

Rediscovering older information

Bob Stembridge
Thomson Scientific
November 2006

A number of information providers have recently extended coverage of their collections back in time to provide electronic access to historical research and news. In 2005 Thomson Scientific added a new archive — Century of Science[®] — to the Web of Science[®], which has extended coverage of scientific research information back to the beginning of the last century. This article explores the impact of providing access to older information. It is based on a presentation given at the ACS 232nd meeting during September 2006¹ (<http://www.acscinf.org/>).

Dramatic discoveries

The early part of the 20th Century was a time of dramatic progress in scientific discovery — Einstein's Theory of Relativity, Marie Curie's work on radioactive elements, the extraction of insulin and the discovery of penicillin were just a few of the discoveries that caused fundamental shifts in scientific understanding. These famous examples are the tip of a valuable iceberg of lesser known work that is increasingly being exposed through information provider's activities in capturing and making this research electronically available. To what extent is the increased availability of this information impacting current research? That is the question we set out to explore in this article.

The *Century of Science* archive comprises over 850,000 fully indexed journal articles from more than 200 scientific journals published from 1900 to 1944. This addition to the *Web of Science* has made it possible to comprehensively analyze citation information across the early part of the 20th Century. As we'll see, there are several highly cited works in this so-called "golden era" of scientific discovery, especially in Chemistry and Physics. Interestingly, many of those are being increasingly cited today.

Albert Einstein

A clear example of this is the work of Albert Einstein. An analysis examining the number of times his 1905 paper on the Special Theory of Relativity² has been cited during the period 1900 – 2006 shows clearly that it has been cited much more heavily in recent times. That is interesting in itself, but particularly so if you compare this with the trend for other articles published in the same journal, *Annalen der Physik* (figure 1). The citation frequency over time of papers, other than Einstein's, published in 1905 decreased steadily over the first 40 years to its current steady low level. So the citation trend shown by Einstein's 1905 paper is directly opposite to the general trend for other papers published in the same journal from that time.

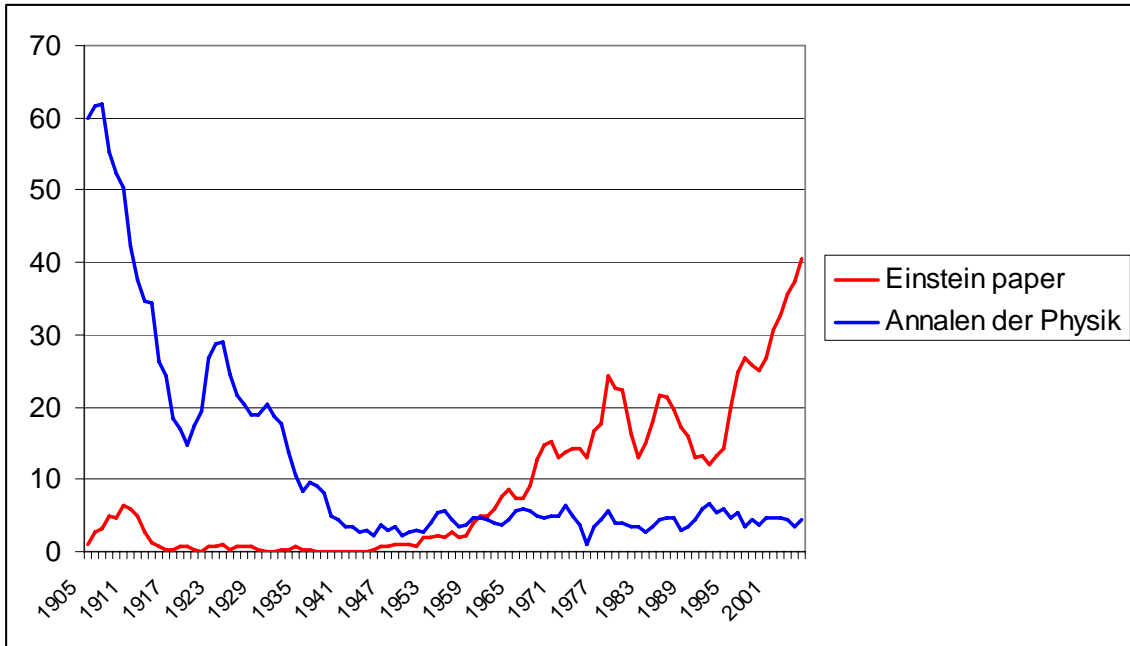


Figure 1: Citation frequency of Einstein's paper compared to other papers published in Annalen der Physik

There may be several possible reasons for this:

- It could be because research in the field of relativity has not yet reached a satisfying conclusion. Einstein spent the last 30 years of his life in pursuit of the Unified Field Theory to bring together generalized descriptions of electromagnetism and gravitation. Others continue that search to the present day
- Another possibility (raised by Marx and Cardona³) is that theories discussed in the original work are so revolutionary that they cannot be fully tested until many years after publication
- It may also be for historical reasons. 2005 was the 100th anniversary of the publication of this paper and also the 50th anniversary of Einstein's death, so that may have had an impact on the recent citation rate, albeit a temporary one
- Lastly, it could be due to the increased visibility and accessibility of Einstein's work

Top five cited papers from *Century of Science*

Not all articles follow the same trend. Consider the citation rates over time for the top five cited papers from the *Century of Science* (figure 2)

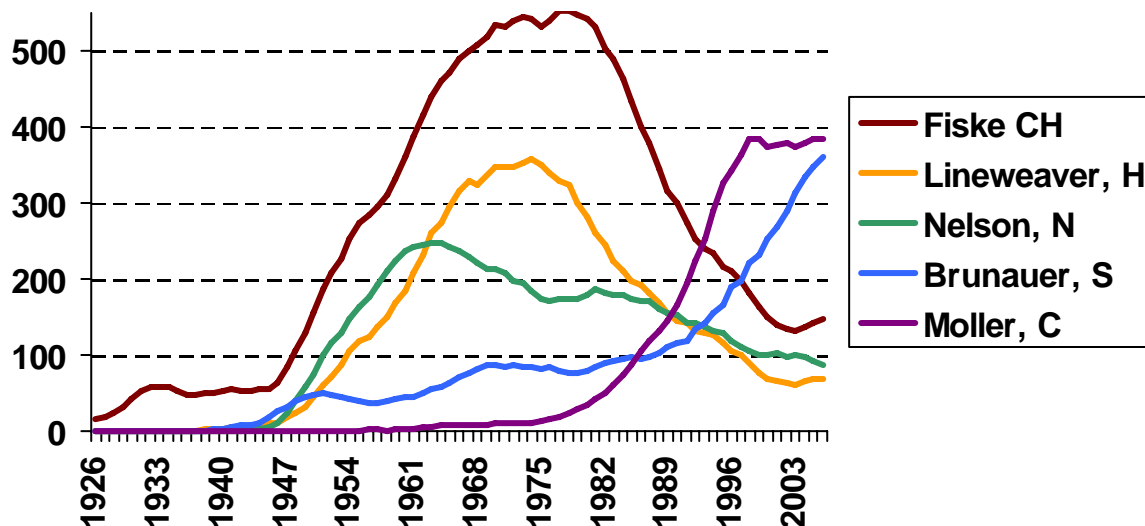


Figure 2: Citation frequency for the top five papers from *Century of Science*

The first three of these papers (Fiske, Lineweaver, & Nelson, all of which have declining citation rates in recent times) are concerned with determination of physical properties. The trend for the last two papers (Brunauer, & Moller) is more interesting. They show increasing citation rates over the last 20 years that far exceed the historic rate. The titles indicate that they are relevant to current research — “Adsorption of gases in multimolecular layers” and “Note on approximation treatment for many-electron systems”.

In particular, the Brunauer, Emmett and Teller (BET) article⁴ on gas adsorption published in 1938 has been cited much more frequently in recent times than in the past. This paper describes a method for calculating the surface area of solids through surface adsorption of gas. There are many applications of this technique reported in recently published articles citing this paper, including technologies such as nanotechnology, pollution control and catalysis.

An analysis of just the last couple of years' worth of papers citing the BET article shows that geographical coverage of institutes involved in this research is international, including France, Spain, China, Italy and Brazil. It would appear this article is well-known throughout the scientific community.

User survey

Beyond looking at citation analysis, we wanted to discover further evidence about the impact of older information in electronic archives through conducting a user survey.

A questionnaire was developed with questions concerning the BET article as follows:

1. Where did you first hear about "Adsorption of gases in multi-molecular layers"?
2. How did you then locate a copy of the full text of the article?
3. Aside from the historical perspective of this article, how relevant is it to your research now?

The survey was sent to 200 practicing chemical scientists in this field from whom 72 replies were received. Responses came mainly from academics from a geographically wide spread area, but also from government and industry organizations as well.

In reply to the first question, 50% of respondents said they retrieved this reference from another article or book, and only 1% from a web search or bibliographic database. For those that responded that they became aware of this article from other sources, all of them reported that they first learned of this article as part of an academic course, one commenting *"Actually, I read about it when I was a graduate student, in the 1950s"*.

In reply to the second question, only 15% responded that they obtained a copy from an electronic full text archive, with over 50% saying they obtained the print copy from their library. This is surprising, since other evidence suggests that electronic archives are being well used. That being the case, if electronic archives amongst this constituency are only being used to the small extent suggested, there may be much potential for increased usage out there.

In reply to the third question, 86% responded that the research is either very or somewhat relevant to current research.

Clearly this historic research published in 1938 is still relevant to research today.

Patent citation rates

Citation analysis of patents was also conducted to see whether the availability of older information in electronic archives is having an impact on later published research.

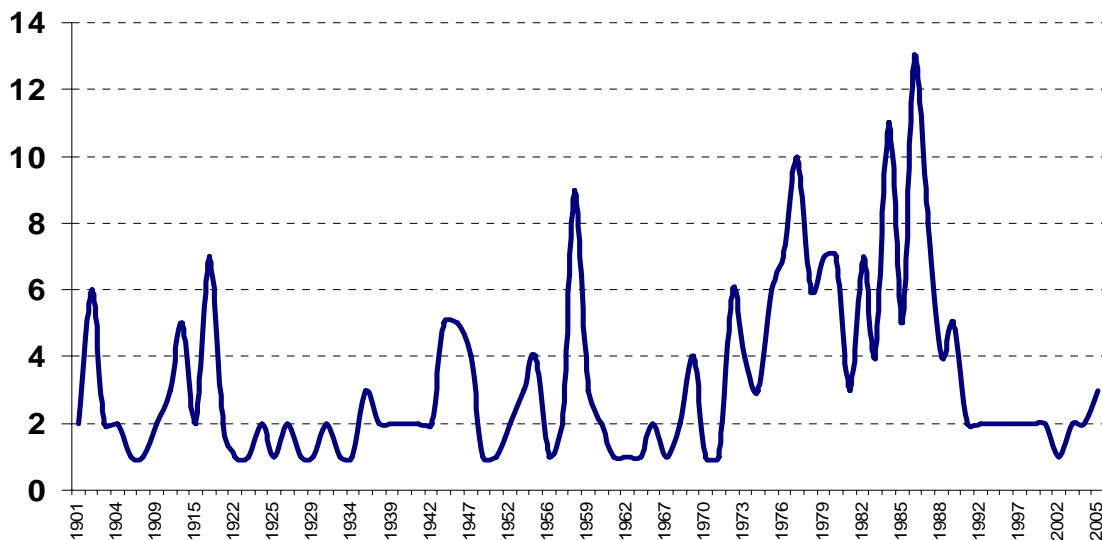


Figure 3: Citation frequency for patents published in 1900

Analysis of citations to patents published in 1900 in later published scientific papers (Figure 3) shows no clear trend, so we might be tempted to conclude that there is no significant increase in citation frequency with the availability of older information. However, this may also be because the technology described in patents from 1900 is predominantly chemical, and in that technology area a method discovered in 1900 will work as well and is as useful today as it was then.

As an example of this, one of the papers we examined is a research article published in 2005 on the porosity characteristics of an activated carbon filament produced by a heat treatment process. There are 14 references cited in this paper. One of these is a British patent from 1900 describing a three step process for activation of charcoal. Clearly, the method described in the 1900 patent of heat treatment is still a good method over a hundred years later in 2005. Interestingly, there is also a citation to the 1938 BET method of estimating surface area through measuring gas adsorption showing that this research also makes use of this historic science.

Summary

In summary, we see that:

- some classic historic research is increasingly heavily cited
- a survey of practicing researchers shows that
 - the majority found the 1938 BET article from a cited reference of bibliography of another work, and web searching does not seem to have played a major role in the awareness of this article
 - the majority of researchers still access the full text of this article via the print journal, which may reflect a lack of access to the electronic copy
 - the majority of researchers still believe that this work is relevant to their research
- a citation analysis of the relationship between patents and articles shows that historic patents are also still being cited, but there is no clear evidence of an increasing citation rate
- retrospective coverage of both journals and patents is valuable to current research

Conclusion

Other research presented at the ACS symposium suggested that overall usage of research information sources increases when older archival information is added (for example, the findings presented by Helen Barsky Atkins⁵) However, our research shows no conclusive evidence that this increased access and usage is yet to translate into significant impact on published research as measured by a general increase in citation rates of older information.

This may be a consequence of timing. The addition of the *Century of Science* archive to the *Web of Science* took place in early 2005, a little over eighteen months ago. For research papers and patents written today, there is a significant lead time until publication. Certainly for patents that are publishing today, applications would have been submitted at the beginning of 2005 — or in other words about the time that the *Century of Science* was introduced. It's quite possible that the introduction of this resource has not yet had time to show an impact in terms of increased citations to this older material. It will be interesting to repeat this research in a couple of years' time to measure the impact then.

References

- ¹ Pratt, S.M. and Stembridge, R., *An analysis of citations in scientific and patent literature to historical research from the first half of the 20th century and the relationship to the accessibility of these works on electronic archives*. CINF Symposium, American Chemical Society 232nd National Meeting. 2006 San Francisco
- ² Einstein, A., *The Electrodynamics moving body*. Annalen der Physik, 1905. **17**(10): p. 891-921.
- ³ Marx, W. and Cardona, M., *Blasts from the past*. Physics World, 2004. 17(2): p. 14-15.
- ⁴ Brunauer, S., Emmett, P.H., and Teller, E., *Adsorption of gases in multimolecular layers*, Journal of the American Chemical Society, 1938. 60(Jan-Jun): p. 309-319
- ⁵ Barsky Atkins, H., *If you build it, will they come? Experience with journal backfiles*. CINF Symposium, American Chemical Society 232nd National Meeting. 2006 San Francisco