50th-59th)....	27
Table 8. Works in Scottish nat'l presence most widely held in lib. collections (90th-99th)....	28
Table 9. Most widely held Scottish Enlightenment works (1740-1800) .....	29
Table 10. Most widely held works in the Scottish national presence, published after 1950* ..	30
Table 11. Most widely held works by Scottish authors, published after 1950* .....	31
Table 12. Most widely held works in Scottish national presence: Scotland, US, and Australia .....	32
Table 13. Works in Scottish national presence with largest work clusters in Hathi Trust collection.....	33

## Figures

Figure 1. Two distinct publications of the same work by Robert Louis Stevenson.....	10
Figure 2. Publications in the Scottish National Presence .....	16
Figure 3. Works in the Scottish national presence.....	17
Figure 4: Materials published in Scottish-Gaelic and Scots.....	19
Figure 5: Distribution of publication dates in the Scottish national presence (percent) .....	20
Figure 6: Global diffusion of Scottish national presence (holdings).....	22

## Acknowledgements

The author thanks Caroline Brazier, Andrew Green, John MacColl, and OCLC colleagues Merrilee Proffitt, Titia van der Werf, and Lorcan Dempsey for reading and commenting on earlier drafts of this report. Special thanks to JD Shipengrover for preparing the graphics shown in figures 2 and 6; also to Lorcan Dempsey for recognizing the potential for using library data to analyze broader cultural topics, and encouraging this research.

## Introduction

Big data sets and powerful computing capacity have transformed scholarly inquiry across many disciplines. While the impact of data-intensive research methodologies is perhaps most distinct in the natural and social sciences, the humanities have also benefited from these new analytical tools. For example, a new field of study—“culturomics”—employs computational methods to identify interesting cultural patterns in digitized texts (Wikipedia 2013). The literary scholar Franco Moretti coined the term “distant reading” to describe literature studies based not on “close” (i.e., human) reading, but computational analysis of massive aggregations of digitized works (Schulz 2011).

While full-text data is necessary to study topics such as lexicographical patterns or the distinctive features of certain literary genres, other types of analysis can be performed using only bibliographic descriptions of a corpus of works—i.e., metadata that includes the identity of the author, publication information, subject classifications, and so on. This report uses the millions of bibliographic descriptions in the WorldCat database to identify and characterize the Scottish presence in the published record.

The sum total of published knowledge is, for the most part, contained within the sixteen billion volumes residing in the collections of a million libraries worldwide (OCLC 2003, 5). The aggregation of global library holdings can therefore serve as a proxy for the published record. In practice, our view of the published record is incomplete: no single aggregation of data completely describes it. However, the WorldCat database can serve as a rough approximation for the global published record.<sup>1</sup> The database contains bibliographic descriptions of more than 255 million distinct publications representing nearly 1.8 billion library holdings worldwide.<sup>2</sup> While WorldCat includes materials of all types, books are particularly well-represented; moreover, WorldCat coverage tends to be most complete *vis-à-vis* North American library collections, and is only partial in other parts of the world. Nevertheless, WorldCat is the best representation of the global library resource available, and therefore the closest approximation of the published record.

This report uses Scotland as a case study to illustrate the concept of a national presence in the published record. The concept of a national presence is defined and operationalized in WorldCat data through a methodology that emphasizes machine processing with minimal manual intervention. The Scottish national presence in the published record is extracted from the global library resource represented in WorldCat, and characterized along a variety of

dimensions. In addition, library holdings data is used to track the international diffusion and impact of the Scottish national presence. Analysis of Scotland's impact on the published record illustrates how massive aggregations of bibliographic data can be used to conduct research on cultural patterns and trends. National libraries, as well as other memory institutions, would benefit from a view of national contributions to scholarship and culture reflected in the published record, given their mission to collect, make available, and preserve their country's cultural and intellectual heritage. In addition, scholars may find the corpus of materials comprising a national presence to be a fruitful area for research.

## A National Presence in the Published Record

The first task in identifying a national presence in the published record is to define it. An obvious starting point is the output of the national publishing industry, which, in a sense, is the most direct and easily measurable contribution to the published record. Another important aspect of the national presence is the intellectual or creative works produced by the people of a particular country; these can be published inside or outside the home country. For example, the Scottish-born author Ian Rankin is considered a key figure in the "Tartan Noir" genre of Scottish crime fiction; his best-selling Inspector Rebus series (set in Edinburgh) is published by the London-based Orion Publishing Group. Scotland's presence in the published record would seem incomplete without Rankin's work.

The distinction between works published in a country, and works published by the people of that same country, finds a parallel in the economic concepts of gross domestic product (GDP) and gross national product (GNP). GDP is the value of all goods and services produced within a country. GNP measures the value of goods and services produced by the nationals of a particular country, regardless of where the production occurred. In practice, the difference between GDP and GNP is usually small, but not trivial.<sup>3</sup> In contrast, the difference between publishing "GDP" (published in the country) and "GNP" (published by the people of a country) may be significant, especially in countries with a domestic publishing industry that is small or projects a modest global profile. Consequently, the definition of a national presence should include materials published in a particular country, as well as materials published by the people of a particular country.

There is yet another element that should be included in the definition of a national presence: materials about a particular country, regardless of their origin. One metric of a country's impact on the published record is the intensity with which materials about the country in question are published worldwide. Returning to the Scotland example, the works produced by the luminaries of the Scottish Enlightenment—e.g., Adam Smith, David Hume, James Hutton, Robert Burns—would surely all be considered part of the Scottish national presence. But what about the corpus of materials written about the Scottish Enlightenment? Such materials form



what might be viewed as second layer of national presence that forms around the direct contributions of a nation's intellectual and creative production (i.e., materials published in the country, or by the country's people). Materials published about a country reflect the global influence exerted by its ideas, institutions, history, and culture.

This interpretation of a national presence in the published record aligns well with the stated missions of national libraries, which are usually the stewards of their nation's cultural and scholarly heritage. For example, the National Library of Australia's Service Charter states that the "Library's role. . . is to ensure that documentary resources of national significance relating to Australia and the Australian people, as well as significant non-Australian library materials, are collected, preserved and made accessible. . . ." <sup>4</sup> Similarly, the mission of the National Library of Ireland "is to collect, preserve, promote and make accessible the documentary and intellectual record of the life of Ireland. . . ." <sup>5</sup> while the Swiss National Library focuses on "Helvetica", including "Swiss publications and foreign publications dealing with Switzerland and its inhabitants as well as publications by Swiss authors that have been published abroad, including translations." <sup>6</sup> The task of the National Library of Poland "is to acquire, store and permanently archive the intellectual output of Poles, whether the works of citizens living on Polish soil, the most important foreign works, or publications related to Poland and published abroad." <sup>7</sup>

In summary, a national presence in the published record is defined to include materials *published in the country, published by the country's nationals, and published about the country*. A variety of issues emerge in drawing boundaries around each of these categories, and operationalizing them via data available in a bibliographic record. These issues are discussed in the next section.

## Identifying the Scottish National Presence in the Published Record

This paper uses Scotland as a case study for illustrating a national presence in the published record. The purpose of the case study is to demonstrate that:

- the concept of a national presence can be operationalized in the form of a methodology operating on bibliographic data;
- the methodology can be designed such that it can be re-purposed without significant modification to almost any country, with only minimal manual intervention.

Ease of implementation is essential: a methodology that must be hand-crafted to fit the circumstances of a particular nation, and includes pain-staking manual review would likely be too cumbersome to implement. But it is not without trade-offs. The higher accuracy that comes with customization and manual review must be weighed against the ease—but lower accuracy—of machine processing. Of course, no methodology—not even one which relies heavily on manual review—can exhaustively identify a national presence. An endless series of refinements can be devised to reduce the incidence of “Type I” and “Type II” errors—that is, materials that are incorrectly accepted, or incorrectly rejected. For our purposes, the goal is to construct a methodology that produces reasonably good results in the absence of significant customization and manual intervention.

The findings reported in this paper are based on WorldCat bibliographic and holdings data from January 2012. Other data sources employed in the analysis are cited later in the study.

### *Some definitions*

The following terminology is helpful in understanding the methodology and analysis described in this report:

- *Work*: a distinct intellectual creation. For example, *Treasure Island* is a work by Robert Louis Stevenson.
- *Publication*: a distinct edition or imprint of a work. For example, the work *Treasure Island* has appeared as many different publications, two of which are shown below (These would be counted as two distinct publications in the analysis in this report).



**Figure 1. Two distinct publications of the same work by Robert Louis Stevenson**

- *Holding*: an indicator that a particular institution (e.g., a library) holds at least one copy of a particular publication in its collection. Note that a holding says nothing about the number of physical copies owned by the institution, other than at least one copy is available. For example, according to their catalog, the Dallas Public Library owns three copies of the Penguin publication of *Treasure Island*. All three copies are represented in WorldCat by a single holding associated with the Dallas Public Library.<sup>8</sup>

## Materials Published in Scotland

The first component of the Scottish national presence includes materials published in Scotland. This component is largely straightforward to operationalize in bibliographic data, since well-defined fields and codes exist to record country of publication in a MARC-format<sup>9</sup> bibliographic record; this in turn facilitates machine processing. A value of “stk” in the 008 field (bytes 15-17) indicates the material described in the record was published in Scotland. A second criterion is to check all instances of the 044 field/subfield-a, which is used to record additional countries of publication when the material is published simultaneously in multiple places. If “stk” appeared in the 008 field (bytes 15-17), or any instance of the 044 field/subfield-a, the record was flagged as describing something published in Scotland.

One complication specific to the Scottish case is that some materials that were published in Scotland are cataloged using the more general designation of the United Kingdom, which includes Scotland as well as England, Wales, and Northern Ireland. To address this, the same procedure for identifying materials explicitly cataloged as published in Scotland was used to identify materials cataloged as published in the UK (the relevant code is “xxk”). For these materials, the 260 field/subfield-a (place of publication) was parsed and analyzed. Typically, the information in this subfield denotes the city in which the material was published. Machine processing of this information is complicated by the fact that values are recorded as “free text”, rather than with standard codes. Consequently, a variety of conventions, abbreviations, and spellings are encountered. To overcome this problem, a matching algorithm was developed which compared normalized words or groups of words in the field with a table of the top 50 Scottish cities (by population)<sup>10</sup>. If a match was identified, the record was flagged as describing material published in Scotland.

The procedure described above was implemented in an algorithm which was run against the WorldCat database. This yielded 966,234 materials explicitly cataloged as published in Scotland, and a further 23,628 cataloged as published in the UK, but actually published in Scotland, for a total of **989,862 materials published in Scotland**.

## Materials Published by Scottish People

The second component of the Scottish national presence includes materials authored or created by Scottish people.<sup>11</sup> Compiling a list of Scottish authors/creators can proceed in several ways. One is to use an existing list. Investigation of this option made clear that there was no single, easily obtainable list of Scottish authors/creators that could be reasonably construed as comprehensive. Some lists tend to focus on historical authors; some on contemporary ones. Some include only writers. All are to a greater or lesser degree incomplete. Of course, comprehensiveness is likely to be unachievable in any case, but the lists examined were either too incomplete to be of use, or focused on a particular category of Scottish authors/creators to the exclusion of others.<sup>12</sup> Another approach is to compile a list by hand, drawing on existing lists and other sources. While this may result in a more comprehensive list of names, it would also involve a labor-intensive, time-consuming process, which would have to be repeated for every country to which the methodology is applied.

For the purposes of this report, a methodology was developed for identifying Scottish authors/creators that draws on the publicly available data sets provided through DBpedia, an initiative aimed at transforming the information in Wikipedia into structured data sets.<sup>13</sup> The DBpedia data sets offer several advantages that recommend them for use in building a list of authors/creators associated with a particular country. The data sets are machine processable, which reduces the need for manual analysis; moreover, the fact that the data is structured enhances the scope for re-purposing the processing algorithms for other countries besides Scotland. The data sets offer a reasonable approximation of comprehensiveness, in that most Scottish authors and creators of at least modest visibility are likely to be represented in Wikipedia. Finally, the “crowd-sourced” nature of Wikipedia content suggests a natural consensus for situations where a person’s nationality is uncertain or in dispute.

The DBpedia file<sup>14</sup> containing structured data about all persons with an entry in the English-language version<sup>15</sup> of Wikipedia was processed to identify all records with a “birthplace” field containing the string “Scotland”. The file was also checked for any birthplace field populated with a string ending in one of the top 50 Scottish locales<sup>16</sup>, or one of the 32 Scottish council areas. This procedure identified 6,097 distinct names of Scottish persons. In addition, the DBpedia file containing the short abstracts for every entry in the English-language version of Wikipedia was processed to identify all entries that contained the word “Scottish”.<sup>17</sup> This produced a list of 23,788 entries.

Each entry in the persons and abstracts files contains a unique identifier in the form of a link—for example, [http://dbpedia.org/resource/Adam\\_Smith](http://dbpedia.org/resource/Adam_Smith). Entries in the two files will share the same link identifier if they pertain to the same Wikipedia page. The links for the entries extracted from the abstracts file were compared to the links from all entries in the

persons file; only those links from the abstracts file that identify a person (by virtue of a matching link in the persons file) were retained. This reduced the list of entries from the abstracts file to 11,075.

The individuals identified as Scottish from the abstracts file fall into one of three categories:

- Those who have already been identified as Scottish from the persons file via birthplace data.
- Those who have no birthplace data in the persons file. These names were retained in the list as being Scottish, by virtue of the abstracts data alone.
- Those who have non-Scottish birthplace data in the persons file. These individuals were discarded from the list.

The lists of individuals identified as Scottish from the persons and abstracts files were then merged, yielding a final list of 11,604 distinct Scottish people in DBPedia. To validate the accuracy of this list, a 1 percent random sample of names was extracted from the list and manually checked; 91 percent were indeed Scottish—i.e., were born in Scotland. Of the remaining 9 percent, some fell afoul of the extremely narrow test of Scottishness. These individuals were identified solely through the abstracts file; manually verifying their birthplace from other data sources revealed that they were born outside of Scotland to Scottish parents. If these persons are also considered Scottish, the accuracy rate of the sample rises to 93 percent.

This methodology for identifying the Scottish people in Wikipedia has the advantage of being conducted entirely through machine processing, with a result that can make a reasonable claim to comprehensiveness. However, it suffers from several drawbacks. As noted, it relies on a narrow definition of Scottishness. In addition, it only includes people who have merited an entry in Wikipedia. Checking the sample can only assess the incidence of names that were included that should not be. It says nothing about names that should be included but are not.

At this stage, we only have a list of Scottish people, not Scottish people who are also authors/creators. To impose the last filter, the names in the list were compared to the data underpinning the WorldCat Identities service.<sup>18</sup> WorldCat Identities provides a summary page for every author or creator whose work is represented in WorldCat, including information about their body of work as it is represented in WorldCat. Matching the names from the list of Scottish people derived from DBpedia to the individuals represented in WorldCat Identities produced a list of all publications in WorldCat associated with a Scottish author/creator.<sup>19</sup> This list of publications represents the second component of the Scottish national presence: **642,427 publications by Scottish people.**

## Materials Published About Scotland

The third and final component of the Scottish national presence is material about Scotland. Identifying materials in WorldCat that are about Scotland is challenging, in that the concept of being “about” something is itself not well defined. It is easy to stray into definitions that are both too narrow and too expansive. For example, being about Scotland is more than just material written specifically about the country of Scotland, such as travel guides. The 2010 biography *Adam Smith: An Enlightened Life*; the acclaimed children’s fiction book *Always Room for One More* (set in Scotland); the personal journals of Scottish missionary David Livingstone: all are, at some level, about Scotland. But a line must be drawn somewhere. A book about the Associate Reformed Presbyterian Church in the United States is not really about Scotland, even though this denomination is of Scottish origin. And a picture of a Scots pine is not about Scotland!

Several approaches can be taken to identify a cohort of materials about Scotland. In choosing a strategy, the key trade-off is between precision and ease of implementation—or put another way, between a labor-intensive process of constructing a precise identification of materials about Scotland, and a largely automated process that requires less effort to implement, but at the price of less precision in results. For this study, an automated approach was chosen, in keeping with the goal to construct a methodology that minimizes manual intervention. However, some manual review was still required, as noted below.

First, the WorldCat database was scanned to identify all records that contained at least one FAST subject heading<sup>20</sup> that included a direct reference to Scotland: “Scotland”, “Scottish”, “Scot”, along with several other variations. All eight FAST subject facets were analyzed: topical, geographic, chronological, personal names, corporate names, events, form, and genre. These headings constituted a core set of Scotland-related subject headings. Any record with a Geographic Area Code of “e-uk-st” in the 043 field/subfield-a was also flagged. This process yielded 395,508 records; all of these were deemed to describe materials about Scotland.

The next step was to extract all of the FAST headings that co-occurred with the core Scotland-related headings. These were sorted by facet, and then ranked by frequency of occurrence (i.e., number of records in which each heading co-occurred with a core Scotland-related heading). All of the headings in each facet that co-occurred ten or more times with one of the core Scotland-related headings were reviewed to assess whether they were also Scotland-related in their own right. For example, the co-occurring FAST heading “Burns, Robert” describes something about Scotland (a Scottish poet), as does “Covenanters” (a Scottish Presbyterian movement). Headings were discarded if they did not reference something about Scotland (e.g., “ballads, English”) or were too general (e.g., “universities and colleges”).<sup>21</sup>

Given the list of subject headings that survived the review, a second scan of WorldCat was performed to identify all records that contained any of these headings, regardless of whether or not they co-occurred with one of the core Scotland-related headings. The resulting record set was then merged with the first record set produced by the core Scotland-related headings. Finally, an additional set of records were added based on WorldCat Identities data. This latter group of records was a by-product of the process of identifying materials published by Scottish authors/creators (described in the previous section). In addition to mapping the publications described in WorldCat records to the individuals responsible for authoring or creating them, WorldCat Identities also maps publications in WorldCat to identities they are about. Given the list of Scottish people described in the previous section, a set of records was identified describing publications that included as a subject one or more of the individuals on the list. Many of these publications had no co-occurring FAST heading that directly referenced Scotland or a Scotland-related subject (other than the identity of the Scottish person whom the material was about), and so were not identified through the FAST headings analysis. Taking all of these records together and eliminating duplicates yielded **515,146 publications that are about Scotland.**

## The Scottish National Presence in the Published Record

Combining the materials published in Scotland, by Scottish people, or about Scotland—and then removing duplicates—yields a Scottish national presence in the published record of 1.8 million distinct publications (figure 2). To lend a sense of proportion to this number, note that the size of the Scottish national presence exceeds that of the library collections of each of the four ancient Scottish universities.<sup>22</sup> Or to put it another way, imagine a medium-sized research library filled with nothing but materials published in Scotland, authored or created by Scottish people, or about Scotland. As these examples suggest, the Scottish national presence is a resource of significant proportions.

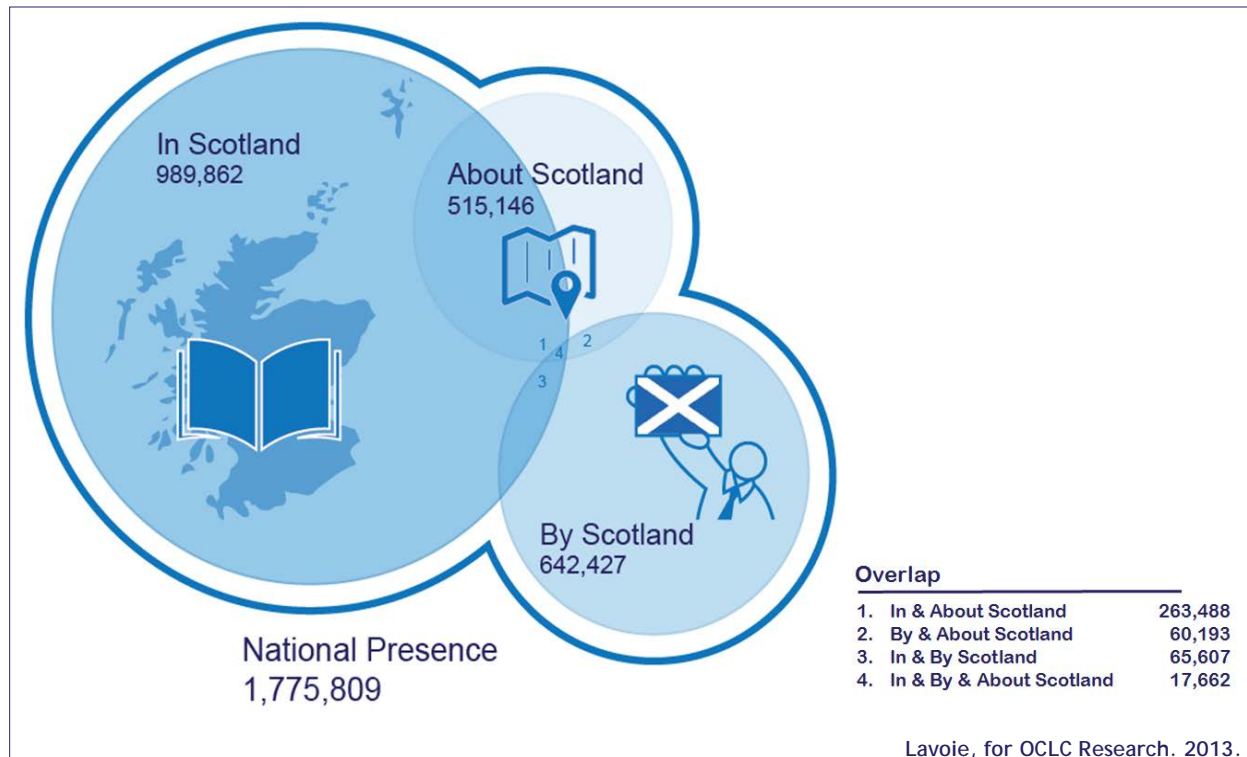


Figure 2. The Scottish Presence in the published record

Table 1 reports the bi-lateral overlap across the three components of the Scottish national presence.

Table 1. Overlap across components of the Scottish national presence\*

	Published In	Published By	Published About
Published In	1.00	0.07	0.27
Published By	0.10	1.00	0.09
Published About	0.51	0.12	1.00

\*Each result represents the percentage of materials identified by row heading that also are included in materials identified by column heading

Only 7 percent of the materials published in Scotland are created or authored by Scottish people. This suggests two possible interpretations: first, that most Scottish authors publish domestically, but are heavily exceeded in number by non-Scottish authors who choose to publish under Scottish imprints; or, that few Scottish people choose to publish domestically. Other information in table 1 favors the latter view: only 10 percent of materials published by Scottish people is published domestically. Similar findings are associated with materials



published about Scotland: roughly a quarter of Scottish domestic publishing output comprises materials about the home country—and of the materials published about Scotland, more than half are published in Scotland—yet relatively little of this material is created or authored by Scottish people. Only 9 percent of the material published by Scottish authors is about their native country; similarly, only 12 percent of the material published about Scotland was created or authored by Scottish people. In short, the data in table 1 suggest that Scottish authors and creators largely look beyond Scotland both for publishing venue and subject.

The publications in the Scottish national presence can be linked to distinct works. Figure 3 indicates the number of works that can be attributed to the national presence as a whole, as well as each of its three components.

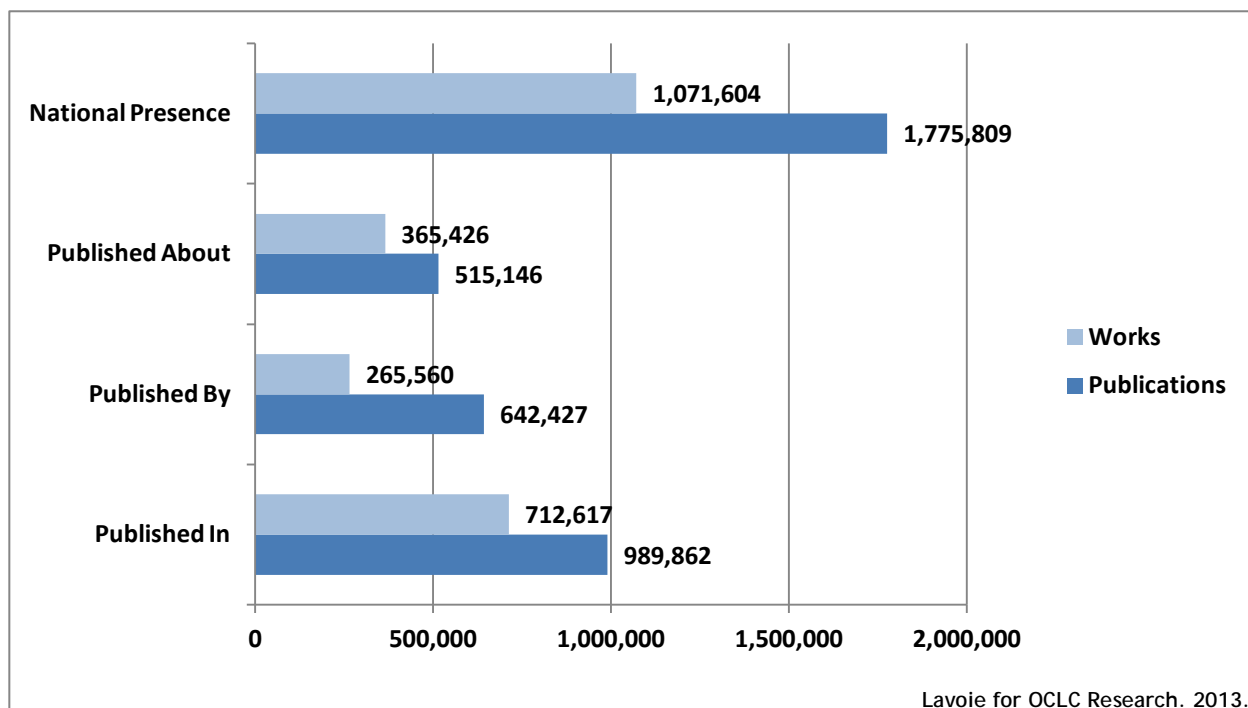


Figure 3. Works in the Scottish national presence

An interesting feature of the data in Figure 3 is that while the “Published In” and “Published About” components—and the national presence overall—reflect similar ratios of publications to works (1.4, 1.4, and 1.7, respectively), the ratio for the “Published By” component is significantly higher (2.4). This suggests that works published by Scottish people tend to be republished more often than works published in Scotland or about Scotland. This difference cannot be explained with the data used in this study, although we can speculate that many Scottish authors who publish abroad do so because they have achieved some degree of international renown; this in turn suggests that their work is sufficiently popular to warrant republication in new editions or translations. On the other

hand, it may be the case that those who are republished are more likely to be renowned, and therefore more likely to have a Wikipedia page—and therefore more likely to have their works identified as produced by a Scottish person (given the methodology used in this report).<sup>23</sup> In any case, it seems that works by Scottish authors have a particularly significant impact in terms of extending the national presence.

The Scottish national presence includes materials in 218 languages; materials published in Scotland include 130 languages; materials published by Scottish people, 174 languages; and materials about Scotland, 123 languages. The vast majority—87 percent—of the Scottish national presence is published in English, the nation’s primary language. However, while 92 percent of the materials published in Scotland, and 93 percent of the materials published about Scotland, are in English, only 79 percent of the materials published by Scottish people are in English. This provides further evidence that Scottish authors and creators tend to publish overseas. Presumably some of these overseas publishing venues are in non-English speaking countries, and would therefore increase the proportion of non-English language materials *vis-à-vis* the other two components of the Scottish national presence.

Table 2 reports the five most frequently-occurring languages other than English for the Scottish national presence and its three components.

**Table 2. Five most frequently-occurring languages other than English**

Published In	Published By	Published About	National Presence
Latin	German	French	Latin
Scottish Gaelic	French	Latin	German
Scots	Latin	German	French
French	Spanish	Scottish Gaelic	Spanish
Spanish	Japanese	Spanish	Scottish Gaelic

An interesting feature of the data in table 2 is the prominence of Latin-language materials in the Scottish national presence, as well as each of its three components. More than 26,000 publications published in Scotland are in Latin, as well as nearly 9,000 materials published by Scottish people, and about 4,500 materials about Scotland. All told, more than 35,000 distinct publications in the Scottish national presence are published in Latin. The median publication date for these materials is 1786, suggesting that these materials are generally quite old, and are likely valued not just for their content but also as historical artifacts.

Although Scotland is primarily an English-speaking country, it possesses two languages native to its people: Scottish Gaelic and Scots. Figure 4 shows the number of publications in these languages in the Scottish national presence and its components.

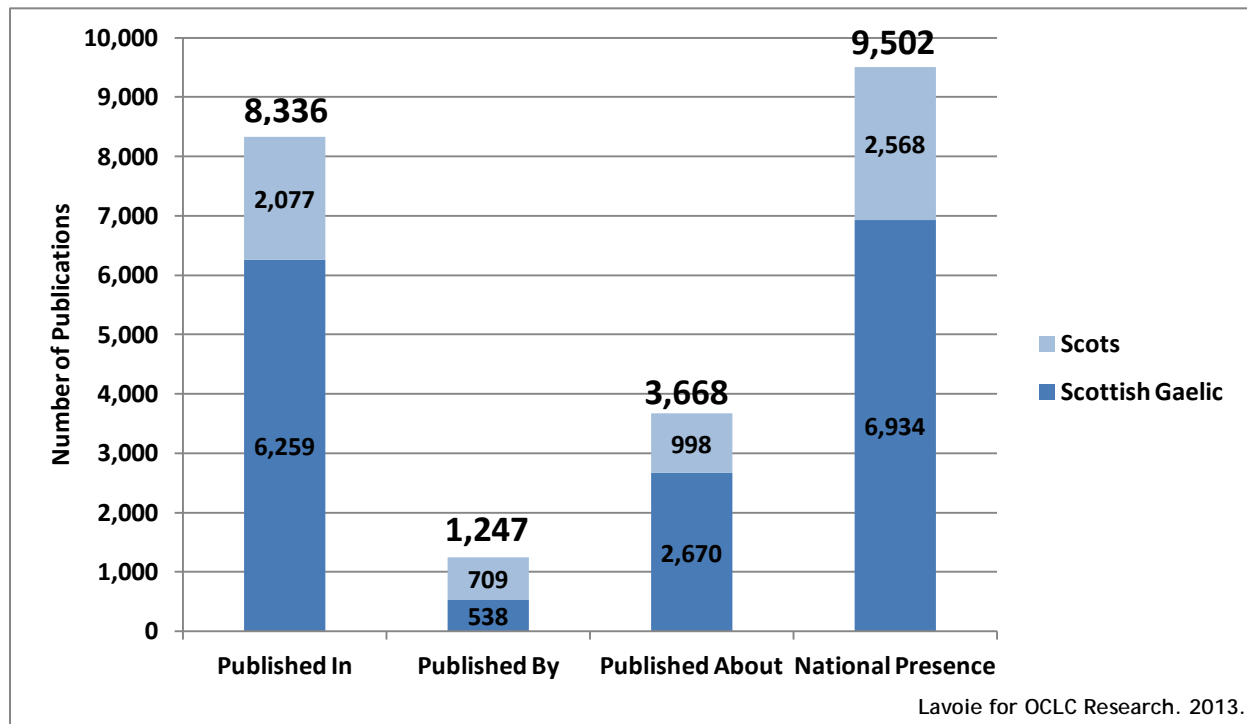


Figure 4: Materials published in Scottish-Gaelic and Scots

The vast majority of Scottish Gaelic- and Scots-language materials in the Scottish national presence are published domestically (although constituting less than 1 percent of all materials published in Scotland). What is perhaps curious is that only a small fraction of the Scottish Gaelic- and Scots-language materials are authored or created by Scottish people. One possible explanation is that these materials are translations from other languages. In fact, 1,335 of the Scottish Gaelic- and Scots-language materials are indeed cataloged as translations from other languages. In the cases where the original language was cataloged, English is predominant (755); other languages include Ancient Greek (91), Latin (28), Hebrew (24), French (16), and Welsh (12). Cataloging for 148 publications indicated Scottish Gaelic was the original language, and so are translations into Scots, while cataloging for 14 publications indicated Scots as the original language, and so were translated into Scottish Gaelic.

While translations may provide a partial explanation for the apparent paucity of Scottish people publishing in the two native Scottish languages, another explanation may be that there are in fact many Scottish people publishing in these languages, but they have not achieved sufficient renown to merit a Wikipedia page. If so, the methodology used in this study would not have identified them as Scottish authors. However, it is doubtful that the

number of Scottish authors omitted in this way would be significant, because as the data in figure 4 suggests, most Scottish Gaelic- and Scots-language materials are published domestically. Therefore, publications by lesser-known Scottish authors in these languages would likely have been identified among the materials published in Scotland.

Another question related to Scottish Gaelic- and Scots-language materials is whether there are places in Scotland that specialize in publishing in these languages. The leading centers (by number of publications) for Scottish publishing in these languages are Edinburgh and Glasgow. However, the data revealed several other places with significant publishing output in Scottish Gaelic and Scots, including Inverness, Aberdeen, Sterling, and Perth. London was the leading location outside of Scotland. The Isle of Lewis (in particular, Stornoway) was also a significant producer of these materials.

Figure 5 reports the distribution of publication dates for the materials in the Scottish national presence.

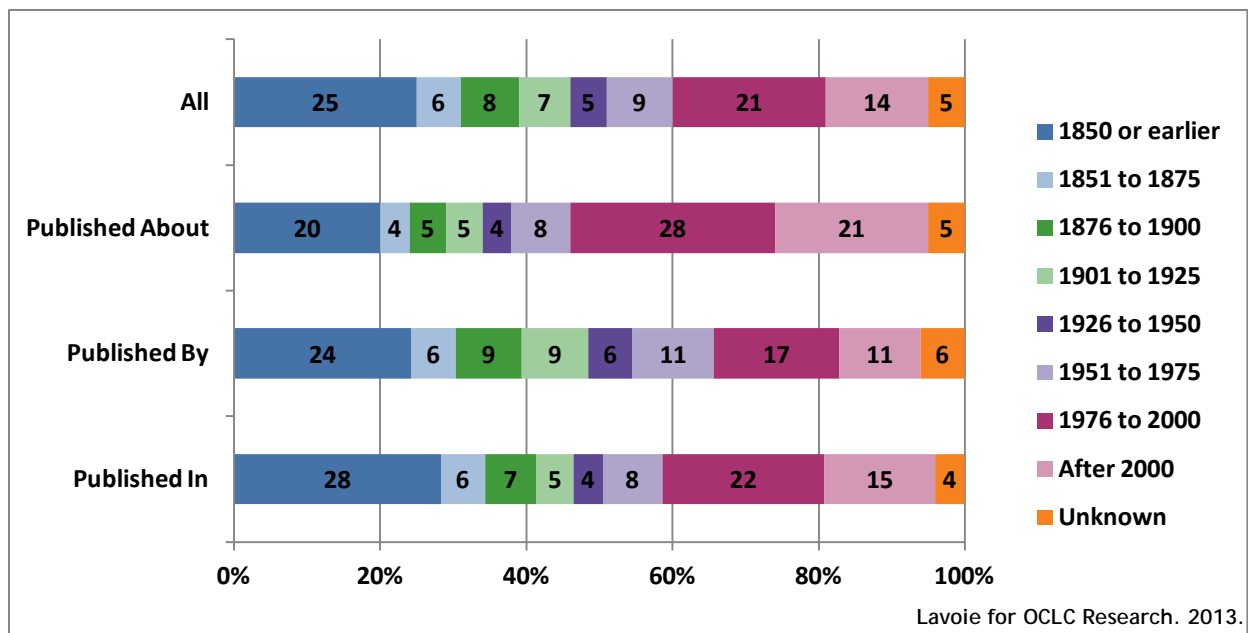


Figure 5: Distribution of publication dates in the Scottish national presence (percent)

For materials published in Scotland or by Scottish people, as well as the Scottish national presence as a whole, publications dating from 1850 or earlier constitute the largest share of the collection. The fact that fully a quarter of the Scottish national presence comprises materials published no later than 1850 is remarkable: for comparison, the global library resource as a whole (as represented by the WorldCat database) exhibits only a 6 percent share of these materials. In many cases, pre-1850 materials are valued not only for their

content but also as historical artifacts, and may receive special curatorial interpretation, preservation, and security.<sup>24</sup>

The median age (years since publication) for materials in the Scottish national presence as a whole is about 65 years. In terms of the individual components of the Scottish national presence, materials published about Scotland seem to have the highest degree of currency, while materials published by Scottish people have the least. The median age of a publication about Scotland is approximately 35 years. In the case of publications by Scottish people, the median age is approximately 85 years. For materials published in Scotland, the median age is about 70 years.

## Global Diffusion of the Scottish National Presence in the Published Record

A nation's cultural and intellectual heritage exerts its influence in many ways. We are accustomed to acknowledging this influence in areas such as language, cuisine, and the media. But as we have seen, a national presence can be identified within the published record, and this presence too has a role in projecting a country's culture and ideas worldwide. Measuring the international diffusion of a national presence in the published record can be approached from a variety of perspectives; this study uses the presence of published materials in library collections around the world as a signal of wider cultural, educational, and scholarly influence.<sup>25</sup>

### International Patterns of Diffusion of the Scottish National Presence

Table 3 reports the number of library holdings worldwide for the materials in the Scottish national presence and each of its components.

**Table 3. Worldwide holdings of the Scottish national presence**

	Holdings	Holdings Per Publication
National presence	19,028,307	10.7
Published In	6,432,538	6.5
Published By	8,478,076	13.2
Published About	7,525,566	14.6

The materials comprising the Scottish national presence account for nearly 20 million holdings in library collections worldwide. While in absolute terms this is certainly a large number, it represents only about 1 percent of the nearly 1.8 billion holdings attached to the global library resource approximated by WorldCat. It is difficult to find a benchmark against which

to assess whether the Scottish influence, as represented by library holdings, is large or small. However, it is worthwhile noting that the average number of holdings per publication in WorldCat is about 7.2, compared to 10.7 for a publication in the Scottish national presence. This suggests that the “influence”—as measured by intensity of holdings—of the average publication in the Scottish national presence is higher than that of the average publication in the global library resource.

A more revealing metric of the international influence of Scottish materials may be the holdings-per-publication ratios for the three components of the Scottish national presence. This ratio is relatively low for materials published in Scotland, compared to both the WorldCat average and to the other two components of the national presence. In contrast, materials published by Scottish authors/creators, and materials published about Scotland are collected quite heavily worldwide, with holdings-per-publication ratios approximately double that of the average for the global library resource. This suggests that it is primarily through these channels that the Scottish national presence in the published record diffuses itself worldwide.

Further insight on this point can be obtained by examining the distribution of Scottish national presence holdings across countries. Figure 6 reports these results.

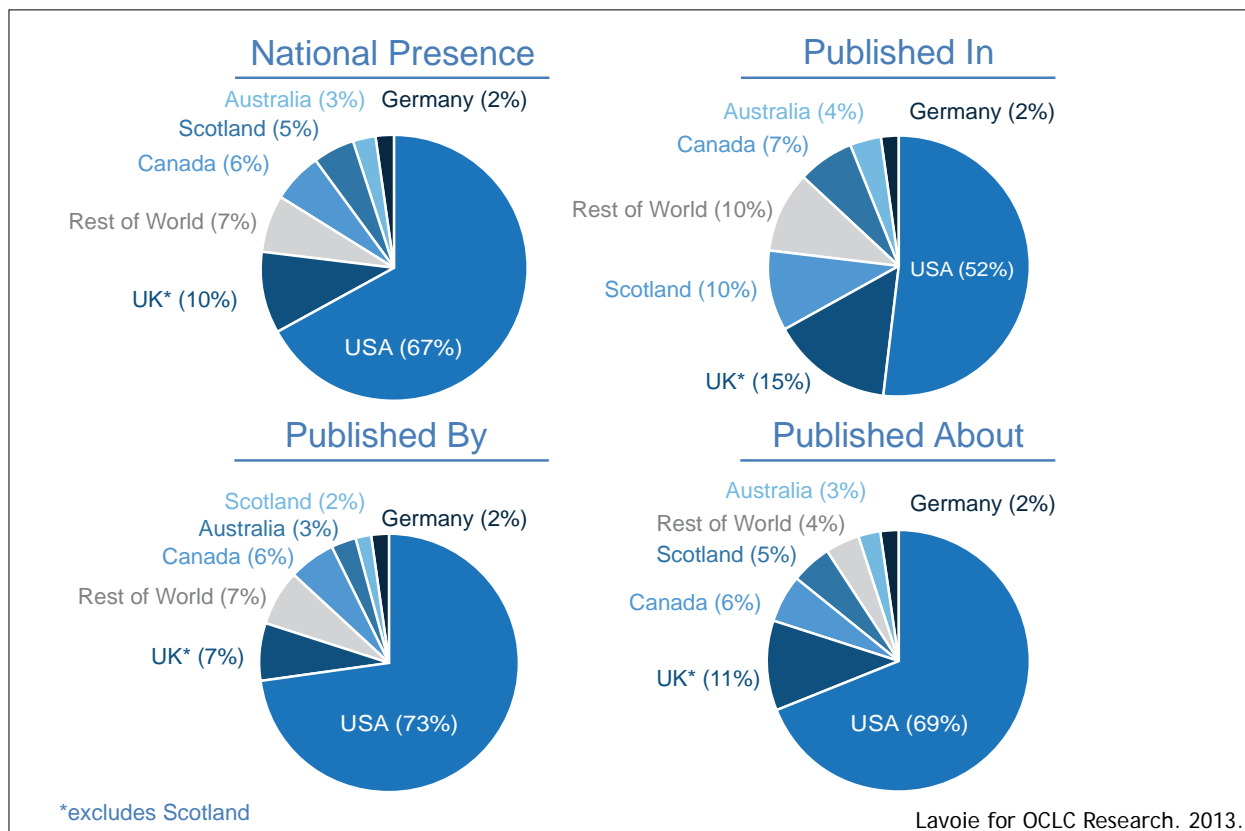


Figure 6: Global diffusion of Scottish national presence (holdings)

The data in figure 6 indicate that the Scottish national presence is manifested (in the form of materials held in library collections) chiefly in places other than Scotland. The United States is the location for the majority of holdings associated with the Scottish national presence or any of its components, with the UK (excluding Scotland), Canada, and Australia also highly ranked. Scottish holdings account for between two and ten percent of the holdings. This is not unexpected—Scotland’s size in comparison to other nations is quite small—but nevertheless underscores the point that Scotland’s national presence in the published record is manifested primarily outside of Scotland.

Materials published in Scotland appear to have a higher tendency to be collected domestically, with Scottish institutions accounting for 10 percent of the holdings associated with Scotland-published materials, compared to only 5 percent for materials published about Scotland and 2 percent for materials published by Scottish people. Several factors may account for this result. Smaller publishing houses are perhaps more likely to have only national or regional audiences, and therefore would be more likely to have their publishing output collected by domestic institutions rather than those overseas. Moreover, Scottish libraries and other collecting institutions are likely to pay special attention to collecting the output of the domestic imprint; indeed, the National Library of Scotland has legal deposit privileges for all printed materials published in the UK or the Republic of Ireland.

Diffusion of the Scottish national presence around the world can also be tracked through the identification of concentrations of Scotland-related materials in library collections. Table 4 reports the largest “Scotland centers” around the world in the context of the Scottish national presence as a whole, and for its three component parts.<sup>26</sup>

**Table 4. Largest concentrations of materials in the Scottish national presence, worldwide**

National Presence	Published In	Published By	Published About
Nat. Lib. of Scotland	Nat. Library of Scotland	British Library	Nat. Lib. of Scotland
British Library	British Library	Natl. Library of Scotland	British Library
U. of Edinburgh	U. of Edinburgh	Harvard University	U. of Oxford
U. of Glasgow	U. of Glasgow	Yale University	U. of Edinburgh
U. of Oxford	U. of Oxford	U. of Toronto	U. of Glasgow
U. of Cambridge	U of Cambridge	U. of Michigan	Harvard University
Harvard University	U. of Aberdeen	U. of Cambridge	U. of Cambridge
Yale University	Harvard University	New York Public Library	Yale University
U. of Michigan	Yale University	U. of Oxford	Library of Congress
U. of Toronto	U. of Michigan	Library of Congress	U. of Strathclyde

In terms of the national presence as a whole, the list is dominated by Scottish institutions, with the National Library of Scotland ranking first as the largest concentration of Scotland-related materials in the world. This result is not unexpected, nor is the British Library's ranking as the second-largest concentration. Both of these institutions—the NLS in particular—would view the acquisition of Scotland-related materials as a key component of their collecting mission. Large concentrations also exist at some of the leading Scottish universities, as well as Oxford and Cambridge. The list is rounded out by three American universities and one Canadian university; however, it is not clear what inference we can draw from their presence on the list. All are members of the Association of Research Libraries (ARL), and in 2011 they represented four out of the five largest ARL member library collections in terms of volumes held.<sup>27</sup> In light of this, it is not clear whether their high ranking reflects an emphasis on collecting Scottish materials, or is simply proportionate to the large size of their collections.

While Scottish and other UK-based institutions are well-represented in the rankings for the Scottish national presence as a whole and for materials published in Scotland and about Scotland, the third component of the national presence—materials published by Scottish people—includes a high proportion of non-UK institutions. This suggests that concentrations of Scottish materials outside Scotland and the UK may exhibit a heavier emphasis on the works of Scottish authors or creators than their Scottish/UK-based counterparts. This result aligns with several other findings mentioned earlier: the inference from table 1 that Scottish nationals largely look outside Scotland both for publishing venue and subject; and the results from figure 6 indicate that materials published by Scottish people exhibit the smallest proportion of holdings by Scottish institutions. The implication seems to be that of all the components of the Scottish national presence, it is materials published by Scottish authors and creators that project the most influence abroad. We return to this hypothesis, and consider some additional evidence bearing on its legitimacy, in the next section.

## Core Works in the Scottish National Presence

One way to characterize the Scottish influence in the published record is to assess the global ubiquity of particular Scottish works. A variety of methods can be used to do this; we will focus on two approaches that are suited to the data sources used in this study. First, we will examine which works in the Scottish national presence have been republished the most over time. Second, we will look at which Scottish works are the most widely held in library collections around the world. Underpinning this analysis is the idea of a core work: that is, a work within a particular national presence that projects an exceptionally large influence in the global published record.



A work with many distinct publications associated with it has been republished many times, in the form of new editions, translations, and so on. This serves as a signal of the work’s enduring popularity and influence. Table 5 lists the largest work clusters—measured as the number of distinct publications associated with a particular work—in the Scottish national presence.

**Table 5. Top 10 largest work clusters in the Scottish national presence**

Works	Publications
Treasure Island	3,456
An Inquiry into the Nature and Causes of the Wealth of Nations	1,829
Dr. Jekyll and Mr. Hyde	1,700
The Hound of the Baskervilles	1,553
The Life of Samuel Johnson	1,519
Adventures/Memoirs of Sherlock Holmes	1,440
The Wind in the Willows	1,350
Kidnapped	1,346
Lectures on Rhetoric and Belles Lettres	979
Peter Pan	931

A key feature of the works listed in table 5 is that all were authored by Scottish people. This further reinforces the proposition that it is materials authored by Scottish authors/creators that exert the greatest Scottish influence on the published record. Robert Louis Stevenson appears to be of particular significance in this regard, with three works on the list, including the top-ranked entry. Arthur Conan Doyle is the only other author with multiple entries on the list.

Another way to measure a work’s influence in the published record is to calculate how many libraries around the world hold a publication of that work in their collection. The presence of Scottish works in library collections serves as a signal of their broader cultural and scholarly influence. Table 6 reports the works in the Scottish national presence most widely held in library collections worldwide.

**Table 6. Top 10 works in Scottish national presence most widely held in library collections**

Works	Total Global Holdings
Treasure Island	44,742
An Inquiry into the Nature and Causes of the Wealth of Nations	30,580
The Wind in the Willows	29,863
Dr. Jekyll and Mr. Hyde	26,210
Kidnapped	24,807
Adventures/Memoirs of Sherlock Holmes	22,403
Peter Pan	21,352
Macbeth	20,563
The Life of Samuel Johnson	20,125
The Hound of the Baskervilles	19,079

The ranking of works in table 6 closely tracks the ranking in table 5, with some re-ordering of the entries. One new entry appears in the list: *Macbeth*—the only entry without a Scottish author, and the most widely held work globally that is about Scotland not written by a Scottish author.

Table 6 offers few surprises in terms of a list of “core” Scottish works, but exploring the lower levels of the ranking of most widely held works in the Scottish national presence reveals some titles that are perhaps not quite as familiar. For example, table 7 reports the ten works in the Scottish national presence ranking 50th through 59th in terms of global library holdings.

**Table 7. Works in Scottish national presence most widely held in library collections (50th-59th)**

Works	Author	Total Global Holdings
The Poems of Ossian	James Macpherson	5,395
The Daughter of Time	Josephine Tey	5,331
Forbes	[periodical]	5,277
At the Back of the North Wind	George MacDonald	5,270
Casebook of Sherlock Holmes	Arthur Conan Doyle	5,247
The Poetical Works of Robert Burns	Robert Burns	5,243
Harry Potter and the Sorcerer's Stone	[movie]	5,206
Essays: Moral, Political, and Literary	David Hume	5,196
The Lord of the Rings: The Two Towers	[movie]	5,114
To the Hilt	Dick Francis	5,094

Table 7 includes both familiar and perhaps not-so-familiar entries. Once again, works by Scottish authors dominate the list: Macpherson, Tey, MacDonald, Doyle, Burns, and Hume. The inclusion of Forbes, a business periodical, perhaps requires explanation: Forbes was founded by the Scottish financial journalist B.C. Forbes. *To the Hilt* is a novel by the Welsh-born mystery writer Dick Francis, and is partially set in Scotland. Finally, the inclusion of the Harry Potter and Lord of the Rings movies in the list can be accounted for by a cataloging convention of listing the actors in a movie as "authors" in the bibliographic record. The algorithm identifying materials published by Scottish authors or creators therefore flagged both movies because of the presence of Scottish-born actors in the cast: for example, Robbie Coltrane in the Harry Potter movie, and Billy Boyd in the Lord of the Rings movie. We leave to the reader to judge whether this is sufficient grounds for inclusion of these and similar movies in the Scottish national presence!

Digging even deeper into the rankings of most widely held works in the Scottish national presence, table 8 reports the ten works falling in slots 90 through 99.

**Table 8. Works in Scottish national presence most widely held in library collections (90th-99th)**

Works	Author	Total Global Holdings
Past and Present	Thomas Carlyle	4,239
The Touch	Colleen McCullough	4,234
Waverly	Walter Scott	4,212
London Journal, 1762-1763	James Boswell	4,210
His Last Bow	Arthur Conan Doyle	4,147
Harry Potter and the Prisoner of Azkaban	[movie]	4,077
Men of Mathematics	Eric Temple Bell	4,053
The Encyclopedia of Mammals	David W. Macdonald	4,049
Harry Potter and the Chamber of Secrets	[movie]	4,019
John Paul Jones, A Sailor's Biography	Samuel Eliot Morison	4,015

As we move deeper into the rankings, more contemporary works begin to appear. Six of the ten works listed in table 8 were originally published in the 20th century or later. As with previous portions of the rankings, works by Scottish authors predominate, although the biography by Samuel Eliot Morison, an American historian, appears on the list by virtue of his subject: John Paul Jones was born in Scotland.

As the lists presented in the three previous tables suggest, the most widely held works in the Scottish national presence—the “core works”—tend to be ones that were originally published long ago. The median year of publication for the materials in the Scottish national presence is 1950. Compiling a list of the ten most widely held works in the Scottish national presence originally published before 1950 yields a ranking identical to the one presented in table 6. This ranking yields few surprises; many readers would have been able to predict most of the works on the list, if not the precise ranking. What is perhaps of more interest is the segment of the pre-1950 period that is of particular importance to Scotland: the Scottish Enlightenment. It was during the Scottish Enlightenment that Scotland's intellectual and cultural influence on the rest of the world was at its zenith, and much of this influence was projected through works published by Scottish authors during this period. The boundaries of the Scottish Enlightenment have been variously defined; we confine our attention to the period 1740-1800. Table 9 provides a list of the most widely held works by Scottish authors originally published during the Scottish Enlightenment.

Table 9. Most widely held Scottish Enlightenment works (1740-1800)

Work	Author	Total Global Holdings
An Inquiry into the Nature and Causes of the Wealth of Nations	Adam Smith	30,580
The Life of Samuel Johnson	James Boswell	20,125
Lectures on Rhetoric and Belles Lettres	Hugh Blair	8,921
Enquiries Concerning the Human Understanding and Concerning the Principles of Morals	David Hume	6,835
Roderick Random	Tobias Smollett	6,611
The Theory of Moral Sentiments	Adam Smith	5,441
The Complete Poetical Works of Robert Burns	Robert Burns	5,430
The Poems of Ossian	James Macpherson	5,395
The Poetical Works of Robert Burns	Robert Burns	5,243
Essays, Moral, Political, and Literary	David Hume	5,196
Dialogues Concerning Natural Religion	David Hume	5,030

Readers will note the appearance of two seemingly identical Robert Burns works on the list. Robert Burns' poems are generally published in collections, each of which might be considered a distinct work in that the collections will have different editors, annotations, commentary, and so on. But because they often have very similar or even identical titles, and the principal author is usually given as Robert Burns, the algorithm that clusters publications into works tends to view these as different publications of the same work, and clusters them accordingly. Thus, the two Burns' entries in table 9 represent two classes of materials: collections of Burns' poems that share the title "The Complete Poetical Works of Robert Burns", and those that share the title "The Poetical Works of Robert Burns". The slight difference in titles is enough for the algorithm to categorize them as different works. This poses a dilemma as to whether these two "works" should be combined in the rankings, or kept separate. For the purposes of this analysis, the two "works" are treated as distinct, on the grounds that while one explicitly states that it includes the complete poetical works of Burns, the other does not, and therefore may include collections comprised of various combinations of poems selected from Burns' complete corpus.

The works listed in table 9 are widely acknowledged as classics. The fact that they are still prominent in library collections today is testimony to the durability of interest in them worldwide. But what about newer works? Are there new Scottish classics emerging? While we cannot use the "test of time" as a metric to gauge contemporary works' potential for enduring influence, we can at least make a preliminary exploration of the question by

examining which of the newer works in the Scottish national presence are heavily collected by libraries around the world. Table 10 lists the most widely held works in the Scottish national presence published after 1950.

**Table 10. Most widely held works in the Scottish national presence, published after 1950\***

Work	Author	Total Global Holdings
The Sunday Philosophy Club	Alexander McCall Smith	6,362
44 Scotland Street	Alexander McCall Smith	5,974
Whiteout	Ken Follett	5,944
Mary Queen of Scots	Antonia Fraser	5,873
Dougal Dixon's Dinosaurs	Dougal Dixon	5,852
Friends, Lovers, Chocolate	Alexander McCall Smith	5,837
The Professor and the Madman	Simon Winchester	5,753
Outlander	Diana Gabaldon	5,716
The Daughter of Time	Josephine Tey	5,331
To the Hilt	Dick Francis	5,094
The Sunday Philosophy Club	Alexander McCall Smith	6,362

*\*Movies excluded*

One key feature of the list in table 10 is that at least<sup>28</sup> half of them make the ranking on the basis that they are in some way about Scotland, but they are not authored or created by a Scottish-born author. This is in contrast to the previous lists, where works authored or created by Scottish people predominate. In considering contemporary Scottish influence in the published record, there might be particular interest in widely held works by currently or recently active Scottish authors. Table 11 presents the most widely held works by Scottish authors, published after 1950.

Table 11. Most widely held works by Scottish authors, published after 1950\*

Work	Author	Total Global Holdings
Dougal Dixon's Dinosaurs	Dougal Dixon	5,852
The Daughter of Time	Josephine Tey	5,331
The Prime of Miss Jean Brodie	Muriel Spark	5,071
The Ascent of Money: A Financial History of the World	Niall Ferguson	4,393
Exit Music	Ian Rankin	4,295
The Naming of the Dead	Ian Rankin	4,267
The Encyclopedia of Mammals	David W. Macdonald	4,049
Food in History	Reay Tannahill	3,866
Fleshmarket Close <sup>29</sup>	Ian Rankin	3,702
After Virtue: A Study in Moral Theory	Alasdair C. MacIntyre	3,625
Dougal Dixon's Dinosaurs	Dougal Dixon	5,852

\**Movies excluded*

Many readers will notice that the author Alexander McCall Smith has disappeared from the list in table 11. This may be surprising, as many would associate Smith with Scotland, and indeed his Wikipedia page indicates his nationality is Scottish. As it turns out, the omission of Smith is a consequence of the methodology used in this study to identify Scottish people in the DBpedia data. The methodology relied primarily on birthplace to determine if someone was or was not Scottish. Smith was born in what was then Rhodesia, which was explicitly noted in his DBpedia data. The algorithm therefore concluded he was not Scottish. Smith's work appears in table 10 not because he was identified as a Scottish author, but because the works noted are about—i.e., set in—Scotland. Conspicuously missing from table 10 are Smith's popular No. 1 Ladies' Detective Agency mystery novels, which are set in Botswana. As discussed earlier, the benefits of the methodology used in this study is that it is automated and therefore relatively easy to apply. The drawback is that nuanced cases like Alexander McCall Smith may be inappropriately categorized.<sup>30</sup> Of course, manual refinements can always be added to the algorithm's results, but access to structured data that explicitly notes an individual's nationality would be the ideal solution.

Two observations seem appropriate concerning influential contemporary works in the Scottish national presence. First, Scottish mystery writing, in the form of novels written by Scottish authors like Ian Rankin, or set in Scotland, like the Isabel Dalhousie novels by Alexander McCall Smith, seem to be especially prominent, suggesting that it is this genre that forms the nucleus of the "new classics"—or contemporary core works—of the Scottish national presence. This seems to be corroborated by the recent emergence of "Tartan Noir" as an internationally recognized form of detective fiction.<sup>31</sup> Second, there seems to be a discernable increase in

the influence of works about Scotland. Consider that the most widely held works in the Scottish national presence published before 1950 (which, as noted earlier, corresponds to the list in table 6) were all written by Scottish-born authors. In contrast, half of the list of most widely held works in the Scottish national presence published after 1950 (table 10) were works that were either set in Scotland or about a Scottish subject—but written by a non-Scottish author (we do not include the works by Alexander McCall Smith in this total). This suggests that Scotland may be increasingly projecting its influence in the published record by lending itself as place, as well as its history and culture, to non-Scottish authors and creators as inspiration for their works. This is perhaps part of a broader trend of globalization in culture, arts, the media, etc; one might contrast this with earlier times, when outputs in these areas were more local in context and focus.<sup>32</sup>

Finally, we examine the diffusion of the Scottish national presence worldwide by considering how that diffusion varies across countries. Table 12 shows the most widely held works in the Scottish national presence in three countries: Scotland, the US, and Australia.

**Table 12. Most widely held works in Scottish national presence: Scotland, US, and Australia**

Scotland	US	Australia
Treasure Island	Treasure Island	Treasure Island
Wealth of Nations	The Wind in the Willows	The Wind in the Willows
The Life of Samuel Johnson	Wealth of Nations	Wealth of Nations
The Poems of Ossian	Dr. Jekyll and Mr. Hyde	Macbeth
Gentle Shepherd	Kidnapped	Dr. Jekyll and Mr. Hyde
Kidnapped	Adv./Mem. of S. Holmes	Peter Pan
Dr. Jekyll and Mr. Hyde	Peter Pan	The Life of Samuel Johnson
The Expedition of Humphrey Clinker	Macbeth	Kidnapped
The Wind in the Willows	The Life of Samuel Johnson	A Child's Garden of Verses
Roderick Random	A Child's Garden of Verses	Adv./Mem. of S. Holmes

The salient feature of table 12 is that while the US and Australian lists contain the same works (albeit with different ordering), the Scottish list is considerably different, containing four works which do not appear on the other two lists. These results suggest an interesting question: as a general rule, are the works in a given national presence that are most influential domestically significantly different from those most influential abroad? The lists in table 12 also suggest some similarities across countries in the perceived core works of the Scottish national presence. In particular, *Treasure Island* is the clear favorite in all three countries, while *Wealth of Nations* also ranks highly. This result, combined with other data



reported earlier in the study suggests that *Treasure Island* may be the most influential work internationally in the Scottish national presence.

Worldwide access to the Scottish national presence in the published record would be facilitated by the availability of digitized copies of print materials. To explore this point, the publications in the Scottish national presence were compared to the digitized texts in the Hathi Trust Digital Library, which is a corpus of digitized print books.<sup>33</sup> Only a small percentage—3 percent, or 51,669 distinct publications—of the Scottish national presence is currently represented in the Hathi Trust corpus. Since the Scottish national presence includes materials in a variety of formats, not everything would be eligible for inclusion in Hathi; restricting the Scottish national presence to print books only, the coverage is slightly higher at 5 percent. These results are in no way indicative, of course, of the full availability of the Scottish national presence in digital form; it merely represents the overlap with one corpus of digitized materials. However, Hathi Trust is a significant digital library in North America, and therefore useful as a means of exploring the characteristics of the Scottish national presence within a large collection of digitized materials located outside of Scotland.

Table 13 lists the works in the Scottish national presence with the most publications in the Hathi Trust collection.

**Table 13. Works in Scottish national presence with largest work clusters in Hathi Trust collection**

Work	Author	Publications
The Life of Samuel Johnson	James Boswell	99
Wealth of Nations	Adam Smith	83
Lectures on Rhetoric and Belles Lettres	Hugh Blair	65
The Poems of Ossian	James Macpherson	60
The Poetical Works of Robert Burns	Robert Burns	45
The Complete Poetical Works of Robert Burns	Robert Burns	40
The French Revolution: A History in Three Parts	Thomas Carlyle	37
Critical and Miscellaneous Essays	Thomas Carlyle	35
Schiller's Mary Stuart	Friedrich Schiller	32
Treasure Island	Robert Louis Stevenson	29

Boswell's *The Life of Samuel Johnson* is the work in the Scottish national presence most abundantly represented in the Hathi corpus, with 99 distinct publications. Smith's *The Wealth of Nations* also enjoys prolific representation, with 83 distinct publications. As with lists we have seen earlier, table 13 is dominated by Scottish-born authors—with one exception, the

German playwright Friedrich Schiller. An interesting characteristic of the list in table 13 is that it is quite different from that in table 5, which ranks the overall size of the work clusters in the Scottish national presence. The list in table 13 seems to be composed almost entirely of works that are primarily of scholarly interest; works of a more popular interest, like *The Wind in the Willows*, *Sherlock Holmes*, and *Peter Pan*, prominent in table 5, have disappeared in table 13 (although works by Robert Burns and Robert Louis Stevenson remain). This undoubtedly reflects the nature of the primary contributors to the Hathi Trust collection, which are academic research libraries. In this sense, the list in table 13 may provide some indication of the Scottish works that are particularly influential in scholarly circles, reflected in the collecting behaviors of the academic and research libraries that serve them.

## Conclusion

The Scotland case study illustrates how the concept of a national presence in the published record can be operationalized in library data, and used to define patterns of analysis characterizing the general contours of the national presence and its diffusion worldwide. The case study proposes and tests a methodology for identifying a national presence in library bibliographic data that relies primarily on automated processing with minimal manual intervention, and can be re-purposed without extensive customization for most countries.

Application of this methodology to Scotland yielded a number of insights into the Scottish national presence in the published record. It is a resource of significant size, widely held in library collections around the world, and containing a large proportion of older (and perhaps historic) materials. Republishing rates are, on average, higher for works in the Scottish national presence when compared to the average work in the WorldCat database.

Works by Scottish-born authors seem to be a particularly influential component of the Scottish national presence, as measured by a variety of indicators. This suggests that it is Scotland's authors and creators, rather than its domestic publishing output, or itself as a subject, which is most effectual in promulgating the national presence overseas. However, the analysis also suggests that works about Scotland, or that use Scotland as a setting, are becoming more numerous in contemporary literature, and may be emerging as another key channel for diffusing Scotland, its culture, and its intellectual heritage around the world.

Most holdings of materials in the Scottish national presence are by institutions outside of Scotland, which reminds us that a national presence in the published record may be primarily manifested outside the home country's borders. Analysis of global library collecting activity provides a means of identifying works in the Scottish national presence that have achieved an enduring presence in the published record. Most of these core works are familiar classics in literature, arts, and science written by Scottish authors like Smith, Hume, and Stevenson.

More contemporary works signal new channels through which Scotland “exports” itself through its national presence—for example, popular historical novels set in Scotland, or the “Tartan noir” mystery genre. However, the Scottish example suggests that there may be significant differences between what is perceived as a core work domestically and overseas. A variety of metrics in the analysis suggest works by Scottish-born authors exhibit a somewhat lesser presence domestically, and a greater presence overseas.

Finally, the analysis reveals that Robert Louis Stevenson’s *Treasure Island* consistently tops the rankings across a variety of indicators measuring the relative impact of works in the Scottish national presence. *Treasure Island* is the most widely held and most widely republished work in the Scottish national presence, and its popularity seems to be consistent both domestically and overseas. In short, *Treasure Island* appears to be the most globally influential work in the Scottish national presence. Rather than Scotch whisky, perhaps it is the pirates’ legendary “bottle of rum” that we should toast as the iconic drink of Scotland!

The methodology used to identify a national presence in library data has a number of shortcomings, several of which have already been noted. In addition to these, three others are of particular significance. The bibliographic data used in this study does not include journal articles written about Scotland, or by Scottish authors.<sup>34</sup> The study would be improved if it could expand to include these materials. Second, while WorldCat is the closest approximation available for the global published record, as represented in library collections, its coverage of materials and library holdings is not exhaustive. Gaps in WorldCat coverage will therefore be reflected in a national presence analysis, with the resulting impact on inferences commensurate with the severity of the gap. Cataloging inconsistencies and errors may also have an impact on the data.

How much of a concern are these issues? The answer is mixed. Of course, better and more comprehensive data would make for better inferences; however, from a service perspective—that is, services operating on WorldCat as a data layer—the picture of the global library resource presented by WorldCat is the reality that people see. Elements of a national presence not represented by data accessible to services are, for all intents and purposes, invisible in the context of the global library resource. Nevertheless, the methodology described in this study can certainly be refined and improved to enhance both its accuracy and ease of implementation. The component of the methodology identifying individuals from a particular country will also improve as the data available from DBpedia improves, and as links between DBpedia and other data sources are strengthened. For example, the recent VIAFbot project experimented in creating reciprocal links between biographical Wikipedia articles and the Virtual International Authority File.<sup>35</sup>

The methodology and analysis described in this report would be of interest to cultural heritage institutions in any country interested in exploring a national presence in the published record in the context of collection development strategies, prioritization of digitization activities, and “gap analysis” for library collections, and would likely be of special interest to national libraries, academic institutions, and public libraries tasked with the responsibility to collect the intellectual and cultural contributions of a particular country. A national presence in the published record would also offer a fertile data set for scholarly research. For example, an interesting application of the national presence concept would be to map concentrations of Scottish materials, such as those listed in table 4, against the pattern of diffusion of the Scottish diaspora. Do areas characterized by a heavy influx of Scottish immigrants generally have access to a geographically proximate concentration of materials in the Scottish national presence? This question may be of special relevance to public libraries interested in providing access to materials relating to the cultural and genealogical history of their patrons. It would also be useful to explore whether the patterns and inferences drawn from the Scottish case study can be generalized over many countries. In short, the concept of a national presence in the published record is a valuable tool to benchmark the scope and diffusion of a country’s literary, scholarly, and cultural heritage in an increasingly globalized information landscape.

## Notes

- <sup>1</sup> The collections of other institutions besides libraries are also represented in WorldCat, although libraries predominate.
- <sup>2</sup> As of January 2012
- <sup>3</sup> For example, US GDP at the end of 2010 was about \$14.9 trillion, while GNP was about \$15.1 trillion. See <http://research.stlouisfed.org/fred2/categories/106>.
- <sup>4</sup> See <http://www.nla.gov.au/service-charter>.
- <sup>5</sup> See <http://www.nli.ie/en/about-the-library.aspx>.
- <sup>6</sup> See <http://www.nb.admin.ch/sammlungen/helvetica/index.html?lang=en>.
- <sup>7</sup> See <http://bn.org.pl/en/>.
- <sup>8</sup> Readers familiar with the FRBR entity relationship model will recognize that a publication is equivalent to a FRBR manifestation, and a physical copy to a FRBR item.
- <sup>9</sup> MARC (Machine-Readable Cataloging) is a standard for encoding bibliographic data in a machine-readable record format. See <http://www.loc.gov/marc/bibliographic/ecbdhome.html>.
- <sup>10</sup> See <http://www.gro-scotland.gov.uk/files2/stats/population-estimates/08mye-localities-table2.pdf>. The list was truncated to include only the top 50 Scottish cities, because some of the smaller towns on the list shared names with locales outside of Scotland, resulting in a number of false matches. For example, “California”, “Springfield”, “Houston”, and “Alexandria” are all Scottish towns whose

names coincide with non-Scottish locales with relatively large populations. Truncating the list to include only the top 50 Scottish cities eliminated most of this problem. While this may result in excluding a small quantity of records that truly describe Scottish-published materials, this number is exceeded by the false matches generated by using the full list. It is probably reasonable to assume that the vast majority of Scottish publishing activity occurs in the larger urban areas represented on the top 50 list.

- <sup>11</sup> A “Scottish person” is defined as a person born in Scotland. There are certainly other categories of people who would also be considered Scottish—for example, naturalized citizens, or even the first generation children of Scottish immigrants. However, for simplicity, the most straightforward definition was chosen
- <sup>12</sup> See, for example, Wikipedia’s List of Scottish Writers [http://en.wikipedia.org/wiki/List\\_of\\_Scottish\\_writers](http://en.wikipedia.org/wiki/List_of_Scottish_writers). Goodreads has an interesting list of books set in Scotland or by a Scottish author, but the list includes only 128 books and is far from complete [http://www.goodreads.com/list/show/2103.Best\\_Scottish\\_Fiction](http://www.goodreads.com/list/show/2103.Best_Scottish_Fiction).
- <sup>13</sup> <http://dbpedia.org/About>
- <sup>14</sup> Data set used was DBpedia 3.7, based on Wikipedia dumps from July 2011. The two files used for this study were “persons\_en.nt” and “short\_abstracts\_en.nt”. See <http://wiki.dbpedia.org/Downloads37?v=u9u> for more information.
- <sup>15</sup> The English-language version of Wikipedia is the largest and most comprehensive, and it is likely that use of non-English language versions would produce different results. In some contexts, however, non-English language versions may even be better: for example, the German-language version of Wikipedia may be more accurate in the context of identifying German nationals.
- <sup>16</sup> Limiting the list to the top 50 Scottish cities will likely have a bigger impact here than in the context of identifying publication location (see above), since any locale can yield a significant author/creator, while small locales are unlikely to be publishing centers. However, the expanding the list leads to the same “false positive” problem described in relation to identifying materials published in Scotland.
- <sup>17</sup> The idea here is that the abstracts contain statements like “Adam Smith was a Scottish moral philosopher. . . .”, which would correctly signal that Adam Smith was Scottish. Problematically, they can also contain statements like “Joe Smith enjoyed walking his Scottish terrier. . . .”, which would register as a false positive; however, it turns out this method is remarkably robust, with relatively few errors of this kind.
- <sup>18</sup> See <http://www.oclc.org/research/activities/identities.html>.
- <sup>19</sup> The author thanks his colleague Ralph LeVan for conducting this matching process.
- <sup>20</sup> FAST (Faceted Application of Subject Terminology) is a streamlined, simplified version of the Library of Congress Subject Headings schema. For more information, see: <http://www.oclc.org/research/activities/fast/>.
- <sup>21</sup> Determining whether a subject heading referenced something “about Scotland” was usually straightforward, but sometimes required judgment: although the heading might have some connection to Scotland, did it describe something that was primarily about Scotland? An interesting example is the heading “Stevenson, Fanny Van de Grift”, which references the wife of Scottish writer Robert Louis Stevenson. Mrs. Stevenson was American by birth, was married to Stevenson for fourteen years, and upon his death returned to the United States. Is this enough to make her “about Scotland”? In the author’s opinion, the answer is no, but one could reasonably argue otherwise!
- <sup>22</sup> As represented in WorldCat in January 2012. The four ancient Scottish universities are Aberdeen, Edinburgh, Glasgow, and St Andrews.

- <sup>23</sup> The author thanks his colleague Merrilee Proffitt for this point.
- <sup>24</sup> The author thanks his colleagues Jackie Dooley, Merrilee Proffitt, and Jennifer Schaffner for clarifying this point.
- <sup>25</sup> The author thanks his colleague Lorcan Dempsey for this phrasing of the role of library collections in representing the diffusion of a national presence in the published record.
- <sup>26</sup> While table 4 reports the rankings for Scotland-related concentrations, it does not report the specific number of publications for each collection. It is the practice of OCLC Research not to publicly report statistics attributable to a particular institution without the institution's permission. As with all of the statistics reported in this study, the rankings reflect institutional collections as they are represented in the WorldCat database.
- <sup>27</sup> See <http://interactive.arlstatistics.org/home>.
- <sup>28</sup> The qualifier "at least" is used because of some complications regarding Alexander McCall Smith; see explanation below
- <sup>29</sup> This work was released under the title *Fleshmarket Alley* in the US.
- <sup>30</sup> It should be noted that indiscriminately discarding individuals born outside of Scotland also helped improve the accuracy of the list of Scottish nationals used in this study. Some individuals were tentatively categorized as Scottish because the word "Scottish" appeared in their DBpedia short abstract, when in reality the reference was to a context other than nationality (e.g., "Person X was a Scottish terrier enthusiast"). While the omission of a prominent author like Alexander McCall Smith is unfortunate, it is the author's belief that not imposing this criterion would have made the final list of Scottish nationals far less accurate.
- <sup>31</sup> See [http://en.wikipedia.org/wiki/Tartan\\_Noir](http://en.wikipedia.org/wiki/Tartan_Noir).
- <sup>32</sup> The author thanks his colleague Lorcan Dempsey for this point.
- <sup>33</sup> The author thanks his colleague Constance Malpas for this data.
- <sup>34</sup> While the WorldCat database used in this study includes bibliographic data on journal titles, it does not contain data on the individual articles published in these journals.
- <sup>35</sup> See Max Klein's hangingtogether.org blog post "VIAFbot Debriefing" (2012) for a summary of the VIAFbot project.

## References

Klein, Max. 2012. "VIAFbot Debriefing." *Hangingtogether.org* (blog). 28 November. <http://hangingtogether.org/?p=2306>.

Library of Congress, Network Development and MARC Standards Office. 2013. *MARC 21 Format for Bibliographic Data*. Last updated 29 April 2013. Washington: Library of Congress, Cataloging Distribution Service. <http://www.loc.gov/marc/bibliographic/ecbdhome.html>.

OCLC. 2003. *Libraries: How They Stack Up*. Dublin, Ohio: OCLC. <http://www.oclc.org/content/dam/oclc/reports/librariesstackup.pdf>.

Schulz, Kathryn. 2011. "What Is Distant Reading?" *The New York Times*. 24 June. [http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html?pagewanted=all&\\_r=2&](http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html?pagewanted=all&_r=2&).

Wikipedia. 2013. "Culturomics." Last modified 5 August. <http://en.wikipedia.org/wiki/Culturomics>.