HOME | ABOUT D-LIB | CURRENT ISSUE | ARCHIVE | INDEXES | CALENDAR | AUTHOR GUIDELINES | SUBSCRIBE | CONTACT D-LIB

Evaluating the Impact of the FWF-E-Book-Library Collection in the OAPEN Library: An Analysis of the 2014 Download Data

Ronald Snijder
OAPEN Foundation
r.snijder@oapen.org

Printer-friendly Version

Abstract

The FWF-E-Book-Library is the Open Access repository for all stand-alone publications funded by the Austrian Science Fund (FWF). This collection of e-books is also made available through the OAPEN Library. This paper analyses the usage of the FWF-E-book collection in the OAPEN Library during 2014, in order to measure scholarly impact and societal relevance in the humanities and social sciences. Every time a reader downloads a document, the Internet Protocol address of the provider—an organisation through which the reader accesses the web—is recorded. By combining the usage data and information about the provider, we can make an assumption about who is using a specific monograph. The influence of language is quite profound: books written in German are much more likely to be read within Germany, Austria or Switzerland, while books written in English have a far greater chance to be used all over the globe. Most of the usage is international; only 11% of the total downloads is national. The role of Germany and Switzerland is quite large, amounting to 42% of the total usage. The remaining 47% of the downloads originate from the rest of the world. The role of academic readers is relatively large, compared to governmental, business or non-profit usage. Yet, the biggest group of users have accessed the collection through an ISP. If the mean downloads per subject are analysed, we see large differences per subject: not all subjects enjoy the same amount of 'popularity'. It is clear that the collection has a wider reach than academics, and has been read not only in the German-speaking countries, but world-wide.

1 Introduction

Measuring scholarly impact and societal relevance in the humanities and social sciences can be done in several ways. Here we will look at a collection of e-books from the FWF-E-Book-Library, which is made available through the OAPEN Library. The Austrian Science Fund (FWF) is Austria's central funding organization for basic research. Its purpose is to support the ongoing development of Austrian science and basic research at a high international level. The FWF-E-Book-Library is the Open Access repository for all stand-alone publications funded by the FWF ("Phaidra-FWF-Der Wissenschaftsfonds," n.d.).

The OAPEN Foundation is dedicated to Open Access publishing of academic books. OAPEN operates two platforms: the OAPEN Library and the Directory of Open Access Books (DOAB). At the time of writing, the OAPEN Library contains over 2,350 freely accessible and Open Access academic books from 82 publishers. OAPEN works with publishers to build a quality controlled collection of Open Access books, and provides services for publishers, libraries and research funders in the areas of dissemination, quality assurance and digital preservation (Open Access Publishing in European Networks, 2010).

In the spring of 2013, FWF and OAPEN agreed that all FWF eBooks will be deposited in the OAPEN Library. A large part of this collection is available in the OAPEN Library, and can be accessed here. This paper analyses the usage of the FWF-E-book

collection in the OAPEN Library during 2014. The methodology used is described in the Appendix and in further detail in the article *Measuring monographs: A quantitative method to assess scientific impact and societal relevance* (Snijder, 2013).

Every time a reader downloads a document, the Internet Protocol (IP) address of the provider—an organisation through which the reader accesses the web—is recorded. By assessing this information, it is possible to determine the type of organisation and the country of origin. If a researcher of the University of Vienna downloads a book using her or his office equipment, the IP address (for instance 131.130.87.80) of that university will be logged. Basic information such as address and telephone number are publicly available and can be found using the so called 'WHOIS protocol' (Daigle, 2004). By combining the usage data and information about the provider, we can make an assumption about who is using a specific monograph. To put it differently: the type of provider is used to assess the type of reader. In the example used, the reader is affiliated with an academic institution, based in Austria.

It should be noted that the data does not enable us to identify an individual. Only the provider can be identified, which ensures that no privacy rules have been breached.

Not everybody will have an academic organisation as provider; it may be another type of organisation or it will be an Internet Service Provider (ISP). It is useful to define several groups of organisations and here the following categories are used: academic; government; business; non-profit organisations and the general public. Academic users are seen as the main audience for monographs. If the provider is an ISP, the reader cannot be linked to an organisation. This could mean that the reader is not acting as a member of an organisation, and may be categorised as a member of the general public.

Another explanation may be found in a reader working from home. However, if that person is connected to her or his organisation's network, the logged download information will point to the organisation. In other words: from the point of view of the OAPEN Library—or any website—the reader is identified as member of that organisation. This is especially useful for members of academic organisations, who enjoy access to a large collection of paywalled online resources not available outside that university.

Apart from the provider, information about the country from which the data request originated is available, which may be used as an indication of the reader's nationality. This information can be used to classify the usage a bit more: national versus international. As will be explained further, in this case a division between the German-speaking countries (Austria, Germany and Switzerland) versus the rest of the world is used. We could view international usage as an indication of esteem. The percentage of usage outside national borders may give an indication of the importance of the work. This reflects on the authors: the level of international interest in their publications could be seen as an 'esteem indicator'.

Yet, conclusions regarding these statistics must be drawn with caution. First of all, the information found using the WHOIS protocol must be interpreted: what type of organisation is described? If the organisation is a university, it is quite clear. The question where to draw the line between an ISP and another type of commercial organisation is less easy to answer. Also, organisational affiliation does not tell anything about professional roles. For instance, if the provider is a university, there is no way to tell whether the reader is a student or a professor. Likewise, if the provider is an ISP, we cannot be sure the reader used the online monograph for personal or professional reasons. Regarding nationality, this too is not a 100% match: we could easily imagine a Spanish reader downloading a monograph while residing in the USA. The user statistic would then indicate the USA as country of origin.

See also Appendix: Methodology for a more detailed description of the methodology used.

1.1 The FWF E-book collection and the usage data

In 2014, 146 books of the FWF-E-Book-Library collection were made available via the OAPEN Library. Looking at this collection of books, we can describe several aspects. Here we look at subject and language, which are both not very evenly distributed.

In the OAPEN Library, the subject of the books is described using the BIC classification (Book Industry Communication, 2010). Due to its hierarchical nature, the classification assigned to each book can be abbreviated. This results in a larger group of monographs which share the same—broad—subject. When applying this to the collection, the large amount of history books is immediately clear. See *Table 1: Number of Books* for more details.

Subject	Number of titles
History (HB)	60
Literature: history & criticism (DS)	18
History of art (AC)	12

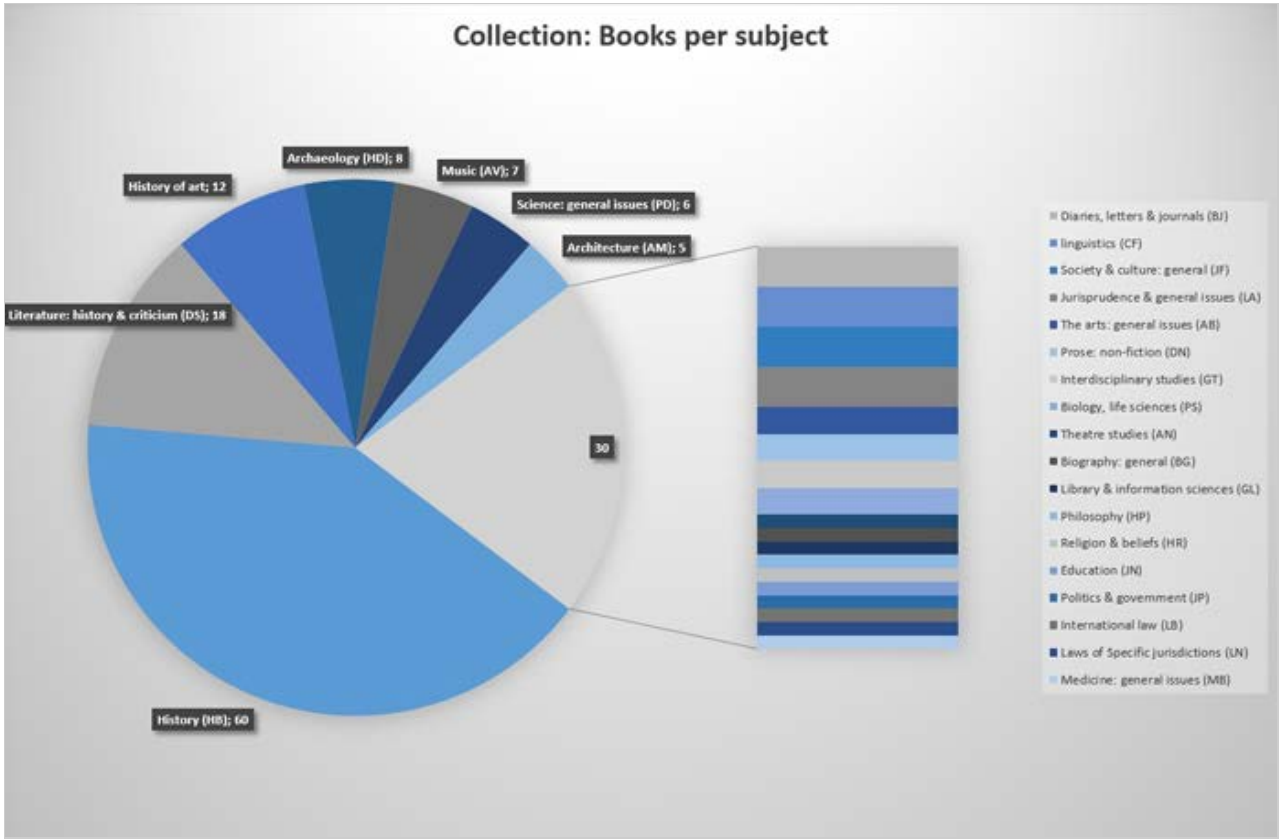|	
Archaeology (HD)	8
Music (AV)	7
Science: general issues (PD)	6
Architecture (AM)	5
Other subjects	30
Total	*146*

Table 1: Number of books per subject

Figure 1: Collection: Books per subject (View larger image).

The collection is mostly written in German; 126 titles, or 86% of the books. As will be described below, the strong emphasis of German affects the usage: most downloads originate from German speaking countries.
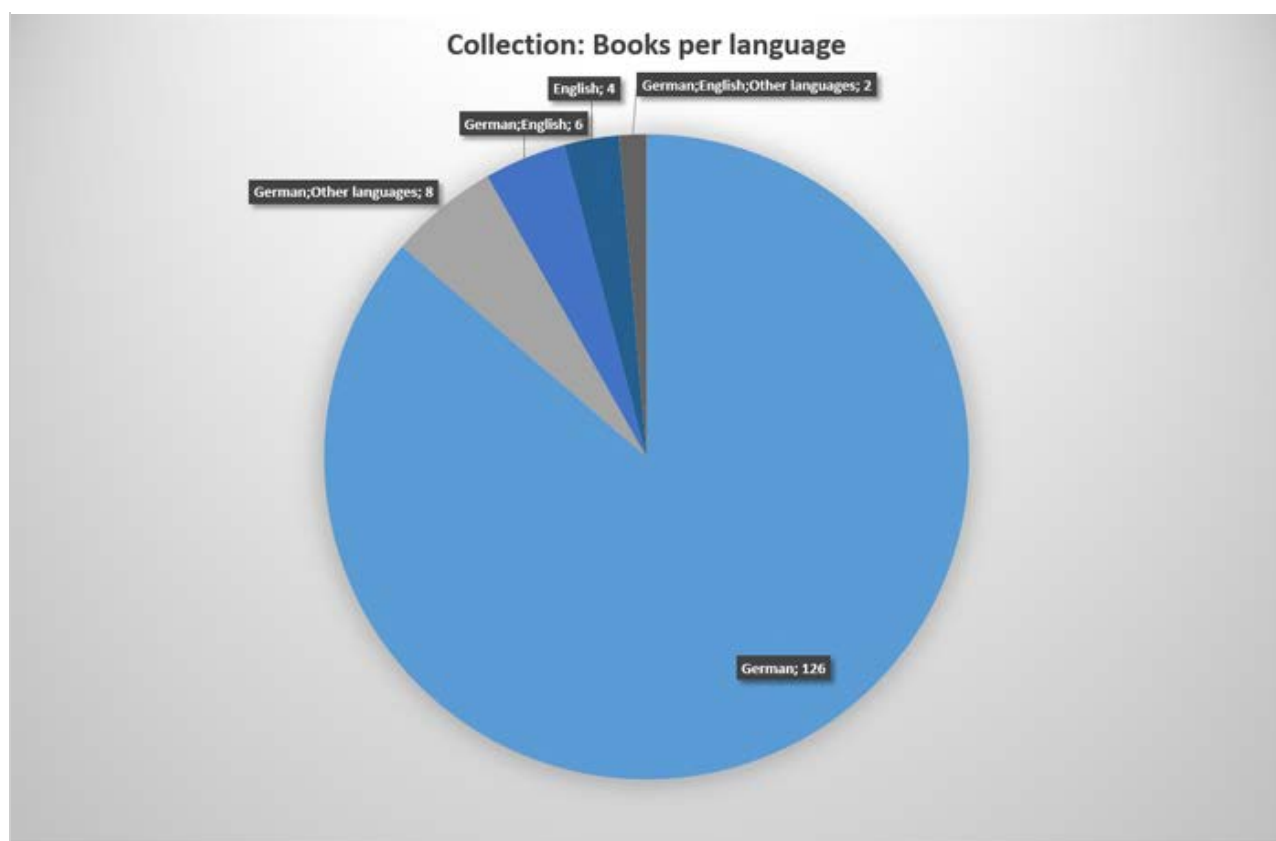
Figure 2: Collection: Books per language (View larger image).

The analysis is based on COUNTER compliant download data. The COUNTER initiative aims to facilitate the recording and reporting of online usage statistics in a consistent, credible and compatible way (COUNTER Online Metrics, 2014). This means that downloads by automated systems ('bots') and other types of suspicious download behaviour is discarded from the reports.

The data of the 28,139 downloads used for this analysis originated from 23,652 IP addresses. It is clear that many providers use several IP addresses: the IP addresses were linked to 2,839 provider names. Where no information about a provider could be found, the download numbers were omitted from the analysis. The omitted data amounted to 6% of the total: 1,955 downloads were not taken into consideration. The data used for this paper is available here.

---

2 German and the DACH countries

Most of the collection is written in German, and this can also be seen when the total number of downloads are charted. Of all downloads, 24,303—or 86%—were of a German language monograph.
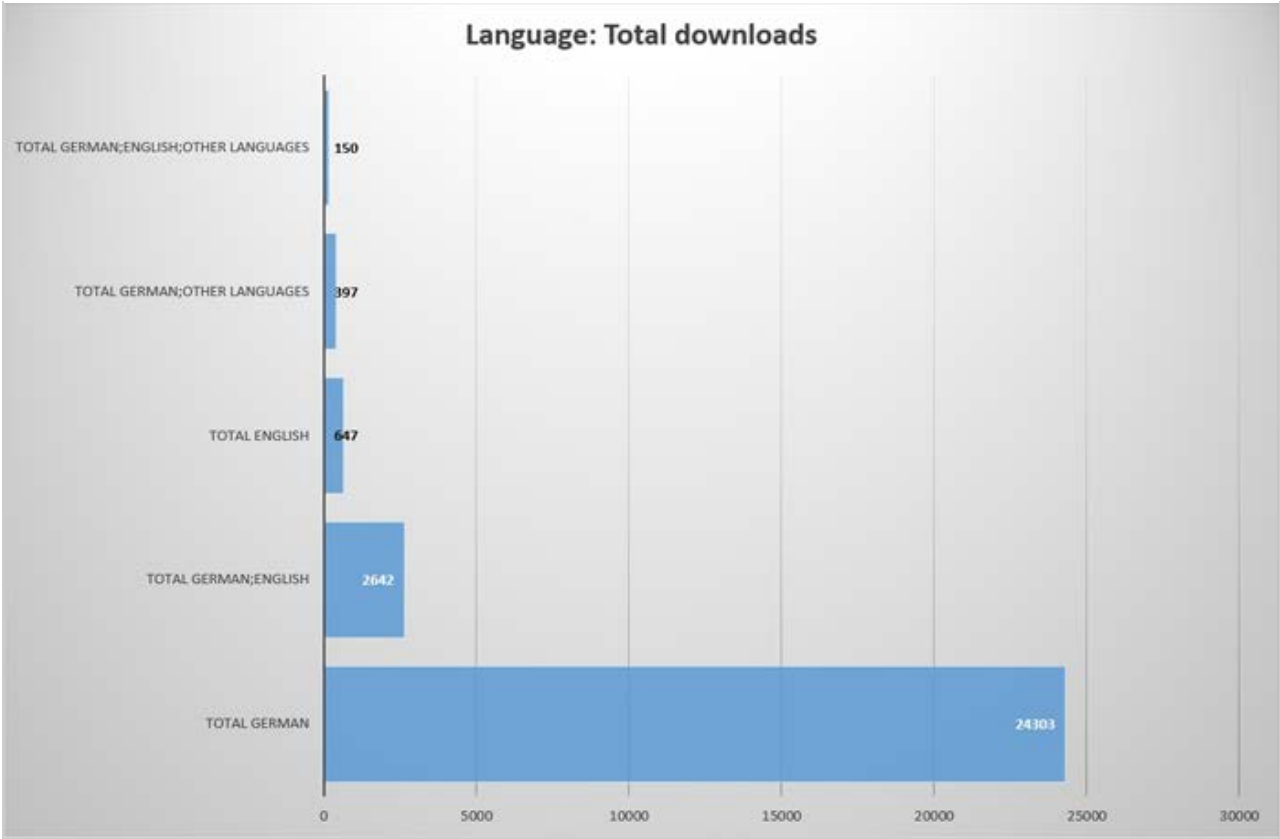
Figure 3: Language: Total downloads (View larger image).

This raises the question of the source of these downloads: from which country do they originate? As charted in *Figure 4: German: Total downloads*, it becomes clear that 55% of all downloads originate from the DACH countries (Germany, Austria, and Switzerland). Because of this, the analysis will use the distinction between the DACH countries and the rest of the world. The data is listed in *Table 2: German: Total downloads*.
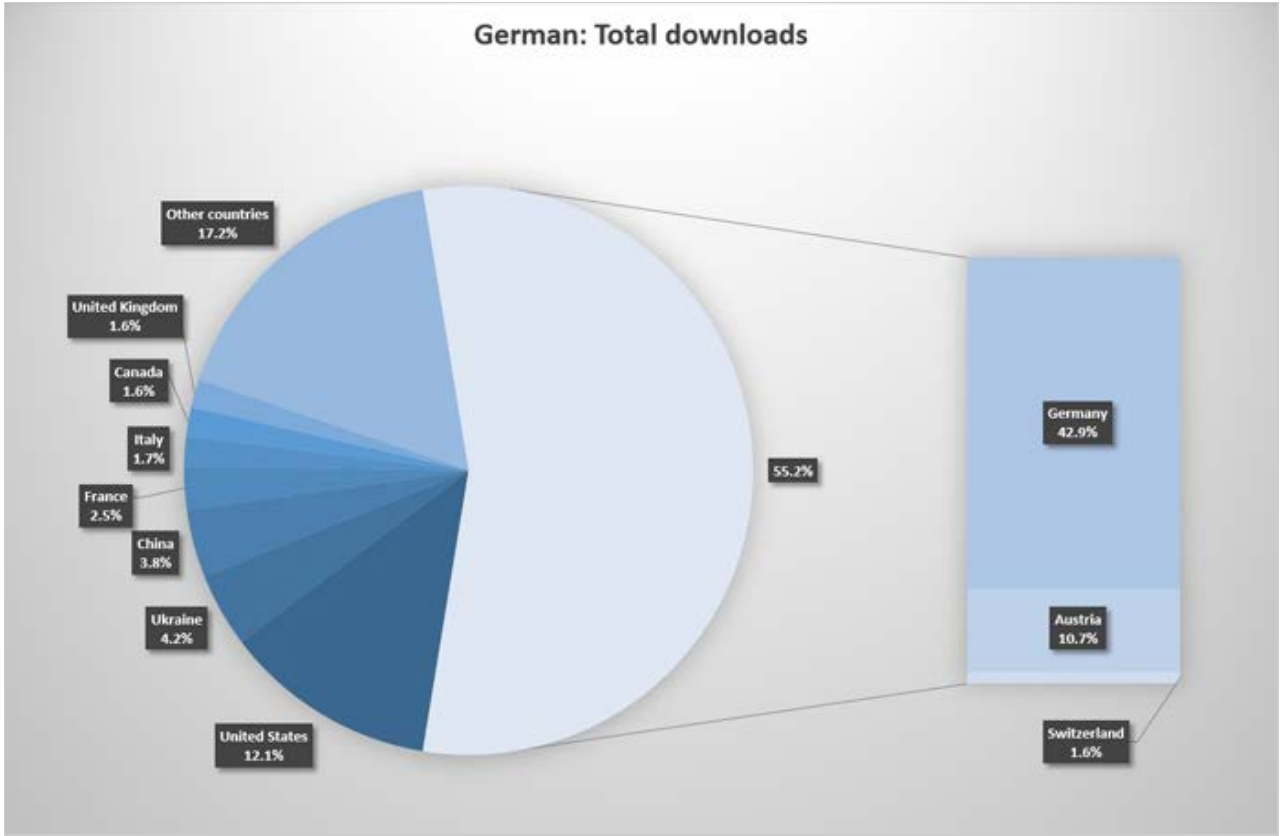

Figure 4: German: Total downloads (View larger image).

Country	German: Total downloads	Percentage
Germany	11,134	43%
United States	3,151	12%
Austria	2,785	11%
Ukraine	1,092	4%
China	1,000	4%
France	650	3%
Italy	433	2%
Canada	428	2%
United Kingdom	416	2%
Switzerland	410	2%
Other countries	4,480	17%
Total	*25,979*	*100%*

Table 2: German: Total downloads

3 Impact abroad: national vs. international usage

In the introduction, we discussed international usage as an estimation of esteem: is the work of the FWF funded researchers used beyond national borders? As can be seen in Figure 5, 11% of the total downloads is national, the rest is international. Of course, the role of Germany and—to a lesser extent—Switzerland is quite large, amounting to 42% of the total usage. The rest—approximately 47% of all downloads—comes from the rest of the world. The data is listed in *Table 3: Total downloads: Type of reader and region*.
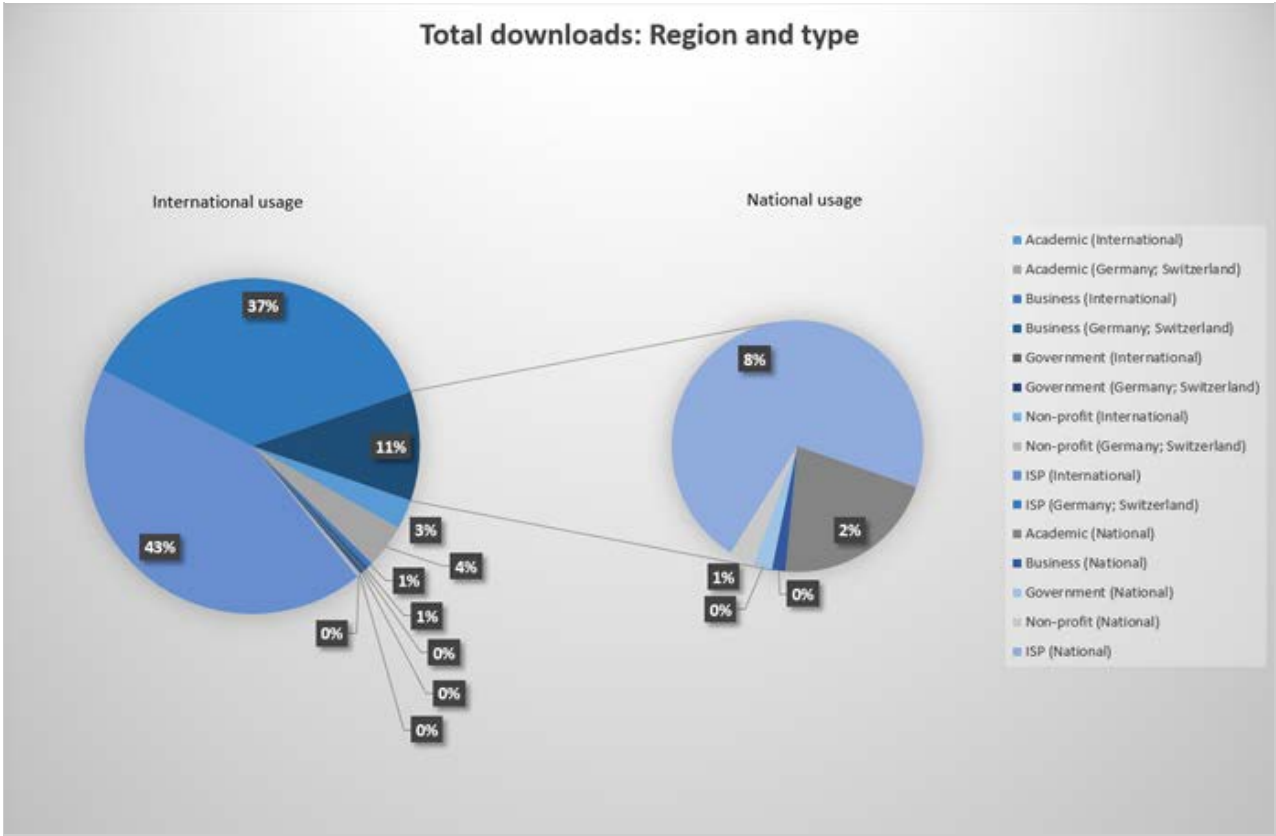
Figure 5: Total downloads: Region and type (View larger image).

The chart also depicts the differences per type of provider. Consistent with Snijder (2013), most of the downloads occur through an Internet Service Provider (ISP)—for instance Vodafone or Deutsche Telekom. Because ISPs function as a gateway for many different Internet users, it is harder to pinpoint the type of reader. However, Austria, Germany and Switzerland are countries with a highly developed Internet infrastructure, where organisations are more likely to 'directly' provide Internet access to their employees. This increases the likelihood that ISP usage originating from Austria, Germany or Switzerland are from people who do not act in an official capacity. In other words: there is a larger possibility that 'ISP users' from the DACH countries are from the 'general public'.

The second largest type of users are academic. Of all downloads, almost 10% originated from an academic institution. Based on this, we might conclude that the collection appeals to scholars. Again in conformance with Snijder (2013), the usage by governmental, business or non-profit organisations is relative low.

The download data—subdivided per user type—is listed in *Table 3: Total downloads: Type of reader and region*.

Region	Type of reader	Total downloads	Percentage
National	Academic	630	2%
	Business	52	0%
	Government	71	0%
	Non-profit	96	0%
	ISP	2,116	8%
Germany; Switzerland	Academic	1,187	4%

	Business	111	0%
	Government	70	0%
	Non-profit	75	0%
	ISP	10,484	37%
International	Academic	799	3%
	Business	146	1%
	Government	73	0%
	Non-profit	52	0%
	ISP	12,177	43%
Total		*28,139*	*100%*

Table 3: Total downloads: Type of reader and region

4 Language analysis

Before, the influence of language on the 'download region' has been discussed and also the usage of the different reader categories. In the following charts the mean downloads per language is shown. The chart in *Figure 6: Language: Mean downloads* describes the relative 'popularity' of the different languages. While the differences seem large, it must be noted that most groups—other than German—consist of less than ten books. On these small amounts, outliers have a large influence. All data can be found in *Table 4: Language: Mean downloads*.
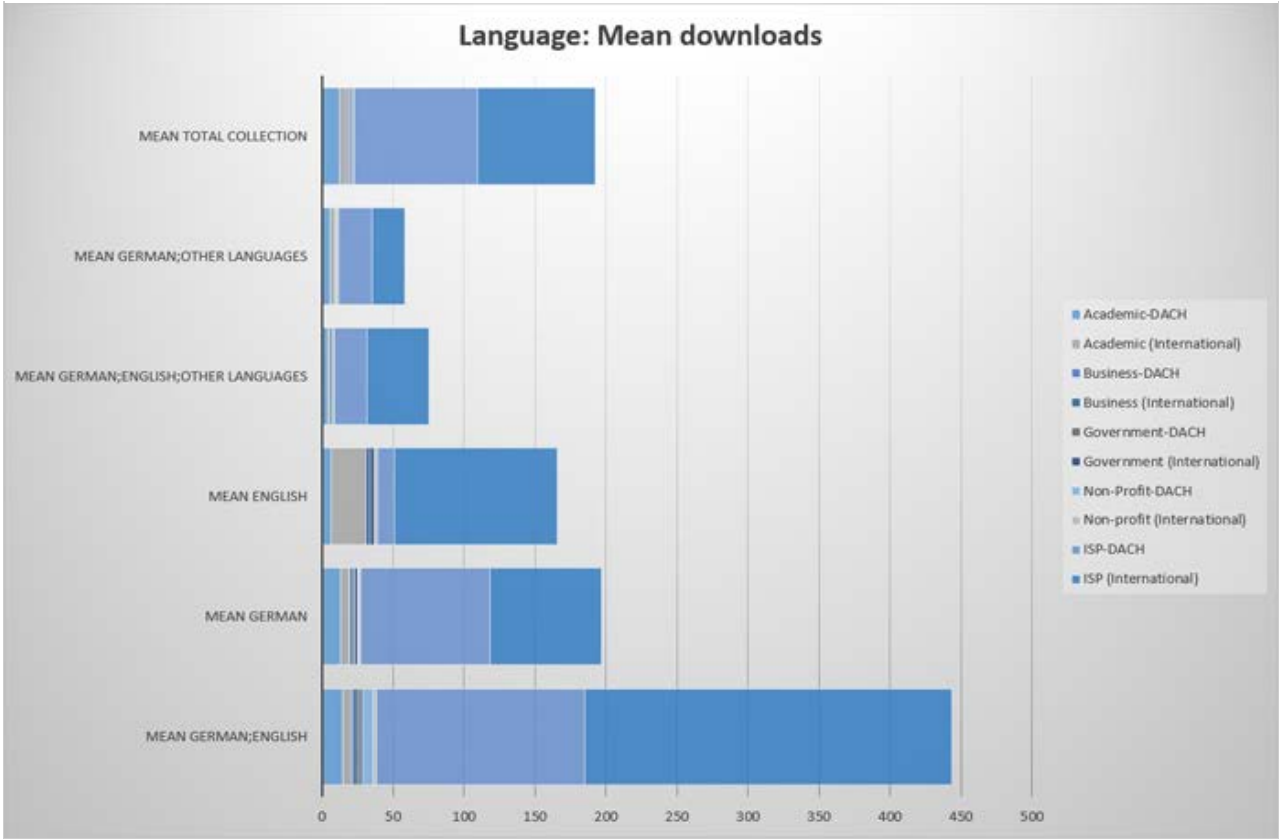
Figure 6: Language: Mean downloads (View larger image).

It is more interesting to look at the usage percentages of the different types of users, which is depicted in *Figure 7: Language: Mean downloads (percentage)*. Here, the percentages for English differ strongly from the rest: the largest portion of academic users and ISP users from countries other than Austria, Germany and Switzerland can be found here. This is another indication of the influence language has on dissemination: publishing in English enhances the usage beyond the DACH countries.
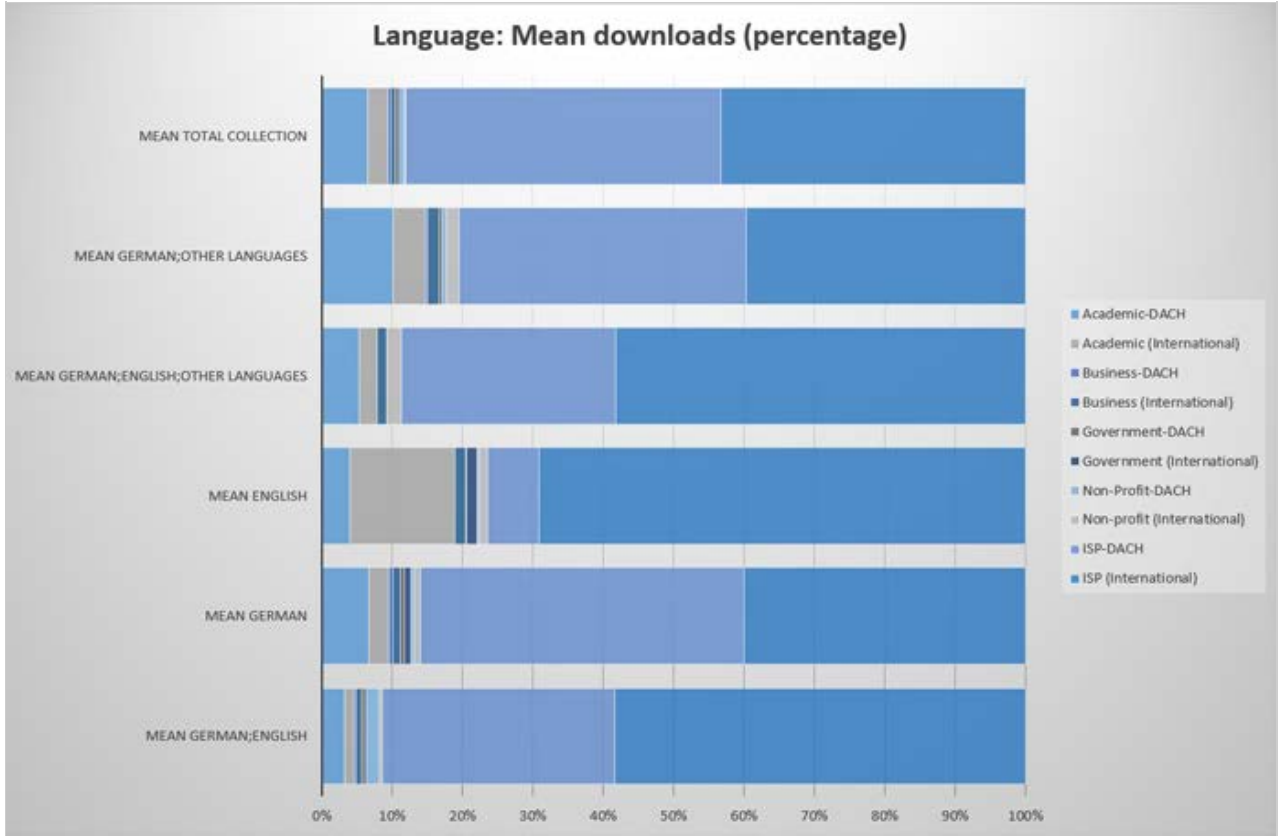

Figure 7: Language: Mean downloads (percentage) (View larger image).

The following table lists the mean downloads per language, plus the mean downloads for the complete collection.

Language	Mean downloads	Downloads	Number of books
Total collection	192.7	28,139	146
German; Other languages	49.6	397	8
German; English; Other Languages	75	150	2
English	161.8	647	4
German	192.9	24,303	126
German; English	440.3	2,642	6

Table 4: Language: Mean downloads

5 Subject analysis

This chapter contains the subject analysis. Section 1.1 describes the classification used, and how it is used to define broad subjects. While the collection contains books on many topics, it holds just a handful of subjects with seven or more books. The group of History books is quite large: 60 books. In contrast, the collection contains just eight Archaeology books and seven titles on Music.

Here we see large differences in the mean number of downloads per subject, where Archaeology in *Figure 8: Subject: Mean downloads* is relatively less 'popular' and the interest for Literature: history & criticism is the highest. However, most groups are quite small, and therefore the mean values are susceptible to outliers.
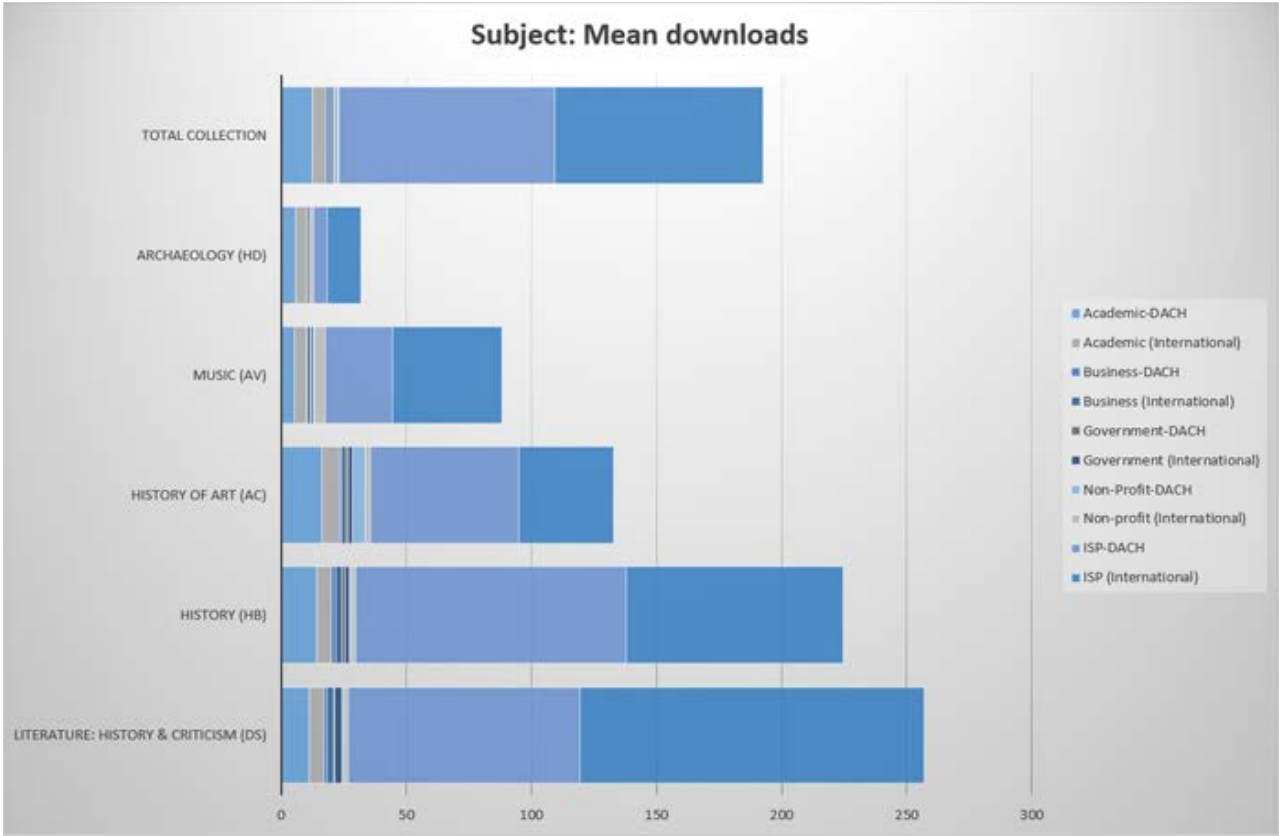

Figure 8: Subject: Mean downloads (View larger image).

When the percentages per subject are depicted in *Figure 9: Subject: Mean downloads (percentage)*, two of the subjects display a different pattern. If the percentages are taken into account, both Archaeology and History of art are downloaded more by academics.
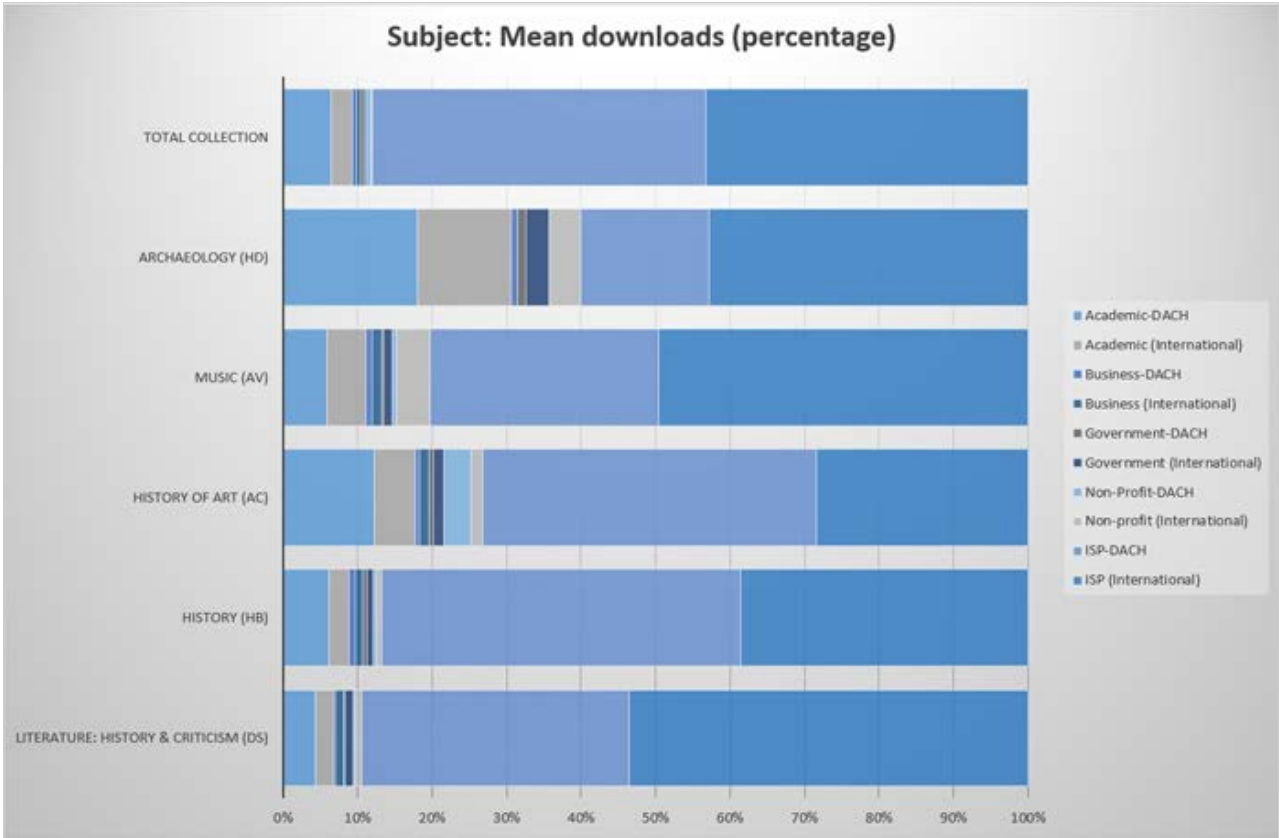

Figure 9: Subject: Mean downloads (percentage) (View larger image).

The data is listed in the table below.

Subject	Mean downloads	Downloads	Number of titles
Total collection	192.7	28,139	146
Archaeology (HD)	32.0	256	8
Music (AV)	87.7	614	7
History of art (AC)	135.4	1,625	12
History (HB)	235.5	14,127	60
Literature: history & criticism (DS)	254.2	4,576	18

Table 5: Subject: Mean downloads

6 Most downloads per provider, per type

Finally, the major users per provider type are listed in *Table 6: Biggest users, per type*, with the exception of ISPs. As ISPs are—per definition—providing Internet access to different users, and the number of downloaded books per ISP are much higher. However, it is not possible to know how many individuals or organisations are serviced by one ISP, which complicates further

analysis.

The relatively large uptake of academic institutions is clearly visible. While the University of Graz or the University of Cologne download 80 or 75 books respectively, the total number of monographs downloaded by the other categories of providers is much lower. In total, the data contains 899 different providers that are not ISPs.

Name	Type	Country	Total downloads
Universitaet Wien	Academic	Austria	295
Universitaet zu Koeln	Academic	Germany	80
Universitaet Graz	Academic	Austria	75
Humboldt-Universitaet zu Berlin	Academic	Germany	73
bartholomaeus.pro	Academic	Germany	68
Ebsco Industries, Inc.	Business	USA	15
PSI-AG, Geschaeftsbereich PS	Business	Germany	15
Intervog SARL	Business	France	14
Microsoft Corporation	Business	USA	11
Backlog Capital, LLC	Business	USA	10
gv.at	Government	Austria	39
Infrastructure datacenter	Government	Austria	13
Network of the Belgian Federal Government	Government	Belgium	9
Informatikzentrum	Government	Germany	8
Steiermaerkische Landesregierung	Government	Austria	7
Stiftung Preussischer Kulturbesitz	Non-profit	Germany	24
Hauptverband der oesterreichischen Sozialversicherungstraeger	Non-profit	Austria	23
Oesterreichische Nationalbibliothek	Non-profit	Austria	23
Kunsthistorisches Museum Wien	Non-profit	Austria	17

Bayerische Staatsbibliothek und Bibliotheksverbund Bayern	Non-profit	Germany	10

Table 6: Biggest users, per type

7 Conclusions

Using the methodology described in Snijder (2013) leads to several conclusions on the usage and the impact of the FWF collection in the OAPEN Library. Most of the usage is international; only 11% of the total downloads is national. The role of Germany and Switzerland is quite large, amounting to 42% of the total usage. The remaining 47% of the downloads originate from the rest of the world. Secondly, the influence of language is quite profound: books written in German are much more likely to be read within the DACH countries, while books written in English have a far greater chance to be used all over the globe.

Also, the role of academic readers is relative large, compared to governmental, business or non-profit usage. Yet, the biggest group of users have accessed the collection through an ISP. It is much harder to draw conclusions about their reasons to download: was it because of an 'official' role or did they act out of non-professional interest? However, a large group of 'ISP users' were based in Austria, Germany or Switzerland. These countries possess a highly developed Internet infrastructure, and this enhances the chance that these readers are members of the general public.

If the mean downloads per subject are analysed, we clearly see differences per subject: not all subjects enjoy the same amount of 'popularity'. Also, in the case of Archaeology and History of art, a relatively big usage by academics was measured.

This analysis helps to understand the impact of the books that have been made freely available by FWF. It is clear that the collection has a wider reach than academics, and has been read not only in the German-speaking countries, but world-wide.

Acknowledgements

References

[1] Book Industry Communication. (2010). BIC Standard Subject Categories—an Overview.

[2] COUNTER Online Metrics. (2014). COUNTER | About Us.

[3] Daigle, L. (2004). WHOIS Protocol Specification.

[4] Open Access Publishing in European Networks. (2010). About OAPEN—Open Access Publishing in European Networks.

[5] Phaidra-FWF-Der Wissenschaftsfonds. (n.d.).

[6] Snijder, R. (2013). Measuring monographs: A quantitative method to assess scientific impact and societal relevance. *First Monday*, 18(5). http://doi.org/10.5210/fm.v18i5.4250

[7] The Wold Bank. (2011). *The Little Data Book on Information and Communication Technology 2011*. Vasa. http://doi.org/10.1596/978-0-8213-9816-6

Appendix: Methodology

The method combines some aspects of the books—subject and language—with metadata of the users. Using web technology to make books available online enables us to collect usage data, such as the number of views or downloads and some information about the 'provider'—the organisation that grants access to the Internet—either the web address, or the IP address. The providers are categorised as 'Academic'; 'Business'; 'Government'; 'Non-profit' and 'Internet Service Provider (ISP)'. Furthermore, the country is also listed. Listing this data for each individual book enables us to draw conclusions on its usage in a certain period: what is the scholarly impact and the societal relevance?

Categorising users

The usage by academic institutions can be used as a proxy for scholarly impact: the total number of downloads; the different number of institutions; whether these institutions are foreign. International usage might be used as an indication of esteem. The percentage of usage outside national borders may give an indication of the importance of the work. This reflects on the authors; one of the 'esteem indicators' is the level of international interest.

The number of downloads by providers in the categories 'Business'; 'Government'; 'Non-profit' could be used as an indication of societal impact. However, most downloads will come from ISPs. Without further refinement, most of the usage is hard to categorize. The question is how to distinguish whether the usage comes from users whose organisation does not provide Internet access or from users who are downloading the monographs 'from home'. The latter group could be seen as the general public. The solution used in Snijder (2013) consists of assessing the Internet infrastructure per country, combined with the percentage of ISPs.

The Internet infrastructure differs from country to country. Presumably, in countries with a highly developed Internet infrastructure, most organisations are capable of directly providing Internet access to their employees. In countries with a weakly developed Internet infrastructure, access to the Internet will almost certainly be provided through an ISP. The World Bank publication *The Little Data Book on Information and Communication Technology* contains several indicators on the state of the IT infrastructure per country (The World Bank, 2011). Countries with 70 Internet users per 100 people or more are considered to possess a highly developed Internet infrastructure. This means that the chances are much higher that users that download books through an ISP are part of the general public.

Finding web addresses

On a more practical level, finding web addresses may be a challenge. The available usage data depends on the infrastructure used to disseminate the books on the web. A much used tool is Google Analytics, where the data can be found via the menu Audience/Technology/Network. In the case of OAPEN, the download data consists of the IP address plus the number of downloads.

For example, the download data of the book *Wien — Geschichte einer Stadt* in January 2015:

Proprietary Identifier	ISBN	Title	IP Address	Reporting Period Total	jan-15
437134	isbn:9783205992677	Wien—Geschichte einer Stadt	131.130.253.60	1	1
437134	isbn:9783205992677	Wien—Geschichte einer Stadt	131.130.87.251	1	1
437134	isbn:9783205992677	Wien—Geschichte einer Stadt	46.245.202.151	2	2
437134	isbn:9783205992677	Wien—Geschichte einer Stadt	77.80.43.171	1	1
437134	isbn:9783205992677	Wien—Geschichte einer Stadt	84.115.1.77	1	1
437134	isbn:9783205992677	Wien—Geschichte einer Stadt	93.128.253.108	1	1

The IP addresses need to be linked to a web address. Here the free lookup tool from xNode has been used: https://xnode.org/page/Bulk_IP_Lookup. (Note that in April 2015, this particular service was no longer available.)

The result below lists three addresses of the University of Vienna, two Austrian ISPs and a German ISP.

[131.130.253.60]

- Vienna, Austria
- univpn.univie.ac.at
- RIPE Network Coordination Centre / LAN University of Vienna / ACONET-LIR-MNT / RIPE-ERX-131-130-0-0

[131.130.87.251]

- Vienna, Austria
- ubop-87-251.ub.univie.ac.at
- RIPE Network Coordination Centre / LAN University of Vienna / ACONET-LIR-MNT / RIPE-ERX-131-130-0-0

[46.245.202.151]

- Austria
- static-46-245-202-151.d-light.at
- D-Light Dynamic Fiber-Customer Pool / MNT-D-LIGHT / D-LIGHT-NET

[77.80.43.171]

- Vienna, Austria
- user-043-171.vpn.univie.ac.at
- Vienna University Computer Center / RIPE-NCC-END-MNT / UNIVIE-II

[84.115.1.77]

- Wiener Neustadt, Austria
- chello084115001077.wrn.surfer.at
- UPC Austria / MNT-LGI / UPC

[93.128.253.108]

- Zella-mehlis, Germany
- x5d81fd6c.dyn.telefonica.de
- NCC#2008040527 / MDA-Z / TEDE-LLU

The data used for this paper contains 2,839 provider names and categories. See: www.persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-7p21-ay.

About the Author

Ronald Snijder has been involved in OAPEN since 2008, first as an employee of Amsterdam University Press (AUP), and now as technical coordinator at the OAPEN Foundation. There he is responsible for the technical development of the OAPEN Library. Before that, he worked in several profit and not-for-profit organizations as an IT and information management specialist. He is currently working on his PhD dissertation on the societal and scholarly impact of Open Access monograph publishing at Leiden University. (E-mail: r.snijder@oapen.org; Twitter: @ronaldsnijder)