Correlation of impact measures of institutional repositories and PBRF ranking

by

Rui Li

Submitted to the School of Information Management,

Victoria University of Wellington

In partial fulfillment of the requirements for the degree of

Master of Information Studies

2011

Acknowledgements

My thanks, first of all, to Alastair Smith for supervising this study, for his helpful guidance along the way, and all the support during the years of my study.

I would like to thank my parents and my wife for their understanding and support during the years, without their support the study will not be possible.

Finally, I would like to thank all the people who have helped me during my research and study.

Abstract

This study examines the correlation of website impact factor of institutional repositories (IR) of all eight universities in New Zealand and the Performance Based Research Fund (PBRF) quality score. The purpose of the research is to find out whether there is correlation between these two figures. The research problems are: the correlation between the IR homepage ranking and the PBRF quality score, the correlation between the IR website inlinks and the PBRF quality score, and the correlation between the IR website impact factor and the PBRF quality score. The research also studied the different web ranking tools and tried to find out whether these tools can be used to measure the quality of IR documents. The research used Yahoo Site Explorer to collect information of inlinks and also use other tools to collect the webpage ranking. The finding of this research are that there is small correlation between the IR website impact factor and PBRF quality score, and the page ranking is not a good tool to exam the quality of IR document as a whole.

Key words: impact factor, webpage rank, institutional repositories, PBRF, inlink

Table of Contents

Introduction

It has been 10 years since the first institutional repository (IR) appeared. During the last decade, IRs have been widely established around the world. For many IRs, people tend to pay their attention to some of the IR impacts which are a tool to help increasing the visibility of academic output because of the open access feature of many IRs, or as a tool to archive the academic output in digital format, in other words a place for preservation. More and more studies have been carried on in the IR field. However, studies to date have tended to focus either on technology - how to build up an IR or on the self-belief – in theory how useful people think their IR is.

Little data has been collected thus far that examines the IR movement as a whole to see whether the overall quality of academic output is reflected to the website ranking, quality or the number of documents of IR website. One of the difficulties of comparing the quality of the documents in the institutional repository is that it is difficult to find some academic output standard to compare with. Fortunately, in New Zealand there is a quality ranking which was taking place since 2003. This research applied the ranking as a comparison figure with the website impact factor of IRs of New Zealand Universities.

1.1 Purpose

In this project, the web ranking of each IR's home page, total number of inlinks for each IR, and the web impact factor of each IR were compared with the quality ranking

that comes out of the Performance Based Research Fund (PBRF) process (Tertiary Education Commission, 2009). PBRF is first introduced in 2003, the second partial round was held in 2006, and the next full round will be held in 2012. The primary purpose of the PBRF is to ensure that excellent research in the tertiary education sector is encouraged and rewarded. The purpose of this project is to test whether there is correlation of the website impact factor of open access institutional repository compare with the PBRF quality score of the universities in New Zealand.

1.2 Design/methodology/approach

The project used online Yahoo Site Explorer to gain the information of inlinks for each IRs and also for individual document record pages so that the website impact factor can be generated. The total number of documents in each IR, this information can be found on many different websites, in this research the Kiwi Research Information Service website - http://nzresearch.org.nz/ was used to collect such information. Once all the information is collected, the project calculated the website impact factor and compare the website impact factor, the rankings of the website and the total number of the inlinks of each IR with the quality score which is produced from the process of the Performance Based Research Fund (PBRF).

1.3 Definition of terms

Institutional repository - it is the digital service in the university where the university library in most of the cases, but sometimes can be faculty or other department in the

university, archives their scholarly output. It includes things such as research papers, PhD theses, master theses, or articles published by the university's academic staff and students. There are different types of IR. Some of them are open access, and others are not. For the purposes of this project, the term "institutional repository" was mainly referred to the open access site. If there are needs to address those "dark IR", which is not public accessible, in this project, special notes will be put into place.

Performance Based Research Fund - is a New Zealand tertiary education funding process, assessing the research performance of tertiary education organisations and then funding them on the basis of their performance.

Website ranking – there are online websites which are holding ranked lists of different websites, generally related by an overall subject. In many cases such websites are directory of related websites which ranks the listed sites by popularity. In this project the tools for generate the ranking was Google Rank. However, other software may also be included in this project if there are needs for more depth analysis.

Website impact factor – it is a tool to measure the impact of websites by the number of links it receives. It can be measured by the ratio of links made to a website to the number of pages at the website (Jalal, Biswas & Mukhopadhyay, 2010, p109). There are different types of website impact factors. In this study it was calculated as:

A = inlinks from other website, B = total documents of the IR website.

Website impact factor = A/B.

Inlink in this project means those websites rather than the IR itself, which have a link or links to either to the IR homepage or to the individual pages in the same IR. In the research, the inlinks were collected by using Yahoo Site Explorer.

The Problem Being Investigated

2.1 Project rationale and significance

By October 2008, there were more than 10000 IRs around the world, containing nearly 200,000 items (Conway, 2008). In 2011 this number has increased to more than 32,000,000 items which is calculated according to the information on ROAR website (http://roar.eprints.org/). However, the real number of items in IRs will be much bigger than this, as there are many IRs where the number of items is not available from the website. In addition to this, there are many "dark" IRs (they are not the focus of this research as they can not be externally linked to) around the world which are not open to the public. In many cases people do not even know they exist. As a result, it is impossible to account the number of items in them. In New Zealand all the universities have their IR. Their names may vary but they all provide the similar services. Although, all the universities in New Zealand have IR, there are hardly any studies which are seeking the correlation between the IR website citation and the PBRF quality score.

The PBRF is an essential part of the universities' funding, and it is based on the output of academic research. One of the purposes of IR in theory is a tool which will represent the quality of academic output. This research is aiming to find out whether the IR homepage ranking, IR website impact factor, and the number of inlinks of IR

are reflecting to the quality of academic output. The finding of the research will be useful for the study of IR and the use and the promotion of the IR services in universities in the future.

2.2 Problem statement

The citation rates of academic output of open access IR have not been well represented in the literature thus far. Where research has been done, it has tended to focus on how to build an IR – technology and methods used, and how good/useful their IR will be.

There are many studies which are only focus on the technology which is used to develop IR. This led Shreeves (2008) to believe IRs were isolated from other library services. Up to date most of the IR studies trend to only focusing on the beautiful side of IR, such as how many items are collected or what IR can do in theory, and people do not want to touch the hard part – everyday work, the services, the usefulness of the service in practice and the quality of the IR. Especially for the quality of the IR, there are hardly any studies have been done in New Zealand. Studies are urgently needed in the areas above.

2.3 Research questions and hypothesis

The hypothesis of this project is the correlation between IR homepage ranking, website impact factor and number of inlinks, and the universities' PBRF quality score, it is expected that data gathered from the project will show whether there is

correlation or not.

The project's research questions are as follows:

- Is there correlation between the IR homepage ranking and the PBRF quality score?

- Is there correlation between the inlinks of an IR and the PBRF quality score?

- Is there correlation between the IR website impact factor and the PBRF quality score? Is this result close to the result of comparing the inlinks of an IR and the PBRF quality score?

- Are there software applications that produce more meaningful and usable rankings than the homepage ranking from Google Page Rank? What is the reason?

Literature Review & Theoretical Framework

Prior to the development of the project concept an integrative literature review was undertaken to determine what research had already been completed in the IR in an academic context.

3.1 Literature review

The Institutional Repository has been developed all around the world during the last decade. People argue about the success of IR. Some think IR has achieved more than they need and some believed it has not met the service level. This review looked at what studies have been done, what can be learned from the studies and what the gaps are.

3.1.1 What is IR?

An institutional repository (IR) is "a set of services and technologies that provide the means to collect, manage, provide access to, disseminate, and preserve digital materials produced at an institution" (Shreeves, 2008). It is a more convenient and more reliable place to hold the output of scholarly communication (Conway, 2008). IRs first appeared around 2000. In 2005, 40 percent of the institutions which offered doctoral degrees in the US had an IR (Shreeves, 2008). The content of IR is growing in an incredible speed, by October 2008, there were more than 10000 IRs around the

world, containing nearly 200,000 items (Conway, 2008), and as discussed in earlier section the number is more than 3 millions in 2011.

Shreeves (2008) believes that the goals and purpose of IRs can vary. IRs can be seen as a result of the serials crisis, but they can also be seen as showcasing an institution's research and scholarship. In Allard, Mack and Feltner-Reichert's (2005) study, 90 percent of the IR literature talked about the definition of IRs, but only one-fourth of them saw IRs as institutionally centred and scholarly. The rest of the definitions varied. This gives a clear message - IRs have different meanings to different people.

The different opinions in defining IR and the different goals and purpose of IRs are crucial factors for people who are studying them. People tend to follow trends. This means when everyone else has an IR, people are following the trend to build their own IR. Especially, when governments started to support it - "In 2004, the Science and Technology Committee of the UK Parliament issued a report recommending that publicly-funded institutions should establish institutional repositories on which their published output can be read free-of-charge online (Chan, Kwok & Yip, 2005)". IRs appear without much thought, without consideration of why an IR is needed, or how it can benefit the institution. When an IR is ready to be used, people simply let it go live.

What is the content of an IR? According to Yeates (2003), IRs cover the following content: pre-prints, other works-in-process, peer-reviewed articles, monographs, enduring research material, data sets, other ancillary research material, conference papers, electronic theses and dissertations, and grey literature. Among the content there is about 13 percent that is already published as peer-reviewed articles (Shreeves, 2008). In Piorun and Palmer's study (2008), there is 41.5 percent of students' work on the IR of Lamar Soutter Library, which is the largest percentage of all contents. These two numbers give people a brief idea of the nature of IRs. It well fits into Chang's (2003) description of IR features: institutionally defined, which means an IR should focus on collecting its own products – scholarly items are created by their staff and students; scholarly content – everything on IRs should only for research and teaching purposes; cumulative and perpetual – means items on IRs should always be kept there. Graham (2005) gave a slightly different definition of IR core features, "digital content, community-driven and focused, and institutionally supported, durable and permanent, accessible content."

3.1.2 Services of IRs

There are two different types of IR users: internal users – faculty, academic staff, or students; and external users – scholars from outside of the institution who is using IRs as information resources. Allard (2005) found that "it [study] still focused on how the IR was established and put into practice, and referred only briefly to activities

associated with assessing the success of its operation". He believed the reason is that IRs were in an evolutionary phase and "this gap in the literature makes it more difficult for librarians to learn how to help an IR initiative succeed".

Internal users are much better researched than outside users, but there is still not enough research done here – "Much has been written about technical aspects of developing a repository, less has been written about the relationships of reference librarians with faculty members and students to encourage them to share their research (Rockman, 2005)". However, there are still some studies on how reference and subject librarians provide services to academic staff. IRs have changed the role of the traditional librarian, and they (librarians) must see themselves as "knowledge managers" (Phillips 2005). The main goal for librarians is to get as much academic staff research output as possible.

3.1.3 Website Impact Factor and Institutional Repository

Website impact factor is measure which is developed from the journal impact factor. It is a way of comparing the impact of websites. The calculation of website impact factor, as discussed earlier, is the total number of inlinks form outside of the website divided to the total number of items on the same website. The two elements of this formula are the total items and the inlinks. The first one is easy to understand, but the second is not very simple. There are many different types of inlinks. Smith (2009) put the inlinks from general website to IR into 12 different groups. They are:

- Formal research impact - 0.74%
- Research or technical report - 0.74%
- Online magazine article - 0.37%
- Informal research impact - 1.78%
- Blog - 7.32%
- Wikipedia or other online reference resource - 7.54%
- Page using image from IR - 1.55%
- Self Publicity - 3.77%
- General navigation, directories etc - 1.04%
- Subject specific navigation - 48.74%
- Directory of IRs - 15.01%
- Links from a virtual repository - 11.39%

The list shows that directory and navigation play very important roles in among all the external inlinks to an IR. However, informal research links also have valuable weight. Informal research, Blog, Wikipedia, page using image from IR, and self publicity together they are 21.96 % out of all kinds of inlinks. This clearly shows that apart from staff of IR, individuals are playing very important promoting role in IR services. Even they may not realise how important they are for IRs but their contribution should not be ignored. Unfortunately, there are not many studies which are focusing on them.

3.1.4 Findings from the literature

It has been 10 years since IRs were first introduced. However, there are many challenges which IRs are facing. McDowell (as cited in Dorothea, 2008, p99) argued that "IRs have not fulfilled their early promise of increased access to the scholarly journal literature through faculty initiative". The reasons of this are varied. First, self-archiving is still just a beautiful dream. Faculty do not like it because of self-archiving may threaten their rights over their work, or damage the relationship between them and their favourite publishers. In addition, self-archiving also put extra work on their shoulder and they do not seen benefits from it; researchers do not like self-archiving, because there is no quality control; librarians do not like it, because they have to clean up the mess in database all the time (Dorothea, 2008). This has resulted in low self-archiving rates (Shreeves, 2008). Second, apart from self-archiving, faculty still stayed away from IRs because IRs offered nothing they valued. This leads to the research question of this project – does quality of IR content reflect to the quality of the academic output?

In addition, there was hardly any research which focused on external users. Does this mean external users have no value to the host institution of an IR? The answer is no: the external users are very important parts of IRs and their future survival. As discussed in the last section, individuals are playing very important role in today's IR world. The reasons why IR needs external users are: for users, knowledge can be shared; for institutions, the quality of their intellectual capital can be highlighted

(Yeates, 2003). To put it simply, IRs can increase the global visibility of institutional scholarship (Buehler, 2005), and the more accessible research output is, the more it is cited (Organ, 2005). This is the main reason why academic staff want to put their research on an IR. The goal of this project is to find out how the quality of the IR websites reflects to the quality of the institution as a whole.

3.2 Theoretical framework

There are not many things we can call "a theory" in IR studies. The reason is: first, people are still not clear about IR in many aspects, e.g. there are many different definitions of IR; secondly, most of IR studies are case studies. Most, if not all studies were focusing on how to solve specific problem rather than general theories. However, this does not mean we can not use any theoretical framework in IR studies. The possible theories we can draw from the early study, according to Maxwell's (2005, p. 42) definition, was the role of IR in the academic libraries. There are two different roles:

- IR is a tool of preservation.
- IR is a tool of self publishing.

This project used the second – IR is a tool of self publishing, which is what the IRs are used in the New Zealand universities in most of the case. There is one thing which needs to be noted here that self publishing does not only mean people publish their research by themselves, but also means the universities try to publish the academic output which are produced by their staff and students.

Research Methods

4.1 Methodology

There are two main methodologies: quantitative and qualititative. This research used the quantitative methodology. Blaxter (2006, p. 64) defined them as "quantitative research is empirical research where the data are in the form of numbers. Qualitative research is empirical research where the data are not in the form of numbers". In addition to this, Blaxter pointed that "quantitative research tends to involve relatively large-scale and representative sets of data, …, qualititative research, on the other hand, is concerned with collecting and analysing information in as many forms, chiefly non-numeric, as possible, … aims to archive depth rather than breadth". This project focused on analysing data which was available to public from the Internet, it was heavily dependent on deskwork and analysing of numbers.

4.2 Research Methods

The project was research of IRs in all 8 universities in New Zealand.

- Auckland University of Technology – ScholarlyCommons@AUT, http://aut.researchgateway.ac.nz/ and http://reposityaut.lconz.ac.nz
- Lincoln University – Lincoln U Research Archive, http://researcharchive.lincoln.ac.nz/dspace/
- Massey University – Massey Research Online, http://muir.massey.ac.nz/

- The University of Auckland – ResearchSpace@Auckland, http://researchspace.auckland.ac.nz/
- University of Canterbury – UC Research Repository, http://ir.canterbury.ac.nz/
- University of Otago - Otago University Research Archive, http://ourarchive.otago.ac.nz/ , the School of Business, http://eprints.otago.ac.nz/ , and Te Tumu, http://eprintstetumu.otago.ac.nz/
- University of Waikato – Research Commons, http://researchcommons.waikato.ac.nz/
- Victoria University of Wellington – ResearchArchive@Victoria, http://researcharchive.vuw.ac.nz/

The research firstly generated rankings as of all IR homepage by using online website ranking tool Google Rank. The reason of choose Google Rank is that many of the online website ranking tool are in fact drawing on the same data from Google Rank. One example is that Page Rank checker, as listed on its website (www.prchecker.info), Page Rank checker is a free service to check Google page rank. The research also used website ranking software, such as, Accurate Monitor for Search Engines, Advanced Web Ranking 7.5, iBusinessPromoter 11, Web CEO, AgentWebRanking to rebate rankings. All of these tools are designed for generating popularity of a webpage. The more external websites cite a webpage, the higher rank the webpage will have.

The second step is to finding the total number of inlinks which from external website

for each IR by using the Yahoo Site Explorer (YSE). Yahoo Site Explorer "allows you to explore all the web pages indexed by Yahoo Search. View the most popular pages from any site, dive into a comprehensive site map, and find pages that link to that site or any page" (Yahoo.com, 2011). In this research, the YSE was used to collect the number and URL of inlinks to the universities' IR website. The project then moved to generate data for website impact factor by using the number of inlinks which is collected above, and the number of total documents in each IR. In this process, website impact factors for each of the IR website were produced. The website impact factor is using the formula - website impact factor = inlinks from other website / total documents of the IR website. The results of comparing the number of the total inlinks with PBRF quality score and website impact factor with PBRF quality score were compared to see which the best way is to measure the quality of the IR content. The data of PBRF quality score was collected from New Zealand Tertiary Education Commission website, http://www.tec.govt.nz/. The last part in this research is to compare the web page ranking from Google Page Rank and those from software to see whether there are different between these two kinds of tools and whether they are possible tools to measure the quality of the documents in IR.

Data Analysis Techniques

5.1 IR homepage Ranking VS PBRF Quality Score

The IR homepage ranking was compared with the PBRF quality score to find out the correlation. A figure was produced to show the result. The figure used in Smith and Thelwall's (2005) research was used as a model for this research. The different between the figures which were used in the model and this research is that, in this research the X axis is the PBRF quality score for each universities and the Y axis is the IR homepage ranking rather than the X axis is Domain inlinks/FTE and Y axis is the PBRF quality score. The reason of this is to find out whether there is correlation between the IR homepage ranking and the quality of academic output.

5.2 Inlinks VS PBRF Quality Score

The number of inlinks for each IR was compared with the PBRF quality score. A figure was also being draw. This time the Y axis is the number of total links.

5.3 Website Impact Factor VS PBRF Quality Score

There will be two different website impact factors for this research question - normal website impact factor and average website impact factor. To collect data for normal website impact factor, a sample of 1000 inlinks were collected by using Yahoo Site Explorer. If the inlinks are less than 1000, all of them were collected. If the inlinks are over 1000, the first 1000 inlinks were collected (an estimated number of total inlinks

to individual documents were calculated based on these first 1000 inlinks, this will discussed in details later in this report). Then each inlink was manually checked to see whether the inlinks from the website was linked to the IR homepage or to an individual document in IR. Website impact factor were produced based on the information collected here. The website impact factor was then compared with PBRF quality score and a figure was produced. In addition, 100 document pages were picked randomly as sample to calculate the average inlinks which were linked to individual documents. The purpose of this is to test whether average number can be used to measure the website impact factor.

The last stage of this part is to compare the result from inlinks vs. PBRF quality score and the result from normal website impact factor vs. PBRF quality score. This showed which way is better when analysing IR website quality. Figures were produced to show the results and differences.

5.4 Analysing Tool comparing

Comparing the result of rankings of IRs, which are generated by using Google Page Rank or software, to see whether there is a more meaningful one. If none of them are useful then find out the possible reason and if it is possible to use website ranking to analysing the quality of IR website. The finding will benefit the future IR study as if there is a good tool, it will save time for people who wants to do IR study in the future. If there is none, it will also stop people wasting time on similar research.

Research process, Findings & Limitations

This section discusses the research process, findings and limitations for each of the research questions.

6.1 Question one: Is there correlation between the IR homepage ranking and the PBRF quality score?

The 2006 PBRF Quality Score according to the New Zealand Tertiary Education Commission (2010) for the eight universities in New Zealand are:

University	PBRF Quality Score
Auckland University of Technology (AUT)	1.8
Lincoln University (LU)	2.94
Massey University (MU)	3.05
University of Auckland (AU)	4.09
University of Canterbury (CU)	4.07
University of Otago (OU)	4.17
University of Waikato (WU)	3.74
Victoria University of Wellington (VU)	3.83

Table 1

The webpage ranking tool is chose for this research question is the Google Page Rank (for the research as a whole, other tools were also used, such as YSE). Although, how the page rank works is not the focus of this research, a brief explanation will help people to understand the find. A page rank, according to Craven (2011), is "a numeric value that represents how important a page is on the web. Google figures that when one page links to another page, it is effectively casting a vote for the other page. The more votes that are cast for a page, the more important the page must be. Also, the importance of the page that is casting the vote determines how important the vote itself is. Google calculates a page's importance from the votes cast for it. How important each vote is taken into account when a page's PageRank is calculated. PageRank is Google's way of deciding a page's importance. It matters because it is one of the factors that determine a page's ranking in the search results. It isn't the only factor that Google uses to rank pages, but it is an important one."

The IR homepage rank of the eight universities are:

University	IR homepage ranking
AUT	6
LU	5
MU	6
AU	6
CU	6
OU	5.7 = (6 + 5 + 6) /3
WU	6
VU	6

Table 2

In the above list there is a university – Otago University, its rank is the average number of 3 webpage ranks. The reason for this is there are three repositories there at Otgao, Otago University Research Archive, the School of Business and Te Tumu.

It is clear that when compare the PBRF quality score and the page rank, these two data are not correlate to each other. In addition, the 6 of the 8 universities which have the same IR homepage ranking, this is not a useful data to measure the order of the ranks. As a result, a conclusion can be made that the webpage ranking is not a good measure for the study of IRs and their PBRF quality score.

The main limitation for this question is that all the IRs have fairly good page ranks - page rank isn't a very discriminating measure at this level. Other limitations included that exactly how Google calculated the page rank is unknown and there are no possible ways to find out as this is a commercial secret. Addition to this, the page rank information is collected by machine, these means there must be some machinery errors, such as people using software or other tools to fool the machine so that they could have a higher page rank.

6.2 Question two: Is there correlation between the inlinks of an IR and the PBRF quality score?

The reason of why inlink is important to website as it is discussed by Craven (2011), the more inlinks to a page, the higher rank it will be on the page rank. To find out how many inlinks are there for the eight repositories, a web tool has been used in this research – Yahoo Site Explorer. Yahoo Site Explore allows people to calculate how many inlinks for a single website. What people need to do to achieve the information of the inlinks which is copy and paste the URL they need on the website http://siteexplorer.search.yahoo.com/, the web tool will generate the results automatically.

The total number of inlinks (both internal and external) of the eight universities' IR website which is generated by Yahoo Site Explore (YSE) is:

University	Total inlinks
AUT	2419
LU	4201
MU	4980
AU	27010
CU	7813
OU	9378 -total of 3 IRs
WU	6096
VU	15259

Table 3

When put this numbers of links in order AU – VU – OU – CU – WU – MU – LU – AUT and compare with the PBRF quality score (see table 4), the result is quite close to each other (apart from AU is the first and VU is the second, these two websites have much more inlinks within their own website compare with the rest. This will be discussed later in this research). This result leads to a supposition of that there may be some correlation between the website inlinks and PBRF quality score. This can be a follow up research question when the result of the next round PBRF quality score is available. If by that time the results are still close to each other, then the order of total

inlinks of an IR website in New Zealand can be seen as the correlation of the universities' PBRF quality score.

PBRF quality score (high to low)	Total inlinks (high to low)
OU	AU
AU	VU
CU	OU
VU	CU
WU	WU
MU	MU
LU	LU
AUT	AUT

Table 4

As noted above, the total inlinks of Victoria University and the University of Auckland are much higher than the rest. The possible reason for this fact is because both IRs have many more internal inlinks within their own website. Internal inlink means the inlink to an individual document which is from a webpage in the same IR. External inlink means the inlink to an individual document which is from webpage outside of the same IR. The assumption is based on the fact which is found during the process of the research question two. While collecting data for the research question two, an inlink checking for 100 random documents for each IR has been carried on.

The finding shows that, most of the sample pages in the University of Auckland's IR which have more than 10 internal inlinks from its own website. Same thing happened to the Victoria University's IR, too. The documents at Victoria University's IR have even more internal inlinks than Auckland. There are only about 5 documents out of 100 samples which have less than 10 internal inlinks. Many of samples have more than 20 internal inlinks and there is one documents even has 44 internal inlinks. On the other hand, the average internal inlinks for the rest of universities' IRs are less than 10. For this reason, it is possible to assume that if Victoria University and the University of Auckland could keep their internal inlinks on the average level the order of the total inlinks for the eight universities should correlate with the PBRF quality score. This means measure the total inlinks of an IR website can be a possible way to value the quality of its academic output.

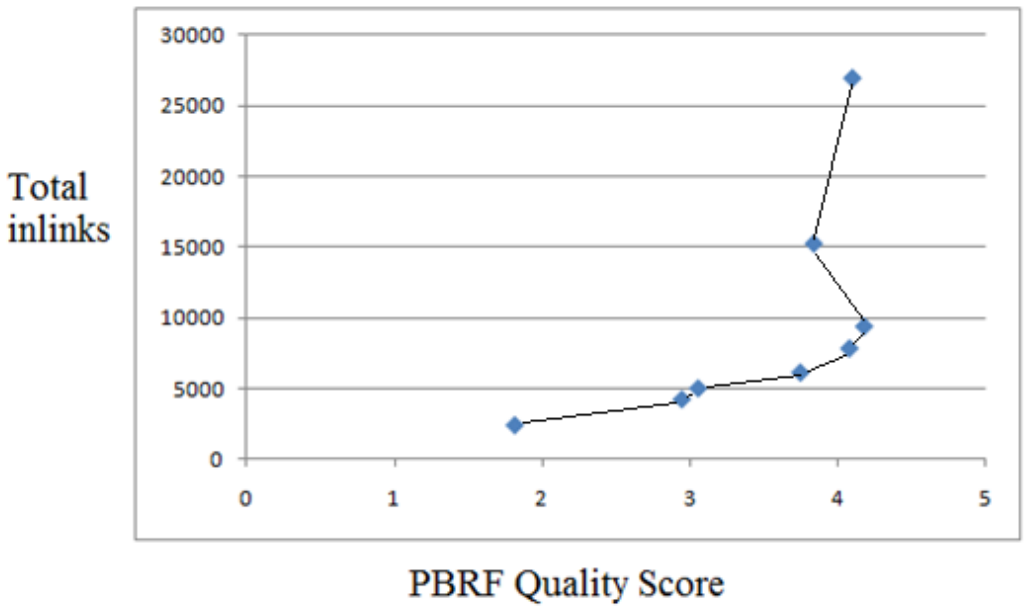

Figure 1

Inlink includes internal inlinks and external inlink. Is there any correlation of external inlinks the PBRF quality score (internal inlinks are not the focus of this research)? The number of total external inlinks of the IRs of the eight universities was collected by YSE:

University	External inlinks
AUT	1273
LU	414
MU	634
AU	3487
CU	991
OU	778 -total of 3 IRs
WU	731
VU	723

Table 5

Compare the number of external inlinks of IRs with the PBRF quality score, as see in figure 2, there is not much correlation between them.

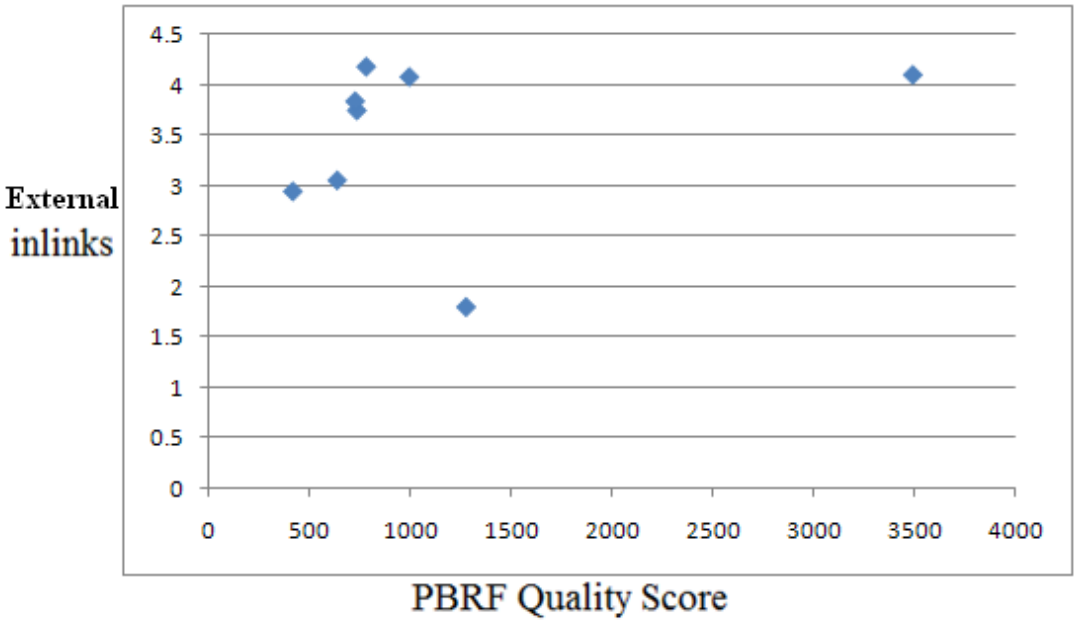

Figure 2

The limitation for this research question is that although it is assumed that total inlinks of IR can show the quality of the university academic output, how relevant of this measure method is which is unknown. Is the correlation between the number of total inlinks and the PBRF quality score just random luck? The method will need to be tested further – more IRs will need to be tested, but because of the time and scope limitations of this research, it can not be done. However, measuring the total inlinks of IR and the quality of its academic output is worth to be followed up in the future research.

6.3 Question three: Is there correlation between the IR website impact factor and the PBRF quality score? Is this result close to the result of comparing the inlinks of an IR and the PBRF quality score?

The website impact factor is calculated as the number of inlinks from outside of the website to individual document within the same website divide by the number of items.

Yahoo Site Explorer is again used by asking it to look for inlinks to the URL of the universities' repositories. For example, the University of Auckland, the URL http://researchspace.auckland.ac.nz/ is put on the YSE website. By explore URL, this repository has 27010 inlinks to the entire site. However, for this research question the inlink needed which is external inlink. The YSE also provides a function to limit the inlikns to "from outside of the website only". This function limits the inlinks to 3487. Within these 3487 inlinks, there are inlinks link to the whole website which means the website has been cited, and also inlinks link to individual documents which means the individual documents have been cited. To find out which inlink is used to cite the website and which link is used to cite individual document, a manual checking process is taking place. Unfortunately, YSE does not provide all the inlinks. This may be able to be solved by submit the website to YSE, but this is impossible to this research, as people must have administrator access to the website before they can submit the website. Fortunately, the YSE allows people to download the first 1000 inlinks. This suited with those repositories with unless than 1000 inlinks, but for those

with more than 1000 inlinks the real number of inlinks to individual items is unknown, but it can be estimated – how to estimate this is discussed in the following part. It is assumed that the inlinks are collected in random order by YSE, but this assumption can not be confirmed.

However, in the eight universities' ten repositories (Otago University has three IRs), only two of them have more than 1000 inlinks – AUT has 1273 inlinks and AU has 3487 inlinks. The first 1000 (if less than 1000, than all the available inlinks) inlinks of each IRs have been download and each link has been manually checked – this means each link has been click to open, and the "Page Source" of each webpage has been carefully searched for whether the inlink links to the repository website or to a individual document. If the link links to the repository, a number "0" is given to the link. If the link links to an individual document, a number "1" is given to the link. After all the 1000 inlinks are checked, the total number was added up. The total number of inlinks was used to generate the website impact factor of the repositories. The following formula is used to calculate the possible total inlinks to individual document for AUT and AU. Total inlinks to individual documents = total inlinks to the individual documents from the first 1000 samples / 1000 x total inlinks to the IR website.

Although, this is not the research question of this research, it worth to note that the higher number of total documents does not necessarily mean a higher PBRF quality

score. Otago University only has 1293 documents which is the 7th place, but it has the highest PBRF quality score. The possible explanation for this is that there are many academic outputs have not been uploaded to IR. This probably because the disciplines that contribute to the quality score (e.g. medicine) do not use the IRs.

The total external inlinks link to individual documents for each repository were manually counted and the number of documents in IRs are collected from the Kiwi Research Information Service (http://nzresearch.org.nz/). Dividing the total number of inlinks link to individual documents with the total documents in each IR, the website impact factor of eight universities' IR is:

University	IR's website impact factor
AUT	236 / 1171 =0.20
LU	221 / 3124 = 0.07
MU	404/ 1668 = 0.24
AU	3051 / 4595 = 0.66
CU	515 / 5003 = 0.10
OU	375 / 1293 = 0.29
WU	331 / 3398 = 0.10
VU	368 / 1407 = 0.26

Table 6

Compare the order of the website impact factor and the PBRF quality score see in figure 3, the result shows that there is a small correlation between each other. There are some universities' website impact factor which is correlated to the PBRF quality score. As discussed earlier, the universities' IR may not hold everything which is published by their academic staff. What if all the IRs have everything their academic staff produced, will the website impact factor correlate with the PBRF quality score? This can only possibly be answer when there is a policy that all the academic output must be put in IR.

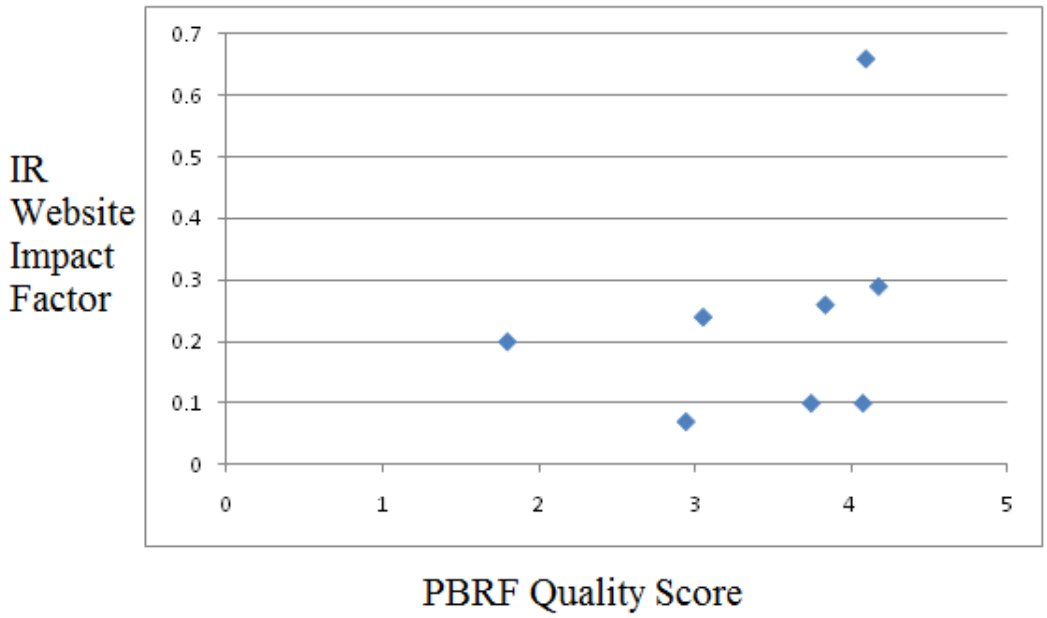

Figure 3

Although the result of the comparison is not very exciting, the finding is interesting. Personal blogs are playing a significant role in promoting repository. In all the

available inlinks, there are 1368 links from blogs. This is number was manually counted from all the inlinks which were collected by YSE (the first 1000 inlinks for each IR). It is about 22% of all inlinks. If excludes the extreme case of AU, there are still 603 inlinks which are from blogs, which is about 11% of the total inlinks. Among these blogs, most of them are used to present people's academic achievement, either by an academic staff or by a student (most of them are postgraduate students). This result is reflected to the finding by Smith (2009) which is discussed in the literature review.

Limitations of this research, first of all is how an individual can affect the entire website impact factor. While collecting the research data, there is an extreme case has been encountered. There is a very active Blog writer, who has been published thousand posts online. On every single of his Blog page, he cited a link to one of the item in AU's repository. As a result, there are 720 inlinks from this person in the first 1000 inlinks of AU's repository. If all of this inlinks are account as 1 then the website impact factor of AU will be significantly reduced from 0.19 to 0.03. In addition to this, among AU's total 3487 inlinks which is to individual documents, there are 2214 inlinks to this particular item. If this 2214 is accounted as 1 again, then total inlinks for AU's individual items will be only 1273 which is more than 63% different. As discussed earlier in this research question, the YSE only provides the first 1000 inlinks, whether the rest of more than one thousand inlinks are from the same person, and if not how many from this person is unknown. This is an alert of IR research,

people should consider the possibility of single document changes the impact factor as a whole.

The other limitation is that while checking the inlinks, there are 600 out of 1000 inlinks from AUT's repository which is requiring login. By search carefully with the information on these websites, the finding is that all of the 600 webpages are from a same AUT postgraduate students forum website. The assumption here is that, the university has built this website for its students and promoting its IR to the student. Since all of the 600 inlinks can not be checked and have to be marked as 0. This is another extreme case. In this case, the website tool Yahoo Exporer somehow picked up those 600 inlinks and put them into the first 1000 inlinks. This leads to a possible mark down on the real impact factor. The example shows how a research can be limited by website setup. It also shows that even a repository is open access, citing from a locked website will not benefit the IR as a while. An open access IR will not be benefit from restricted website, what about those "dark IR" out there which are not public accessible? This is something for the IR developer to think about when they are planning put their documents into the "dark IR".

Second limitation is full access to the inlinks. As discussed above, without full access it can be hard for research to generate accurate information. It could also affect the ranking of an IR website in some ways.

The dynamic website links is another limitation. The information collected will only be accurate for a short period of time after they are produced. The reason is that web links are appearing and disappearing from time to time. The links can break in any second and new citation links can be put online in any seconds. There is no freeze time frame for people to analyze unchanged online information. However, since the number of inlinks is big enough, small changes will not affect the results. The last limitation is the time limited. As the research has to be done within a short period of time, at some aspect an in depth research is not possible.

To expand the research an addition sub-search has been carried on. In this part, 100 documents have been picked up from each IR. The documents were picked randomly. The method used is that the following documents (following the title order) were picked up, no. 1, no. 11, no. 21, … , no. 971, no. 981, no. 991. After the samples have been selected, each of them was checked for inlinks from outside of their IR by using YSE. Then the number of inlinks to individual documents out of 100 samples for each IR was calculated.

The result as follow:

University	Number of inlinks out of 100
AUT	7
LU	1
MU	5
AU	6
CU	2
OU	15
WU	22
VU	9

Table 7

The result shows that a small amount of samples will not be able to provide useful figure, because the bulk of inlinks are linked to the IR, rather than individual document.

When put the results of the total inlinks to IR compare with the PBRF quality score and the IR website impact factor compare with the PBRF quality score together, there are similarity between these two results. The question need to be think about is that because the similarity, is there any other way, people can calculate both information more accurately?

6.4 Question four: Are there software applications that produce more meaningful and usable rankings than the homepage ranking from Google Page Rank? What is the reason?

To answer these questions there are five web rank software have been tested. They are Accurate Monitor for Search Engines, Advanced Web Ranking 7.5, iBusinessPromoter 11, Web CEO, AgentWebRanking. The result is disappointing, all of them are built based on other search engines. They are all a tool to collect information about website/page rank from search engines, such as Google, MSN or Yahoo, rather than produce their own rank. All of them require URL and key words before the rank can be produced. There are many limitations which make these software not suitable for the web ranking and website impact factor in this research. Firstly, none of them can produce a website impact factor, this is understandable as they are not built for this purpose. However, if there is one software can do the job which will be very useful not only for this research but also many other researches. Secondly, they all depend on other search engines. This means if the search engines break down, they will not be able to produce the rank. Third, they are similar software but not very easy to use. In addition, as they are free software, people do not have full function. People have to pay to get it unlocked. The last and also the most important limitation is that they only show the rank of the top 10, top 100 or even top 1000 website. This is a big problem when use these software to rank the eight New Zealand universities' IR, because their website is not in the top rank. This means people can not find a rank by using the software. They may be good software for commercial use

but for a researcher who only need an easy and handle tool, they are definitely not the choice. Therefore, the answer for this research question is that there is no easy way to produce IR website ranking and the software and the Google page are very similar when use them to rank IR. Both of them are not useful

Although the downloadable software is not very helpful for this research, there are some useful website have been found. There are many websites which provide easy to use web rank service, such as PR Checker.Info (http://www.prchecker.info). The most interesting one is the Rank Web of the World Repositories (http://repositories.webometrics.info). The ranks of all eight New Zealand universities' repositories can be found on the website. There are four different parts which contribute to the rank on this website. According to the website (CCHS-CSIC, 2011), they are:

- Size –"Number of pages recovered from the four largest engines: Google, Yahoo, Live Search and Exalead".
- Visibility – "The total number of unique external links received (inlinks) by a site can be only confidently obtained from Yahoo Search and Exalead".
- Rich Files – "Only the number of text files in Acrobat format (.pdf) extracted from Google and Yahoo are considered".
- Scholar – "Using Google Scholar database we calculate the mean of the normalised total number of papers and those (recent papers) published between 2001 and 2008".

The ranking of 8 universities' IR is listed in table 9. It shows that this rank does not correlate with the PBRF quality score.

University	IR ranking	PBRF Quality Score
AUT	554, 1011	1.8
LU	369	2.94
MU	590	3.05
AU	191	4.09
CU	302	4.07
OU	538	4.17
WU	742	3.74
VU	479	3.83

Table 8

This website is useful but also has some limitations. The rank is depended on other research engines a lot. The definition of rich files is to narrow, because if some IR decides to collect documents in the format of Word Document, the documents on the IR will not be included. It only counts the scholar papers published between 2001-2008. The problem is there are many classic articles which were published before 2001. The last problem is that there are two different records for the AUT IR. The possible reason for this error is that the two records have different URL, people

were failed to notice that there were two websites for the same institution when they put the information online.

6.5 Overall limitation

The project is aiming to find out if there is a correlation of website ranking, website impact factor, total number of inlinks and the PBRF quality score. The limitations in general are: first of all YSE can not provide all the information the research needs. Unable to access all the inlinks to AUT and AU resulting on the number of inlinks to individual document have to be calculated based on reasonable assumption. Second major limitation is that the PBRF quality score has been use for 5 years already – last round was produced in 2006. Although, it is believed that the score will not changed much before next PBRF round which is hold in 2012, the real information is unknown. When the next round PBRF information is available the information is used in this research will be out of date. However, the research methods and the ways of analysing data will still be the same. Another limitation is that the websites are not stable, they are changing all the time. As a result, many broken links have been found during the research. The data were collected is only accurate in a period of time. This is something out of anyone's control and there is no easy to fix it. The last limitation major limitation is that open access IR is not necessarily representative of universities research as a whole. For example, there are many research outputs which is held at Victoria University' "dark IR" which is called "RestrictedArchive @ Victoria". The items in this IR are not public accessible, they can only be access by the university

staff and students on campus. The reasons of having a “dark IR” are either the university do not have copyrights to make the documents public available or the sensitivity of the research. The quality of documents on “RestrictedArchive @ Victoria” can not be tested because of the limitation of access. Therefore, the open access IR at Victoria University can not represent the quality of its research output as a while. In addition, there are many research output across the universities which are not held in IRs. The reasons are varied. It can be the willingness of the staff, the sensitivity of the research or not enough promoting of the IR. All of above make the open access IR can not represent the universities’ research output perfectly.

Conclusion

In conclusion, the aim of this research is by comparing different measures of IR website with the PBRF quality score to find out whether there is / are method/s which can be used to measure the quality of an IR. The research shows that website impact factor has a small correlate with the quality of university's academic output. The website ranking does not provide value to represent the quality of IR content. The number of total inlinks, however, is correlating with the quality of academic output.

Other finding of the research is that individuals' role of promoting IR can not be ignored any more. It is clearly shown in the extreme case of the University of Auckland's IR, that one person can accidentally affect the website impact factor by repetitively citing the same website. The find also confirm the finding by Smith (2009) that informal research is playing a major role in today's IR citing activity.

The finding of this project will be useful to the future studies of the quality of open access IR and the quality of academic output, and the studies of analysing and ranking institutional repository.

Reference:

Allard, S., Mack, T. R. & Feltner-Reichert, M. (2005). The librarian's role in institutional repositories: A content analysis of the literature. *Reference Services*

Review, 33/3, 325-336.

Blaxter, L. (2006). How to research, New York: Open University Press.

Buehler, M. A. & Boateng, A. (2005). The evolving impact of institutional repositories on reference librarians. *Reference Services Review*, 33/3, 291-300.

CCHS-CSIC, (2011), The rank web of the world repositories, Retrieved from: http://repositories.webometrics.info.

Chan, D. L. H., Kwok, C. S. Y. & Yip, S. K. F. (2005). Changing roles of reference librarians: the case of the HKUST Institutional Repository. *Reference Services Review*, 33/3, 268-282.

Chang, S-H. (2003). Editorial: Institutional repositories: The library's new role. *OCLC Systems and Services*, 19/3, 77-79.

Conway, P. (2008). Theme article modeling the digital content landscape in universities. *Library Hi Tech*, 26, 342-354.

Craven, P. (2011), Google's PageRank explained and how to make the most of it, Retrieved from: http://www.webworkshop.net/pagerank.html.

Dorothea, S. (2008). Innkeeper at the Roach Motel. *Library Trends,* 57, 98-123.

Graham, J. Skaggs, B. L. & Stevens, K. W. (2005). Digitizing a gap: a state-wide institutional repository project. *Reference Services Review,* 33/3, 337-345.

Kalal, S. K., Biswas, S. C., and Mukhopadhyay, P., (2010) Web impact factor and link analysis of selected Indian universities, *Annals of library and information studies,* 57, 109-121.

Kiwi Research Information Service, (2011), Retrieved from: http://nzresearch.org.nz/.

Maxwell, J.A. (2005). Qualitative research design, UK: Sage Publications, Inc.

Organ, M. & Mandl, H. (2007). Institutional repositories outsourcing open access digital commons at the University of Wollongong, Australia. *OCLC Systems & Services: International digital library perspectives,* 23, 353-362.

Page Rank checker, (2011), Retrieved from: www.prchecker.info.

Phillips, H., Carr, R. & Teal, J. (2005). Leading roles for reference librarians in institutional repositories: One library's experience. *Reference Services Review,* 33/3,

301-311.

Piorun, M. & Palmer, L. A. (2008). Digitizing dissertations for an institutional repository: a process and cost analysis. *Journal of the Medical Library Association*, 96, 223-229.

Registry of Open Access Repositories, (2011), Retrieved from: http://roar.eprints.org.

Rockman, I. F. (2005). Distinct and expanded roles for reference librarians. *Reference Services Review*, 33/3, 257-258.

Shreeves, S. & Cragin, M. H. (2008). Introduction: Institutional Repositories: Current State and Future. *Library Trends*, 57, 89-97.

Smith, A.G. (2009) Linking to Institutional Repositories from the general Web. Proceedings of ISSI 2009, Rio de Janeiro.

Smith, A. G., Thelwall, M. (2005). Web links as an indicator of research output: a comparison of NZ Tertiary Institution links with the Performance Based Research Funding assessment. Proceedings of ISSI 2005, Stockholm.

Tertiary Education Commission, (2009), Performance Based Research Fund,

Retrieved from: http://www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/.

Yahoo Site Explorer, (2011), Retrieved from http://siteexplorer.search.yahoo.com.

Yeates, R. (2003). Over the horizon Institutional repositories. *VINE*, 33/2, 96-99.