

## Genealogy can be reproducible and efficient

*Lars Berglund*

**Abstract** Genealogy has problems with poor quality of source citations and questionable ancestries dating back to ancient times. Our research has gained a reputation for being unreliable. The author believes that serious genealogy is "real" research. Thus, we should adopt the review process that exists in other areas of scholarship. Quality journals where genealogists publish results after review by knowledgeable colleagues would improve our reputation. Publication in a journal is a goal that rewards hard work. If nothing is done there is a risk that our descendants will be forced to reject our results. Genealogy should not deserve to be scorned but taken so seriously that conventional research methods are used.

The intention with this article is to address some problems with genealogy in Sweden but also in other countries. The article also identifies an efficient solution.

One problem area that has been discussed for a long time is to what extent we cite sources. How to specify sources and how to make source evaluation is described by various authors<sup>1</sup>.

Despite this methodological support and the extensive discussions nothing happens in the real world of genealogists. The author made a study on March 30, 2014 among 30 pages of genealogical pedigree charts on the Internet. A site that displays links to private sites was used and the first 30 non-broken links were chosen. At 24 of these sites there was no information on sources. Among them there was a person who offered research for payment but did not specify any source for information about his own ancestry. Four pages contained inadequate information, where only a few data items would have an indication of source like "Parish Name C: 5 birth, consecrated,

---

Copyright 2014 Lars Berglund, Associate Professor. Permission to reproduce is granted if not for profit or commercial advantage, and if the original work is properly cited.

Author's address: lars.berglund@ucr.uu.se Cite as: Lars Berglund, "Genealogy can be reproducible and efficient", *Annals of Genealogical Research* Vol. 10, No. 1 (2014) at <http://www.genlit.org/agr/viewarticle.php?id=56>

deceased 1834-1860" but without page numbers. Two genealogical sites stated their sources admirably and their data could at random be verified<sup>2</sup>.

If this survey is representative, it means that about 80% of published pages are devoid of indications of source and that only 7% of the pages indicate sources so that others can use their data. Most of these 30 genealogists have probably done extensive and solid work. Unfortunately, most of the results are not usable. It may be that some genealogists have registered sources but have chosen not to include them in the presentation. In this case it is even more egregious because their important work becomes of little worth when reported. The reason for not showing sources in presentations may be that genealogy programs handle this information clumsily when web pages are created but then the programs should be changed. The author's own experience implies that it does not take much extra time to register sources and it is an immense advantage, not just when publishing the results, but also for current research work.

Another problem that is tangent to the first is the imagination of researchers. Unfortunately the drive for many genealogists is to find as many and as early relatives as possible. Four examples of this overgrown flora: A researcher writes that he is related in a direct line with the god Odin. A local newspaper described a researcher who has come to the conclusion that she is related to a pharaoh in ancient Egypt. A researcher asks in a forum how to enter dates before Christ in a register program. A researcher boasts that he has the oldest ancestor and more entries than anyone else<sup>3</sup>.

All of this has given Swedish genealogy a reputation for unreliability and even dishonesty. What will happen the day when journalists from general media ask us where we got our information and whether we can trust it? There are many discussions and much advice on methodological issues but few seem to have had the thought that the results of research need to be

examined. There has never been made so much genealogy work as is done today. Imagine in 20 years: Will our children and grandchildren be forced to reject all of this work (or at least 93%!) and start over? We need to improve our reputation so we can attract more serious people to the exciting and rewarding pursuit that genealogy is. We also need to streamline research to avoid duplicate work.

The author's proposal is that we use the same review process that already exists in most scientific fields. Suppose that there are a number of journals where genealogists can publish their results. (We call hereafter the submitted genealogical research an "article" and the genealogist its "author".) When the journal gets a request from an author to publish an article it is sent for review to two or three reviewers. The reviewers should remain anonymous to the author.

There are a large number of reputable genealogists in Sweden. There will be both carrots and sticks to motivate review work. Many of the knowledgeable genealogists will want to act as reviewers to increase the quality of our research, and to improve our tarnished reputation. This work can be used in the reviewer's CV which would be useful for example for the professional genealogists who perform research on commercial basis. Another motivation for doing review work is that it gives a broad knowledge of current research results and methods which may be usable in the reviewer's own research. The stick part is the group pressure (especially in a small community like Sweden where genealogists socialize in small circles). It can be perceived as disloyal to say no and offload work to others.

Nonetheless, the success of the proposal highly depends on the demand from genealogists to get their work examined and the willingness from knowledgeable genealogists to perform article reviews. These factors are largely unknown and will only be brought to light when a number of genealogy journals offer the possibility to publish peer-reviewed articles.

The review shall not examine all details. The examiner assesses what needs to be verified. He or she also makes an evaluation of the presentation, such as how readable and interesting it is. This process will encourage genealogists to do research summaries that can contain general information such as geographical areas, occupations, life span, common diseases and causes of death for the family. Moreover, one can highlight interesting life stories and special events. At present one often comes straight into a pedigree on the Internet without getting the overview that makes the presentation appealing for more people than the closest relatives.

The comments, which are sent to the author, should give constructive criticism and suggestions to improve the article for publication in this or, at rejection, possibly in another journal. A review leads to a recommendation of rejection, minor or major changes or approval for publication. The editor then makes a decision based on the reviewers' recommendations. It is common for open journals to charge for published articles in order to finance the editorial work. By contrast, the reviewers' work is not remunerated. The rationale for this is that it eliminates the risk that individuals undertake many assignments only for financial reasons.

When an article is published, it gets an identifier, such as "Pedigree NN, Journal Name 2014-07-01 p. 599 [www.genealogi.se/ ...](http://www.genealogi.se/)", to be used when other genealogists want to quote information. It may be objected that it is better to use primary sources and this should naturally be done in the first place but this will be a big boost compared to the examples given above. Moreover, not all genealogists need to go to the original sources if this is done once and then stamped quality. The preferred article format is PDF and with possibly a GEDCOM file as supplementary material. Published articles are administered by the journals and thus have a stability which is lacking on Internet sites where individual genealogists are responsible for availability.

After a few years, for each journal a so-called impact factor is computed, i.e. the average number of citations for papers published in the journal. It is common that the citations are calculated for the two years following its publication in a particular year so if the journal in 2014 published 45 articles, and these are cited 158 times by articles in this and other journals during the years 2015 and 2016 the impact factor will be 3.5 for the journal in 2016. It is important that there are several journals that compete with each other. Bibliometric methods for measuring the number of citations are easily accessible and are well proven. There are several existing Swedish journals and associations who can shoulder this mantle. Journals that introduce serious examination processes can seek and receive financial support from research foundations and commercial operators will certainly have a substantial interest in supporting this development. Eventually there will be a pecking order among the journals and serious genealogists will aim to be published in the highest ranked journals.

Reviewers may on own initiative keep statistics on the percentage of false information they find. It can be anything from typos to such serious mistakes that lead to wrong relationships. Everyone makes errors. When this author entered the data for the family history for the five youngest generations, another year was devoted to verify and complete data. A large number of minor errors were found and in one case an assumption of kinship was too vague and had to let go on reflection.

A problem that arises is that genealogy work is not static. Information is continually added but a review refers to a single static set of data and relationships within an article. Active genealogists will want to publish further articles with additional data (and maybe corrections to previously published results). It may be reasonable to first publish the parts of research which the author believes has enough quality, e.g. data for the four youngest generations. A second article covers these data (possibly with a

reference to the first article) and additions made, e.g. older generations or estate inventories, and the reviewer will examine the second article as if it was completely new research.

Nothing will prevent people from posting their work online without being published in journals. However, there will be a significant boundary line between such amateur reporting and quality results published in journals. The unstructured variety with pedigrees from the fairy world will be diminished and respect for our research will increase significantly. We will get much attention in the general media when this process is introduced.

Other types of articles can also be published in these journals, such as articles on methodology and empirical studies as outlined above. There are already plenty of methodology articles but one advantage of this scheme is that such articles get a seal of approval by being reviewed professionally.

Genealogy differs from other forms of science in many respects. Other science fields are financially supported by a large number of disparate organizations. A large number of persons have careers in science via business and industry, education, and government, and those careers often depend on publication and the assessment thereof. Genealogy, in contrast, is of concern or interest to a much narrower slice of society although the number of practitioners is large and growing. It has next to no financial support from charitable, educational, or government organizations. There are few people with a career in genealogy and for these persons publication have had less importance than in other forms of science. Other aspects have traditionally been more important, e.g. advertising, public lectures, personal recommendation, etc.

Although genealogy differs from other science fields the similarities are also large and the current proposal is a path to decrease the gaps. All scientific fields must use reproducible methods and results need to be easily available. For genealogy

research to be acknowledged and used by e.g. professional historians this proposal is crucial.

The journals/associations exist and ready to use bibliometric methods are available. There are already several collections online that present private genealogy pages. The administration is thus there. The journal to be the first to use the review process will have a head start and with an ambitious review process it will likely get a high impact factor.

The method described, the so-called peer-review process, has been used for hundreds of years in most scientific fields and it is extremely successful. Any scientist can post their creations on a website but the scientific community cares only about results that are published in reputable journals. The peer-review process is the difference between science and therapy.

One possible objection is that review already exists, e.g. in the form of comments from teachers of genealogy courses who assess students' first steps and as exchanges between members of genealogical societies. All of this is important, but partly it is far from enough, as my survey above shows, and partly it's like writing in water, i.e. neither the examiner nor the genealogist gets documentation showing that examination was performed and what the outcome was.

One has to ask what the alternative for us genealogists are. To the author's knowledge, there is no organized quality improvement work targeted on published results. If nothing is done, false information will spread and research based on incorrect perceptions of kinship must be restarted from the beginning.

It will take some years to implement this proposal. It will take time before the journals and review process will have found forms that are useful in practice. For example one should build up common assessment criteria that all reviewers can use. Many people will be critical of the proposal, but everyone can

agree that the highlighted problems are significant and growing.  
Critics may well give alternative solutions to these problems.

---

<sup>1</sup> see e.g. Wiki Roots [www.genealogi.se/wiki/index.php/K%C3%A4llkritik](http://www.genealogi.se/wiki/index.php/K%C3%A4llkritik), Michael Lundholm [www.dis.se/sv/publikationer/diskulogen/senastenumren.html](http://www.dis.se/sv/publikationer/diskulogen/senastenumren.html) and Håkan Skogsjö [www.genealogi.se/images/pdf/kvalitet1.pdf](http://www.genealogi.se/images/pdf/kvalitet1.pdf), all in Swedish but available in English with Google translator.

One can find it odd that no one has carried out an empirical study in which one tries to retrieve published data given different methods of source citing, such as different ways to specify the page numbers. If one can find the same data, the time to find it etc. are parameters that can be compared between methods. Why not use existing objective approaches rather than the infected discussions that are now going on?

<sup>2</sup> The addresses to the 30 web sites are available after request to the author.

<sup>3</sup> The purpose of these examples is not to hang out individuals. The author refers to a general discussion of the phenomenon in the forum <http://forum.skalmann.se/viewtopic.php?f=41&t=21514> April 14, 2014.