2 Searching and Reviewing Scientific Literature

2.1 Introduction

This chapter is on searching relevant literature, reviewing and ranking published books and journal articles, and extracting relevant information in the form of text, data, or graphs. In this context, the focus of our book is on Internet resources and literature in an electronic format such as the Adobe *Portable Document Format* (PDF), rather than on printed journals and books in a library. Although some of you might have a printed version of this book in your hands, most people have probably taken advantage of Springer's eBook Collection to read the digital version on a computer.

The advantages of electronic resources are very obvious, as drawers of suspension files filled with hundreds or thousands of printouts of journal articles, or office shelves containing large numbers of heavy books are of little use when working at home or traveling to conferences. This chapter focuses on the most popular open source and commercial Internet resources. The reference search described in the following section demonstrates a typical literature survey by students, starting with the popular online encyclopedias and bookstores, continuing to the generic web search engines to find relevant literature, and ending up in commercial literature data bases designed for professional use.

2.2 Resources for Literature Reviews

In this section, we go through a typical information and literature search, starting with the very popular web-based, open source encyclopedia *Wikipedia* (http://www.wikipedia.org) to find information on earth science topics. Wikipedia was launched in 2001 by Jimmy Wales and Larry Sanger. It attracted 78 million visitors in 2011 and has 91,000 contributors working on about 17 million articles, according its own article about the site. It is written and edited by largely anonymous Internet volunteers, under the supervision of experienced editors who ensure that any edits are cumulative

M. H. Trauth and E. Sillmann, *MATLAB® and Design Recipes for Earth Sciences,*
DOI: 10.1007/978-3-642-32544-1_2,

improvements. Although often criticized for most articles not being edited by experts in the particular topic, the obvious advantage of Wikipedia is the topicality of its content.

Searching for literature, especially when trying to familiarize oneself with a completely new topic, usually starts in a bookstore. We have agreed to use Internet resources in this book and hence the bookstore will be online. The largest online retailer for books, and more recently for many other products including music, consumer electronics, and even food, is the online commerce company *Amazon* (http://www.amazon.com). Amazon was founded in 1995 by Jeff Bezos. In 2007 Amazon launched the Amazon Kindle, an ebook reader that can download content from the web via a wireless network.

Google was set up in 1996 by Larry Page and Sergey Brin as a graduate student research project at Stanford University. In addition to the general web search services, Google also hosts *Google Books* (http://books.google.com), which was introduced in 2004 under its original name of Google Print. This service searches the full text of both original ebooks and scanned books, as well as excerpts from commercial books, provided as previews by publishers. The sister webpage is *Google Scholar* (http://scholar.google.com), which was introduced in 2004 and searches online journals rather than books. The idea behind this project was to provide a freely-available article search engine as an alternative to some of the commercial products such as Elsevier's *Scopus* or Thomson Reuters' *ISI Web of Science*.

The Web of Science was founded by Eugene Garfield in 1960 and subsequently acquired by the Thomson Reuters Corporation, an information company headquartered in New York City. The Web of Science can be accessed by universities and research institutions from the *Web of Knowledge* (http://isiknowledge.com) index. This index includes about 23,000 scientific journals and includes the Web of Science journal listing and citation index, as well as 110,000 conference proceedings and many other academic resources.

A commercial bibliographic data base specializing in earth science literature is *GeoRef* (http://www.agiweb.org/georef) produced by the American Geological Institute (AGI). GeoRef was launched in 1966 as a series of printed catalogues, later evolving into a CD-based system and ending up as an online service. According to the AGI webpage, this data base contains over 3.2 million references including journal articles, books, maps, conference papers, reports, and theses.

2.3 Finding the Relevant Literature

We now use the Internet resources introduced in the previous section to demonstrate a typical search for books and journal articles, using various indices to rank the literature. We first start with commercial online bookstores, proceed to Google Books and Google Scholar, and end up at professional literature search data bases such as the ISI Web of Science.

Searching with Online Bookstores

In our first example, we would like to explore a new field for future research activities, e.g., human evolution. Let us assume that we are planning to spend €200 on relevant books and start with a literature search on *Amazon* (http://amazon.com) or any other online bookstore. After entering *human evolution* in the search field in early 2011, we obtained a list of more than 40,000 books, DVDs and other products, showing the results in order of relevance. A search for *human evolution* in early 2011 produced the following list of books (Fig. 2.1):

```
1. Palmer D (2010) Origins: Human Evolution Revealed. Mitchell
Beazley

2. Stringer C, Andrews P (2005) The Complete World of Human
Evolution. Thames & Hudson

3. Zimmer C (2007) Smithsonian Intimate Guide to Human Origins.
Harper Paperbacks

4. Gibbons A (2007) The First Human: The Race to Discover Our
Earliest Ancestors

5. Sawyer GJ, Deak V, Sarmiento E, Milner R (2007) The Last
Human: A Guide to Twenty-Two Species of Extinct Humans

6. Wood BA (2011) Human Evolution (A Brief Insight). Sterling

7. Shubin N (2009) Your Inner Fish: A Journey into the
3.5-Billion-Year History of the Human Body. Vintage

8. Lewin R (2004) Human Evolution: An Illustrated Introduction.
Wiley-Blackwell

9. Cochran G, Harpending H (2010) The 10,000 Year Explosion: How
Civilization Accelerated Human Evolution

10. Foley RA, Lewin R (2004) Principles of Human Evolution.
Wiley-Blackwell
```

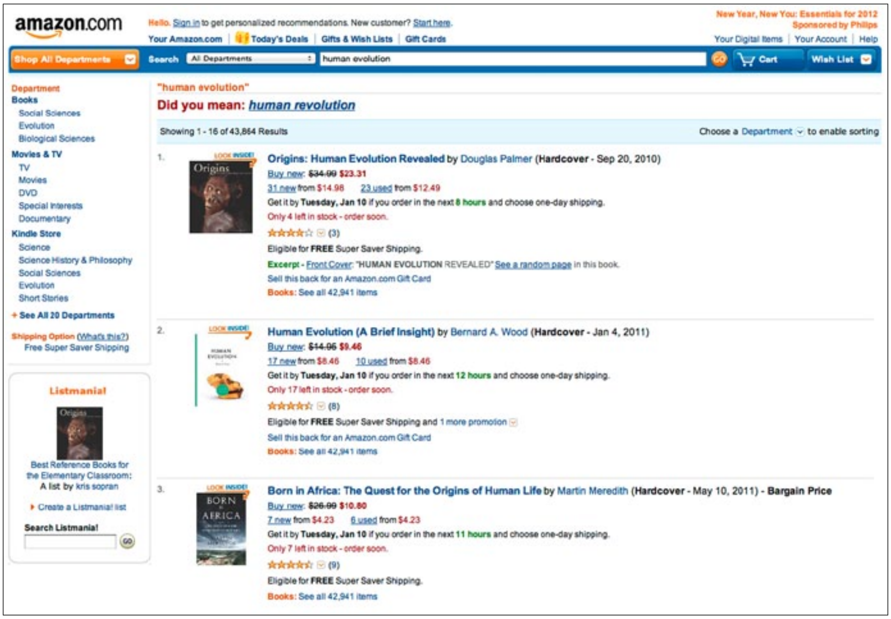

Fig. 2.1 Screenshot from Amazon (http://amazon.com) for the search term *Human Evolution*. The date of screenshot is 9th January 2012, whereas the search experiment described in the text was performed in early 2011, which explains the slight difference in search results.

The search technology is provided by *A9* (http://a9.com), which is owned by Amazon. Amazon also ranks all products by their sales from their own webpages (the *Amazon sales rank*, or ASR, which is updated every hour). Sales, not citations, make a book popular at Amazon: both professionals working in the field of human evolution and interested lay people contribute to the sales of a book, and therefore to the ranking of a particular product.

What makes a customer decide to buy a particular book through Amazon? Why are certain books popular at Amazon and others not? There is much speculation about the exact mathematical algorithm used by the company to calculate the ranking of a book. The obvious factors that may increase the sales of a book are a low price, recent publication date, good title, attractive cover, great illustrations, and positive customer reviews. The reputation of the author is certainly an important criterion within the professional community but might be a secondary consideration for other customers. Douglas Palmer, author of the first-ranked book on human evolution and a very successful author of popular books on paleontology and the earth's history, has not published many original research articles. Ann Gibbons, author of the fourth-ranked book, is a correspondent for *Science* magazine, for which she has covered human evolution for more than a decade. She is certainly in close contact with scientists conducting original research in the field of paleoanthropology and paleontology.

In contrast, Bernard Wood, who authored the sixth-ranked book entitled *Human Evolution (A Brief Insight)* is an anthropologist and the University Professor of Human Origins at the George Washington University. His paperback book of 176 pages, which provides a very brief summary of current knowledge concerning human origins, received mixed but predominantly positive reviews. Robert Foley and Roger Lewin, who authored the tenth-ranked book, are also active researchers in the field of human evolution and have authored numerous highly-cited journal articles and books. Their book, *Principles of Human Evolution*, states very clearly, that it is a textbook for students and professionals rather than a popular science book for an interested lay audience. Robert Lewin also authored the eighth ranked book, *Human Evolution: An illustrated Introduction*, which is a popular book written by an active researcher in the field of paleoanthropology.

These four examples illustrate that authors of science books can range from science writers and journalists with a science background, often within the field of the book's topic, to active researchers writing both popular science books for an educated lay audience and textbooks for students or professionals with a particular interest in the topic. A very good example of a popular book written by an expert is the book

```
Potts R, Sloan C (2010) What Does It Mean to be Human? National
Geographic
```

which was ranked *165* by Amazon in early 2011. Rick Potts is the director of the Smithsonian Institution's Human Origins Program (http://humanorigins.si.edu/), and Chris Sloan is the National Geographic's paleoanthropology expert (http://www.nationalgeographic.com). In this example, the collaboration between an active researcher and a popular science writer has resulted in an excellent book for both experts and lay persons.

These five books together come to less than €200 and cover the full range of book types, from a textbook for students to a popular book for the interested lay person. As demonstrated in this section, however, it may require a lot of background information on the various authors before the best books on a specific topic can be selected. The next subsections provide some more objective criteria with which to rank books and journal articles.

Searching with Google Books and Google Scholar

What alternative rankings are available that are not based on commercial factors and undisclosed mathematical algorithms? Let us try *Google Books*, as a non-commercially biased search service. By once again entering *human evolution* in the search field, we obtained a list of about 2.2 million results in early 2011 sorted by relevance. Here are the top ten items on the Google Books list:

```
1. Wolpoff MH (1995) Human Evolution. McGraw-Hill

2. Stringer C, Andrews P (2005) The Complete World of Human
Evolution. Thames & Hudson

3. Bradshaw JL (1998) Human evolution: a neuropsychological
perspective. Psychology Press

4. Lewin R (2005) Human Evolution: An Illustrated Introduction.
Wiley-Blackwell

5. Regal B (2004) Human evolution: a guide to the debates.
Controversies in Science, ABC-CLIO

6. Campbell BG (1999) Human Evolution: An Introduction to Mans
Adaptations - 4th Edition. Aldine Transaction

7. Maxwell M (1984) Human evolution: a philosophical
anthropology. Columbia University Press

8. Wood B (2005) Human evolution: a very short introduction.
Oxford University Press

9. Cela-Conde CJ, Ayala FJ (2007) Human Evolution: trails from
the past. Oxford University Press

10. Khanna DR (2004) Human Evolution. Discovery Publishing House
```

This certainly is a significantly different type of list, including not only books but also book chapters; it is computer generated by an automated search for keywords without the involvement of any quality criteria. The books on the Amazon list by Roger Lewin, Bernard Wood, and Chris Stringer with Peter Andrews, also appear on the Google Books list, but at least in one case as an older edition with a different publisher. The other items listed are relatively less popular books or monographs on specific aspects, current debates, or perspectives of human evolution. This example demonstrates that, although biased by non-scientific criteria, searching for books on the basis of sales-based rankings seems to be the better option for finding the relevant literature.

Finding relevant journal articles on a specific topic is a much simpler task by far. In this case, the rankings are based on citations by experts rather than on sales to a mixed professional and non-professional clientele. As an example, we aim to find scientific literature on the influence of the Younger Dryas cold reversal in Africa. The online encyclopedia provides some general information on the Younger Dryas as a climate event:

```
http://en.wikipedia.org/wiki/Younger_Dryas
```

Authored by anonymous and possibly non-expert contributors, this article states:

```
The Younger Dryas stadial, also referred to as the Big Freeze,[1]
was a geologically brief (1,300 ± 70 years) cold climate period
between approximately 12,800 and 11,500 years ago (between 10,800
and 9500 BC).[2]
```

citing two journal articles [1] and [2], namely

```
Berger, WH (1990) The Younger Dryas cold spell – a quest
for causes. Global and Planetary Change 3 (3):219–237.
doi:10.1016/0921-8181(90)90018-8.

Muscheler R, et al. (2008) Tree rings and ice cores reveal 14C
calibration uncertainties during the Younger Dryas. Nature
Geoscience 1:263–267. doi:10.1038/ngeo128.
```

W.H. Berger is the author of the first article, which was published in the year 1990, entitled *The Younger Dryas cold spell – a quest for the causes*. The article appeared in Elsevier's *Global and Planetary Change* journal (Volume 3, Issue 3, pages 219–237). DOI stands for *Digital Object Identifier*, which is a character string that enables all kinds of electronic documents and objects to be identified. The second article, authored by R. Muscheler and his co-authors B. Kromer, S. Björck, A. Svensson, M. Friedrich, K.F. Kaiser, and J. Southon, appeared in the *Nature Geoscience* journal in the year 2008 (Volume 1, pages 263–267). Since the article has more than three authors, the co-authors were merged as *et al.*, which stands *et alii*, Latin for *and others*. The exact format, in particular the use of *et al.*, depends on the format required for citations by the particular journal.

We are not yet sure whether the two cited articles are the most relevant references for the Younger Dryas. The articles on Wikipedia are, however, always a good starting point for an in-depth literature research on a specific topic. Knowing that the Younger Dryas, in general, is a time interval between 12,800 and 11,500 years ago that was characterized by a cold climate at least points us in the right direction for our literature review on the topic.

We next use *Google Scholar* for a more detailed search of journal articles (Fig. 2.2). For this we enter *Younger Dryas Africa* in the search field of

```
http://scholar.google.com
```

and get a list of about 7,000 references. According to Google Scholar's webpage, it aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature. Browsing the list of articles retrieved by our search, however, it is not obvious how the articles are ranked, but the number of citations and the occurrence of the search terms in the title certainly seem to have had some influence. The top ten articles on the list are:

```
1. N Roberts, M Taieb, P Barker, B Damnati, M Icole, D Williamson
(1993) Timing of the Younger Dryas event in East Africa from
lake-level changes. Nature 366:146-148 [Citations: 74]
```

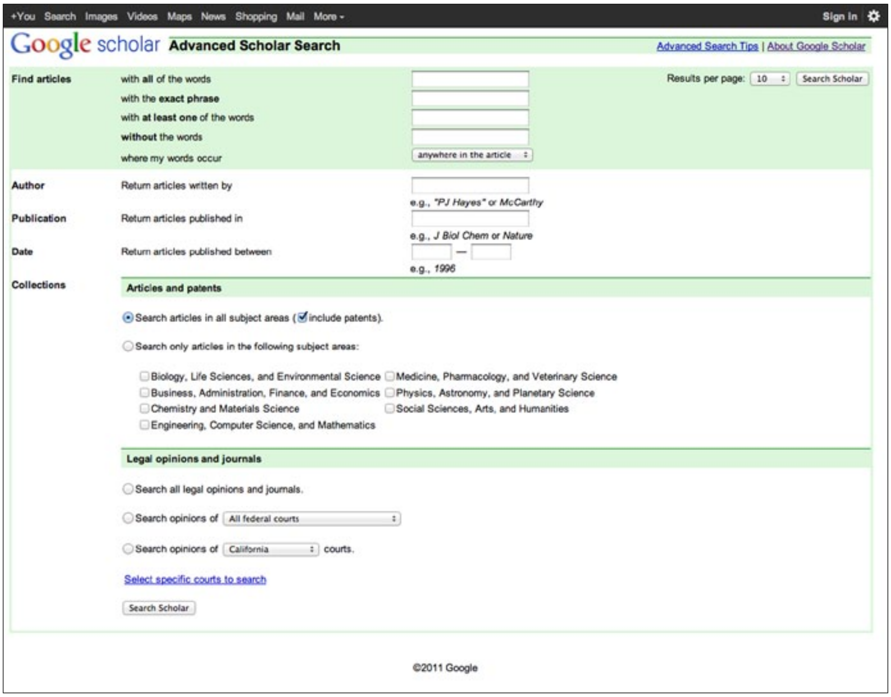

Fig. 2.2 Screenshot from the *Advanced Scholar Search* of *Google Scholar* (http://scholar.google.com), taken in early 2012. Google Scholar was introduced in 2004 and included online journals rather than only books.

2. Y Garcin, A Vincens, D Williamson, G Buchet, J Guiot (2007) Abrupt resumption of the African Monsoon at the Younger Dryas--Holocene climatic transition. Quaternary Science Reviews 26:690-704 [Citations: 26]

3. D Rind, D Peteet, W Broecker, A McIntyre, W Ruddiman (1986) The impact of cold North Atlantic sea surface temperatures on climate: implications for the Younger Dryas cooling (11–10 k). Climate Dynamics 1:3-33 [Citations: 126]

4. RB Alley (2000) The Younger Dryas cold interval as viewed from central Greenland, Quaternary Science Reviews 19:213-226 [Citations: 218]

5. P deMenocal, J Ortiz, T Guilderson, J Adkins, M Sarnthein, L Baker, M Yarusinski (2000) Abrupt onset and termination of the African Humid Period::: rapid climate responses to gradual insolation forcing. Quaternary Science Reviews 19:347-361 [Citations: 290]

6. D Verschuren, KR Laird , BF Cumming (2000) Rainfall and drought in equatorial east Africa during the past 1,100 years. Nature 403:410-414 [Citations: 302]

7. P De Deckker, T Corrège, J Head (1991) Late Pleistocene record of cyclic eolian activity from tropical Australia suggesting the Younger Dryas is not an unusual climatic event. Geology 19:602-605 [Citations: 33]

8. MR Talbot, ML Filippi, NB Jensen, JJ Tiercelin (2007) An abrupt change in the African monsoon at the end of the Younger Dryas. Geochemistry Geophysics Geosystems 8:Q03005 [Citations: 12]

9. W Zhou, MJ Head, Z An, P De Deckker, Z Liu, Y Liu, X Lu, D Donahue, AJT Jull, JW Beck (2001) Terrestrial evidence for a spatial structure of tropical-polar interconnections during the Younger Dryas episode. Earth and Planetary Science Letters 191:231-239 [Citations: 26]

10. H Renssen (1997) The global response to Younger Dryas boundary conditions in an AGCM simulation. Climate Dynamics 13:587-599 [Citations: 24]

Depending on our institution's journal subscriptions, we may be able to directly access the PDFs of these papers from the publisher's website by following the link provided in the Google Scholar list of articles. As we will see later, the first article, which is by N. Roberts, was indeed the earliest publication on the influence of the Younger Dryas cold event in Africa. Published in *Nature* in 1993, it was the result of a French-British collaboration in the 1980s and 1990s led by M. Taieb, and is fairly-well cited with 74 citations. The paper by Y. Garcin is one of many follow-up publications from this project, although the main conclusions are different from those in the Roberts paper (see Section 2.4). Articles no. 3, 4, 9 and 10 on the list provide a more global perspective on the influences of the Younger Dryas, including some comments on Africa, while article no. 7 is on climate change in Australia at that time. Article no. 8 provides some insights into the termination of the Younger Dryas in Africa, whereas the highly-cited article no. 5 is on climate change in Africa over the last 23,000 years, including the Younger Dryas. The well-cited paper by D. Verschuren, listed as no. 6, does not address the influence of the Younger Dryas at all as it only provides a history of climate change in Africa during the last 1,100 years. It does, however, refer to the first article on the topic (by N. Roberts) and for that reason appears in the list of relevant literature.

As we can see from this experiment, our Google literature search gives some reasonable results but with limitations. Whereas articles no. 1, 3, 4, 5, 9 and 10 may provide a good overview of the Younger Dryas in general, and of its influence in Africa in particular, articles 2 and 7 represent more specific articles with controversial discussions and conclusions rather than providing established knowledge on climatic connections between high and low latitudes at the last glacial termination. Article no. 6 is not at all relevant as it does not include any discussion on the Younger Dryas in its analysis of past climate change over the last millennium. The article by N. Roberts is cited by D. Verschuren and co-authors for noting the difference in the influence that solar heating had on moisture levels in Africa between glacial and interglacial periods, and the variations in moisture levels during the transition from the Medieval Warm Period to the Little Ice Age. Article no. 8 might be of interest for comparing Africa with other areas in the subtropics and tropics.

Searching with the Thomson Reuters ISI Web of Science

There are not many possibilities for influencing the search results in Google Scholar. The exact routine that ranks the journal articles for their relevance is not obvious. In our next experiment, we use the commercial Thomson Reuters ISI *Web of Science* literature search data base, which is very similar to Elsevier's Scopus service. For this we enter *Younger Dryas Africa* in the *topic* search field of

```
http://isiknowledge.com/
```

In early 2011 this yielded a list of about 100 articles, ranked by publication date (Fig. 2.3). In contrast to Google Scholar, the ISI Web of Science allows changes to be made to the mode of ranking, which can be, e.g., by the first author, the publication date, the number of citations, and the title of the journal. There is also the option to sort by relevance using a ranking system that considers how many of the search terms are found in each record, according to *Help*. Sorting by relevance, the article

```
Abell PI, Plug I. (2000) The Pleistocene/Holocene transition in
South Africa: evidence for the Younger Dryas event. Global and
Planetary Chance 26:173-179
```

which has only been cited 14 times during the last decade, ranked first in early 2011. Sorting the list of 100 articles on the Younger Dryas in Africa by the number of citations, the article

```
BarMatthews M, Ayalon A, Kaufman A (1997) Late quaternary
paleoclimate in the eastern Mediterranean region from stable
isotope analysis of speleothems at Soreq Cave, Israel. Quaternary
Research 47:155-168
```

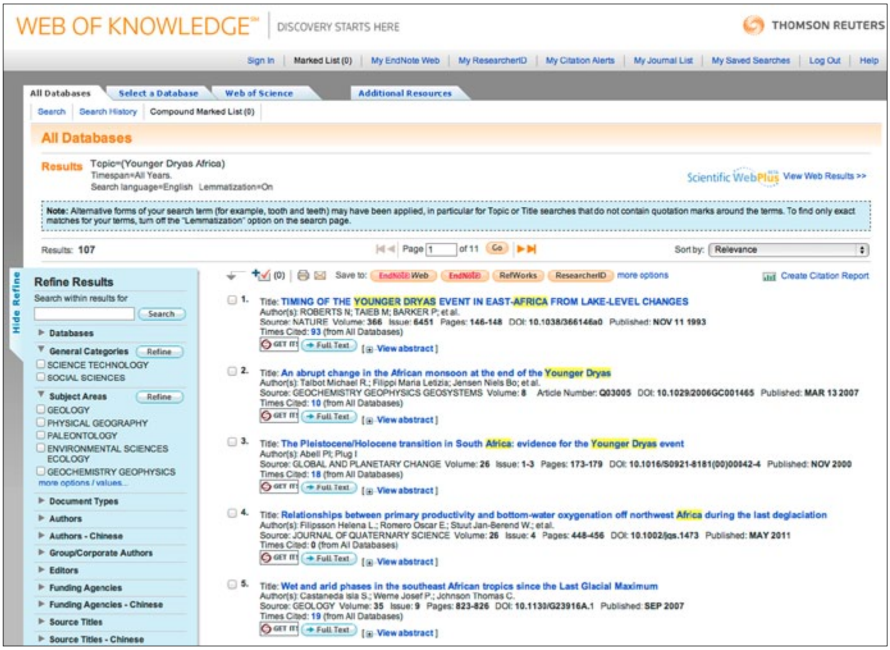

Fig. 2.3 Screenshot from Thomson Reuters' ISI Web of Knowledge (http://isiknowledge.com), showing results of a search for *Younger Dryas Africa*. The date of the screenshot is 9th January 2012, whereas the search experiment described in the text was performed in early 2011, which explains the slight difference in search results.

ranks first, with 185 citations. Sorting by publication date, the article

```
Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D
(1993) Timing of the Younger Dryas event in East Africa from
lake-level changes. Nature 366:146-148
```

which was ranked first in Google Scholar, is the oldest article of relevance on the Younger Dryas in Africa. Clicking on the number *90* opens a new webpage listing all other articles citing *Roberts et al. (1993).* After having identified this article as one of the earliest to provide evidence of a Younger Dryas influence in Africa, we also find

```
Williamson D, Taieb M, Damnati B, Icole M, Thouveny M (1993)
Equatorial extension of the Younger Dryas Event - Rock Magnetic
Evidence From Lake Magadi (Kenya). Global and Planetary Change
7:235-242
```

which was published in the same year. What other criteria would help us to identify the best article on the topic, rather than relying on simple counts of the number of times the search term *Younger Dryas Africa* is found in a record? In addition to the quality of the content, three measures are available to assist us in finding the best article on the topic; these measures are widely used but their validity is intensely debated.

The first measure for the relevance of an article is the *number of citations*, already mentioned when we used Google Scholar to search for literature. The number of citations counts all citations of an article in journals listed on the ISI Web of Science since the article's publication date. It therefore does not count the citations of an article in books or other media. In this regard the article by N. Roberts is more frequently cited (90 times based on ISI Web of Science, 74 times according to Google Scholar) than the article by D. Williamson (22 times). On the basis of this measure, the article by N. Roberts therefore appears to be more relevant, but let us explore some other criteria before we come to a final decision on the most relevant publication.

The second measure for the relevance of an article is the journal *impact factor* (IF). The journal impact factor can be obtained from the *Journal Citation Reports* (JCR) of the ISI Web of Science. The JCR Science Edition contains 7,300 journals in science and technology; the JCR Social Science Edition, not considered here, contains data from over 2,200 journals in social sciences. The JCR lists the number of articles published in the journal per year, the number of citations of that journal in articles published during the same year, the impact factor calculated from that year, and many other parameters. The IF calculated annually is the average number of citations of an article in journals during the two previous years. The IF measures only citations for journals that are listed on the ISI Web of Science and also only counts citations from those journals.

Since being introduced by Eugene Garfield, the founder of the ISI, the impact factor has been under severe criticism. For instance, the IF is highly dependent on the discipline. The journal *Nature*, in which the article by N. Roberts was published, and its subsidiary journal *Nature Geoscience*, are good examples of this dependency. The British journal *Nature* and its American counterpart *Science*, are the world's most respected scientific journals, with 2009 impact factors of 34.480 (Nature) and 29.747 (Science), i.e., articles in these journals were, on average, cited about 30 times during 2007 and 2008. The impact factor of these and other journals can be accessed from the ISI Web of Science webpage by clicking the *Additional Resources* tab, and then following the *Journal Citation Reports* link. There you can select the JCR edition and year, e.g., the *JCR Science Edition* for the year *2009*, and select the option *Search for a specific journal*. Enter *Nature* in the search field and you will get a table with the rank of the journal based on the IF, the total number of citations, the IF, and the five-year IF, as well as many other parameters for measuring the success of a journal.

Going back to the article by N. Roberts, cited 90 times since its publication date, we can analyze the citation record over the last two decades. Again, clicking on the number *90* (the number of citations) opens a new webpage listing all other articles citing *Roberts et al. (1993)*. There we click on *Analyze Results* and get a new webpage with various options for ranking and sorting the results. We then choose *Publication Year* to rank the records and *Selected field* to sort the results. This yields a chart displaying the citations per year where we can see that the article typically receives 5 to 7 citations per year, with a peak of 14 citations in the year 2000. Note that this citation rate is significantly lower than the journal's two-year average citation rate, i.e., its impact factor. However, Nature's citation rate in 1993 was different from that of today. The article by N. Roberts, which certainly made an important contribution to the understanding of the influence of high-latitude climate change in the tropics and provided the first evidence of the Younger Dryas cold event in East Africa, does not receive a large number of citations simply because it covers a relatively restricted field of research in which a low number of active scientists are involved.

The article's citation rate, however, is within the range of the IF of Nature Geoscience, which was launched as a disciplinary sub-journal of Nature in January 2008. According to the Journal Citation Reports, the 2009 IF for this journal is only 8.108, as compared to 34.480 for Nature. H. Langenberg, editor at Nature Geoscience, explained the difference between the IFs of the two journals as being due to the average citation rates per paper varying by a factor of five or six, depending on the scientific field, with the geosciences at the low end compared to other fields such as genetics and stem-cell research in which articles receive many more citations (Langenberg 2010). Browsing through the Journal Citation Reports for other well-respected journals in earth sciences, the third-best ranked journal behind Nature and Science is the *Proceedings of the National Academy of Sciences of the United States of America* (PNAS). PNAS is another interdisciplinary journal publishing original research; it has a 2009 IF of 9.432, and is therefore ranked slightly higher than Nature Geoscience, which is the most highly ranked specialized journal in our field.

Other highly ranked journals in earth sciences are Reviews of Geophysics (IF 8.021), Annual Reviews of Earth and Planetary Sciences (IF 7.581), and Earth-Science Reviews (IF 6.942). Below these, numerous journals are ranked with an IF in the 3 to 5 range, such as *Geology* (IF 4.368), *Quaternary Science Reviews* (IF 4.245), *Earth and Planetary Science Letters* (IF 4.062) and the more specialized journal *Paleoceanography* (IF 3.644). Almost 90 % of all earth science journals, however, fall within the IF < 3.000 range, presumably including articles that are not cited at all. The article by D. Williamson was published in *Global and Planetary Change*, a journal ranking at the lower end of the top 10 % of journals according to its IF of 3.272. Based on the IF of the two journals in which the articles about the Younger Dryas in Africa were published, the article by N. Roberts is again the better choice.

The third measure, which describes publication success, is the *h-factor* of the first author. This parameter, which is even more intensively debated than the previous two, was introduced by J.E. Hirsch to quantify an individual's scientific research output (Hirsch 2005). The index h is defined as the number of papers published by a scientist with citation number $\geq h$. According to J.E. Hirsch, the index h for a given individual should increase approximately linearly with time, i.e., scientists produce papers of similar quality at a steady rate over the course of their careers. Of course, the slope m of h over n years varies between different researchers, again depending not only on publication success but also on the discipline and on co-authorship regulations. On the basis of the slope m, J.E. Hirsch concludes that $m \approx 1$, i.e., h=20 after 20 years, characterizes a successful scientist, $m \approx 2$ characterizes an outstanding scientist, and $m \approx 3$ characterizes truly unique individuals.

To determine the h-factor for the two first authors, N. Roberts and D. Williamson, we click on the *Web of Science* tab on the ISI Web of

Knowledge webpage, and then enter *Roberts N* in the *Author* search field. We enable the *Science Citation Index Expanded* data base and disable the social sciences and arts and humanities citation indices, and then click *Search*. This yields a list of about 550 references, including those for other scientists with the same name as the author of the article on the Younger Dryas. We can *View Distinct Author Sets* by clicking on *Roberts N* above the list to exclude those authors from other fields such as clinical sciences, biology of reproduction, and librarianship, and keep the person authoring articles in the field of physical sciences. This selection reduces the total number of publications to 33, ranking the Nature article first with 90 citations. According to the Journal Citation Reports, the total number of citations of N. Roberts is 703, the average number of citations per item is 21.30, and the *h*-factor is 15. Since we are not sure whether we have indeed included all publications by N. Roberts, we return to the *Results* page and *Refine Results* by excluding *Subject Areas*. We then get a long list of subject areas, in which we activate the field of physical sciences, which includes geosciences, geography, geochemistry and geophysics, environmental sciences, multidisciplinary sciences, limnology, paleontology, oceanography, and archeology, and then click *Refine*.

This then yields a list of 54 publications, but it soon becomes evident that we have again included authors with similar names but from different fields, such as N. Roberts working on environmental trace-metal contamination. Excluding those publications from authors such as Noel Roberts, Nicholas J. Roberts and others, we end up with pretty much the same number of about 35 publications by the C. Neil Roberts that works at the University of Plymouth and led projects on paleoenvironmental change in Turkey after he stopped working in Africa, as we had before. Cross-checking the Journal Citation Reports with the (incomplete) list of publications provided on his webpage (http://www.plymouth.ac.uk/staff/cnroberts), we are very confident that we have arrived close to the true publication record of the N. Roberts who authored the article on the Younger Dryas in Africa. As I know from the author, however, both our search results and Neil Roberts' webpage are missing one important article, which was first-authored by Henry Lamb from the University of Aberystwyth and co-authored by N. Roberts:

```
Lamb, H.F., Gasse, F., Benkaddour, A., el-Hamouti, N., van der
Kaars, S., Perkins, W.T., Pearce, N.J., Roberts, C.N. (1995)
Relation between century-scale Holocene arid intervals in
tropical and temperate zones, Nature, 373, 134-137.
```

According to the Web of Science, this article has been cited more than 175 times. This article is missing from our search list because in the list of authors Roberts is listed as *C.N. Roberts*. Neil Roberts is known by his middle name Neil, but some publications, such as the one published in Nature, also add his first initial C. Note that the same author has also published the very popular textbook

```
Roberts N (1998) The Holocene: an environmental history – 2nd
edition. Blackwell, Oxford
```

which is not mentioned at all in the ISI Citation Reports. Not considering books is of course a significant weakness of the ISI Web of Science. Books do not contribute to any of the measures of the scientific or publication success of a researcher. Recently, however, Thomson Reuters has launched a *Book Citation Index* for the Web of Knowledge, which can help to find the most relevant books on a topic, using similar parameters to those used for journal articles to measure the contribution that books make a particular discipline.

How does D. Williamson's *h*-factor compare with that of N. Roberts? The first search returns about 670 references, but can *View Distinct Author Sets* reduce the articles to only those published by the French researcher D. Williamson? By excluding all publications related to life sciences we reduce the list to about 43 articles, with a total of 1,016 citations, an average citation rate of 24.19, and an *h*-factor of 20. David Williamson, who led the paleomagnetics laboratory at the CEREGE in Aix-en-Provence, obviously benefitted from being a laboratory manager. He is co-author of many highly-cited paleoclimate publications that used paleomagnetics, but also first-authored several well-cited articles on mineral-magnetic proxies in paleoclimate research.

Please note that he is also a co-author on N. Roberts' Nature paper, to which he probably contributed the magnetics record shown in Figure 2 of the article. Comparing the authors listed for both papers

```
Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D
(1993) Timing of the Younger Dryas event in East Africa from
lake-level changes. Nature 366:146-148

Williamson D, Taieb M, Damnati B, Icole M, Thouveny M (1993)
Equatorial extension of the Younger Dryas Event - Rock Magnetic
Evidence From Lake Magadi (Kenya). Global and Planetary Change
7:235-242
```

reveals that other authors are also listed for both articles. An Internet search for some of these authors leads to the conclusion that both N. Roberts and D. Williamson worked under the guidance of Maurice Taieb, who was the leader of the paleolimnological project at Lake Magadi in Southern Kenya.

M. Taieb is a French geologist and paleoanthropologist who worked in Africa for over fifty years, and was involved in the discovery of the fossil fragments of *Lucy* in the early 1970s.

Because of the leadership and co-authorship of M. Taieb on both publications, and because D. Williamson's article is on the paleomagnetic record only whereas N. Roberts' paper includes a discussion of other climate proxies such as diatoms (silica algae), organic carbon, and calcite, and also because Nature is a far better reference than Global and Planetary Change, we come to the conclusion that the article by N. Roberts is, in general, likely to be the more useful publication on the Younger Dryas in Africa. An analysis of our literature research, however, clearly shows how difficult it is to evaluate the relevance of a specific article on the basis of bibliometric measures such as journal impact factors, citation indices, and h-factors. As the demonstration clearly shows, it often requires a lot of background knowledge to find the most relevant papers on a particular topic. Of course having found the most important journal article should not prevent you from also reading some of the less relevant papers. These may well contain interesting details in the text that are more relevant to your own particular research than to that of the citing authors.

2.4 Extracting the Relevant Information from Literature

Having identified N. Roberts' article as the most relevant reference to use as a starting point for research into the influence of the Younger Dryas event in East Africa, we now obtain the electronic version of this article , which was published in Nature (http://www.nature.com) in 1993. The journal's webpage has a search field in which we can enter the terms *Younger Dryas East Africa* and get about 40 references as a result. We see that there are even older articles mentioning the possible influence of the Younger Dryas in Africa based, for example, on analysis of the sapropels in the Mediterranean Sea as indicators of monsoonal rainfall intensity:

```
Rossignol-Strick M, Nesteroff W, Olive P, Vergnaud-Grazzini C
(1982) After the deluge: Mediterranean stagnation and sapropel
formation. Nature 295:105-110
```

The article by N. Roberts also shows up in the 40 references, listed as a *Letter to Nature* type of article. A Letter in Nature is four pages long and has no more than 30 references, whereas an *Article* can be up to five pages long and has up to 50 references; *Brief Communications* are even shorter than Letters.

```
Timing of the Younger Dryas event in East Africa from lake-level
changes
Nell Roberts, Maurice Taieb, Philip Barker, Brahlm Damnati,
Michel Icole, David Williamson
Nature 366, 146-148 (11 November 1993) doi:10.1038/366146a0
Letter
Abstract | PDF | Rights and permissions | Save this link
```

Please note that the author's first name Neil is actually misspelled Nell in the journal's article data base. The person who typed the reference into the data base presumably used the scanned version of the article, which we download later from the journal's webpage, in which the first name indeed looks like Nell rather than Neil.

As an alternative, we can browse the contents of the journal if we know the volume and page numbers for the article. On the main page of Nature, we click on *Publications A–Z* under the *Inside nature.com* header. There we choose the journal *Nature* from the list of publications, then *Archive* from the menu bar of the journal's webpage, select the year *1993*, volume *11 November 1993 366 6451 95–188* and get the contents list for the issue of the journal that contains the article by N. Roberts.

```
Timing of the Younger Dryas event in East Africa from lake-level
changes 146
NELL ROBERTS, MAURICE TAIEB, PHILIP BARKER, BRAHLM DAMNATI,
MICHEL ICOLE & DAVID WILLIAMSON
doi:10.1038/366146a0
First paragraph & References | PDF (355K)
```

The reference contains the title, the page number (*146*), the authors, the DOI (*0.1038/366146a0*), the first paragraph and references, and a link to the full article. The title, list of authors, abstract and references, are freely available without subscription to the journal (Fig. 2.4). The full article, however, is available by subscription only. In many cases, however, the title and the abstract already contain the most important information from the article. We are interested to learn about the influence of the Younger Dryas in East Africa. If the Younger Dryas did have an influence, how did it affect the climate in East Africa and when? The title

```
Timing of the Younger Dryas event in East Africa from lake-level
changes
```

suggests that there was indeed an influence of the Younger Dryas in East Africa, and that this influence has been identified in the water-level record of Lake Magadi in Kenya. Furthermore, the title indicates that the article discusses the timing of the Younger Dryas influence in the region. It does not,

npg nature publishing group

nature.com | about npg | news@nature.com | naturejobs | natureevents | help | site index

nature

my account | e-alerts | subscribe | register

SEARCH JOURNAL Go

Monday 09 January 2012

Journal Home
Current Issue
AOP
Archive

THIS ARTICLE -
Download PDF
References
Export citation
Export references
Send to a friend
More articles like this
Table of Contents
< Previous | Next >

letters to nature

Nature **366**, 146 - 148 (11 November 1993); doi:10.1038/366146a0

Timing of the Younger Dryas event in East Africa from lake-level changes

NELL ROBERTS[& AST;], MAURICE TAIEB[& DAGGER;], PHILIP BARKER[& AST;], BRAHLM DAMNATI[& DAGGER;], MICHEL ICOLE[& DAGGER;] & DAVID WILLIAMSON[& DAGGER;]

[& ast;] Department of Geography, Loughborough University of Technology, Leicester, LE11 3TU, UK
[& dagger;] CNRS Laboratoire de G& eacute;ologic du Quaternaire, Case 907, 13288 Marseille, Cedex 9, France

THE last deglaciation was interrupted by an abrupt cooling event, the Younger Dryas, at 11,000& ndash;10,000 yr BP (uncalibrated radiocarbon timescale)[1]. Originally recognized in climate records from northwest Europe, the Younger Dryas has now been identified in marine and ice-core records worldwide[2& ndash;6]. In the tropics, a broadly contemporaneous change in climate is recorded by decreases in water levels and increased salinity of lakes[7& ndash;9,14], indicating a period of arid climate caused by a reduction in ocean-to-land moisture flux. The exact timing of these changes in relation to the Younger Dryas event in high-latitude records has remained unclear, however. Here we present climate records based on analyses of diatom assemblages, geochemistry and magnetic mineralogy of radiocarbon-dated sequences of laminated lake sediments from Lake Magadi in the East African rift. These records provide a detailed record of climate change in lowland equatorial Africa throughout the last deglaciation (12,800& ndash;10,000 ^{14}C yr BP). We find that lake-level and humidity maxima coincide with the most rapid phases of ice melting in the Northern Hemisphere, and that the climate changes, including the Younger Dryas event, were synchronous at low and high latitudes. Thus, the effects of abrupt climate change appear to be felt at both high and low latitudes without a significant time lag.

References

1. Zahn, R. *Nature* **356**, 744–746 (1992).
2. Fairbanks, R. G. *Paleoceanography* **5**, 937–948 (1990).
3. Dansgaard, W., White, J. W. C. & Johnsen, S. J. *Nature* **339**, 532–534 (1989).
4. Lehmann, S. J. & Keigwin, L. D. *Nature* **356**, 757–762 (1992).
5. Taylor, K. C. *et al.* *Nature* **361**, 432–436 (1993).
6. Bard, E. *et al.* *Radiocarbon* **35**, 191–199 (1993).
7. Street-Perrott, F. A. & Perrott, A. *Nature* **346**, 601–612 (1990).
8. Gasse, F., Téhet, R., Durand, A., Gibert, E. & Fontes, J.-C. *Nature* **346**, 141–146 (1990).
9. Talbot, M. R. & Johannessen, T. *Earth Planet. Sci. Lett.* **110**, 23–37 (1992).
10. Jones, B. F., Eugster, H. P. & Rettig, S. L. *Geochim. cosmochim. Acta* **41**, 53–72 (1977).
11. Baker, B. H. *Geology of the Magadi Area* (Geol. Surv. Kenya, **42**, 1958).

Fig. 2.4 Screenshot from *Nature* (http://nature.com) showing the freely-available title, list of authors, abstract, and references from the paper by Roberts et al. (1993). The date of the screenshot is 9th January 2012 whereas the search experiment described in the text was performed in early 2011, which explains the slight difference in search results.

however, contain information on the nature of the influence (e.g., whether it resulted in a drier or wetter climate) or the exact timing of the resulting climate change in East Africa.

We proceed to the abstract of the article, which typically provides a summary, in most cases without reference to previous articles by other researchers. For articles in Nature, however, the abstract is a fully-referenced first paragraph, in bold format.

> THE last deglaciation was interrupted by an abrupt cooling event, the Younger Dryas, at 11,000–10,000 yr BP (uncalibrated radiocarbon timescale)[1]. Originally recognized in climate records from northwest Europe, the Younger Dryas has now been identified in marine and ice-core records worldwide[2–6]. In the tropics, a broadly contemporaneous change in climate is recorded by decreases in water levels and increased salinity of lakes[7–9,14], indicating a period of arid climate caused by a reduction in ocean-to-land moisture flux. The exact timing of these changes

```
in relation to the Younger Dryas event in high-latitude records
has remained unclear, however. Here we present climate records
based on analyses of diatom assemblages, geochemistry and
magnetic mineralogy of radiocarbon-dated sequences of laminated
lake sediments from Lake Magadi in the East African rift. These
records provide a detailed record of climate change in lowland
equatorial Africa throughout the last deglaciation (12,800–10,000
14C yr BP). We find that lake-level and humidity maxima coincide
with the most rapid phases of ice melting in the Northern
Hemisphere, and that the climate changes, including the Younger
Dryas event, were synchronous at low and high latitudes. Thus,
the effects of abrupt climate change appear to be felt at both
high and low latitudes without a significant time lag.
```

The small superscript letters refer to articles in the list of references. As an example, reference 1 after the first sentence refers to

```
1. Zahn, R. Nature 356, 744–746 (1992).
```

which is the article by R. Zahn

```
Zahn, R (1992) Deep ocean circulation puzzle. Nature 356:744-746
```

on the Younger Dryas event in general. This article is cited after the first sentence of the article by N. Roberts. This sentence introduces the phenomenon of the Younger Dryas as an abrupt cooling event between 11,000 and 10,000 ^{14}C yr BP during the last deglaciation, i.e., during the transition from the last glacial to the present interglacial. The second and third sentences state that the Younger Dryas was a global event that resulted in a drier climate in the tropics. The exact timing of tropical climate change during the Younger Dryas event, however, remains unclear.

Following an outline of the scientific question to be investigated, the second half of an abstract or first paragraph typically starts with the statement *Here we present* ..., introducing the approach, results, and conclusions of the actual research presented in the article. In our example, this reads as follows:

```
Here we present climate records based on analyses of diatom
assemblages, geochemistry and magnetic mineralogy of radiocarbon-
dated sequences of laminated lake sediments from Lake Magadi in
the East African rift.
```

from which the reader learns that the article is about a lake-sediment record from a lake located in the East African rift, and that it includes the analysis of fossil silica algae (diatom) assemblages, as well as chemical and physical sediment parameters. The last three sentences of the abstract

```
These records provide a detailed record of climate change in
lowland equatorial Africa throughout the last deglaciation
(12,800-10,000 14C yr BP). We find that lake-level and humidity
maxima coincide with the most rapid phases of ice melting in the
Northern Hemisphere, and that the climate changes, including the
Younger Dryas event, were synchronous at low and high latitudes.
```

basically list the results – more specifically, the age of the sediments and therefore the time interval contained in the lake-level record, and the information that the climate change coincided with the high-latitude deglaciation – but it does not contain information on precisely how the Younger Dryas affected the climate in the region. Remembering that the first part of the paragraph said that the Younger Dryas was generally dry in the tropics and reading here that humidity maxima coincided with the ice melting, we could conclude that the Younger Dryas cold event was probably dry. The ultimate conclusion of the article is then given in the final sentence of the abstract

```
Thus, the effects of abrupt climate change appear to be felt at
both high and low latitudes without a significant time lag.
```

essentially proposing a link between high and low latitudes based on the correlation of climate records without any time lag.

Reading the title and the abstract (or the first paragraph) of the article has taken us only one or two minutes, but we have already extracted the most important scientific information, despite the minor weakness in this particular article of not being precise about the exact nature of the influence that the Younger Dryas had in East Africa. In our experiment this weakness necessitates an analysis of the full article, or at least parts of it. The article is in a Portable Document Format (PDF). This format was designed by Adobe Systems (http://www.adobe.com/) as self-contained cross-platform document. PDF files contain the complete formatting of vector illustrations, raster images and text, or a combination of all these, and includes all necessary fonts. These files are highly compressed, allowing a fast Internet download. Adobe Systems provides the Acrobat Reader free-of-charge for all computer platforms, for reading PDF files. As we will see later, more recent articles are in a vector format, allowing text fragments and graphics to be extracted and incorporated into lectures, or into books such as this, provided that the copyright of the electronic materials is respected. The article by N. Roberts, published in 1993, is a scanned version of the original printed journal article and therefore only available in raster format.

The article is approximately 2.5 pages long and therefore represents a relatively short piece of text. Nature, Science and the Proceedings of the

National Academy of Sciences (PNAS) typically publish articles with a maximum length of four pages, whereas most other journals allow papers of 10 printed pages, or even more. Most journals also allow the authors to publish an electronic supplement to the printed article, including more text, tables, figures and references. Reading an article of such length requires efficient reading, and we therefore need to know the most likely location of the most relevant information in an article. Fortunately, research articles share a common structure comprising the title and abstract, introduction, setting and methods, results, discussion, and conclusions. Differences exist between journals, however, and Nature, Science, and PNAS each structure articles in a slightly different way. The overall structure of the introduction, methods, results, and discussion/conclusions also applies to these journals but the sections do not have headers. Details about the geographic and geologic setting of the study area, an overview of the methods used in the analysis, and results presented as tables and text are usually published separately from the actual article as electronic supplementary information.

Apart from the information contained in the title and abstract/first paragraph, the most important information typically occurs in the conclusion or the last paragraph of the article, and in the last figure. The last figure of an article often summarizes the results or places a new data set in a larger context by, for instance, correlating the new data with established records of past climate change, or with any other data set from a different location and from a different research project. In our example, Figure 3 compares the water-level record from Lake Magadi in Kenya with other climate records, namely the oxygen-isotope curve from Rotsee in Switzerland by Lotter et al. (1989) and winter temperatures for Britain reconstructed from beetle remains by Atkinson et al. (1987):

```
Lotter A, Zbinden H (1989) Late-Glacial pollen analysis, oxygen-
isotope record, and radiocarbon stratigraphy from Rotsee
(Lucerne), Central Swiss Plateau. Ecologae geol. Helv. 82:191-202

Atkinson TC, Briffa K, Coope GR (1987) Seasonal temperatures in
Britain during the past 22,000 years, reconstructed using beetle
remains. Nature 325:587-592.
```

The figure clearly shows that the water level dropped by about 40 meters during the Younger Dryas, suggesting that tropical East African climate became drier when the oxygen isotopes and beetle assemblages indicate colder temperatures in the mid latitudes. The relevant text in the last paragraph, which typically contains the conclusion of the article, also says that

```
Conditions became warmer and wetter around 12,700 14C yr BP,
```

```
cooler and drier at 11,100-10,700 14C yr BP, with renewed warming
and wetting ~10,000 14C yr BP.
```

and then comments on the timing of this dry event in the tropics, relative to the Younger Dryas cold event

```
The similarity of these and other records indicates that this
climate oscillations was not only a global event, but also
that there were no significant time lags between tropical and
temperate components of the climate system during the last
deglaciation.
```

Having read the most important sections of the article, we read the remaining text and figures in more detail if additional information is required for the article to be used as a reference in our own manuscript (see Chapter 12). In that case, we cite the article in the text as

```
Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D
(1993) Timing of the Younger Dryas event in East Africa from
lake-level changes. Nature 366:146-148
```

in the reference list of our article. The format used for references can vary significantly depending on the requirements of the particular journal, provided in their *Guidelines for Authors*. In the text itself, the article is cited as

```
Roberts et al. (1993) have presented ...
```

where, as explained previously, *et al.* stands *et alii*, which is Latin for *and others*, since the article has more than two authors. If there are two authors only, e.g., N. Roberts and M. Taieb, the reference is cited as

```
Roberts and Taieb (1993) have presented ...
```

or

```
Roberts & Taieb (1993) have presented ...
```

In most cases, a reference occurs at the end of a statement rather than being incorporated in the text itself. The references then reads

```
The Younger Dryas in Africa was dry (Roberts et al., 1993).
```

or in the case of two authors

```
The Younger Dryas in Africa was dry (Roberts and Taieb, 1993).
```

or

```
The Younger Dryas in Africa was dry (Roberts & Taieb, 1993).
```

depending on the format that each journal defines in its guidelines. In many cases more than one reference is cited after a scientific statement, such as

```
The Younger Dryas in Africa was dry (Williamson et al., 1993;
Roberts et al., 1993).
```

with the references listed chronologically, with the oldest first. Bearing in mind the space limitations of most journals, however, we need to decide which references to actually cite for a statement, which brings us back to the original topic of this section. As an example, the maximum number of references in the main text in Nature and Science is thirty or fifty citations, depending on the type of paper, whereas there is no limit to the number of references in the electronic supplement.

2.5 Extracting Text, Data and Graphs from Literature

Having identified the most relevant article on a specific topic, we now wish to extract scientific information from that article, not only by reading and remembering that information, but also by taking text and figures from the electronic document for incorporation into a poster, a presentation slide or any other type of presentation, of course within the limits of copyright regulations. Most posters and presentations contain original research and the figures are therefore newly-designed by the researcher. In the bachelor's level course that inspired this book, however, the students were not conducting their own research and therefore prepared posters, abstracts, and presentations based on previously published work by other researchers. University lecturers, however, cover broader topics, usually complemented by their own original research.

The level of modification that is possible for text and figures extracted from a published manuscript or book, prior to its incorporation into a presentation, depends on the way the document has been electronically saved. Computer graphics are generally stored and processed as either vector or raster data. Most of the data types encountered in the following chapter on processing geoscientific information are vector data, i.e., points, lines and polygons. Drainage networks, the outlines of geologic units, sampling locations and topographic contours are all examples of vector data. In Chapter 6, coastlines are stored in a vector format while bathymetric and topographic data are saved in a raster format. Vector and raster data are often combined in a single data set, for example in order to display the course of a river on a satellite image. Raster data are often converted to vector data by digitizing points, lines or polygons. Conversely, vector data are sometimes transformed to raster data, which requires a degree of discretization and generalization. Images are generally presented as raster data, i.e., as a 2D array of color intensities. Images are everywhere in geosciences. Field geologists use aerial photographs and satellite images to identify lithologic units, tectonic structures, landslides, and other features within a study area. Petrologists and micropaleontologists use microscope images of minerals and fossils from either an optical microscope or a scanning electron microscope.

The PostScript (PS) format is used to store scalable vector graphics, raster images, and text. It was developed by John Warnock at PARC, the Xerox research institute. Warnock was also co-founder of Adobe Systems, where the EPS format was created. The PostScript vector format would have never become an industry standard without Apple Computers. In 1985 Apple needed a typesetter-quality controller for the new Apple LaserWriter printer and the Macintosh operating system, and adopted the PostScript format. The third partner in the history of PostScript was the Aldus company, the developer of the PageMaker software and now a part of Adobe Systems. The combination of Aldus PageMaker software, the PS format, and the Apple LaserWriter printer led to the creation of Desktop Publishing. The Encapsulated PostScript (EPS) format was then developed by Adobe Systems as a standard file format for importing and exporting PS files. Whereas a PS file is generally a single-page format containing either an illustration or a text, the purpose of an EPS file is to allow the inclusion of other pages, i.e., it is a file that can contain any combination of text, graphics and images. The Portable Document Format (PDF) designed by Adobe Systems in 1993 is now a true, self-contained, cross-platform document.

Most journal articles and electronic books are today available as PDF files, which may include editable text in a vector format, vector graphics, and raster images. The PDF file of the article by Neil Roberts, however, is a scanned document in raster format, as is revealed by zooming in on the text and figures. It is therefore not possible to extract text and figures in a vector format for further modification. Text recognition, automated vectorization, or manual digitizing can partly solve this problem, and pixel graphics could also be extracted without further editing and overlaid by vector additions in a separate layer (see Chapter 9).

More recent articles published in Nature, Science, or any other journal are available in a vector format. As an example, we use my own article

```
Trauth MH, Maslin M, Deino MA, Deino A, Strecker MR (2005) Late
Cenozoic Moisture History of East Africa. Science 203:2051-2053.
```

published in 2005 in Science (http://www.sciencemag.org). As with Nature, the Science webpage has a search field in which we can enter the title *Late Cenozoic Moisture History of East Africa* and quickly find the article.

```
Late Cenozoic Moisture History of East Africa
Martin H. Trauth, Mark A. Maslin, Alan Deino, and Manfred R.
Strecker
Science 23 September 2005: 2051-2053.Published online 18 August
2005 [DOI:10.1126/science.1112964]
Abstract Full Text Full Text (PDF) Supporting Online Material
```

As well as the article being available as the abstract only, as the full text online, or as the full text in a PDF, it also includes supporting online material. Once again, the abstract is freely available:

```
Lake sediments in 10 Ethiopian, Kenyan, and Tanzanian rift basins
suggest that there were three humid periods at 2.7 to 2.5 million
years ago (Ma), 1.9 to 1.7 Ma, and 1.1 to 0.9 Ma, superimposed on
the longer-term aridification of East Africa. These humid periods
correlate with increased aridity in northwest and northeast
Africa and with substantial global climate transitions. These
episodes could have had important impacts on the speciation
and dispersal of mammals and hominins, because a number of key
events, such as the origin of the genus Homo and the evolution of
the species Homo erectus, took place in this region during that
time.
```

It provides the most important information and conclusions from the article. Access to the full text, whether online or as a PDF file, requires a subscription to the journal, or alternatively it can be accessed via a pay-per-view option. Having downloaded the PDF file, we quickly notice that it is in a true vector format since, for instance, we are able to mark and copy text fragments and to paste them into word processing software. Differences exist, however, in the way a PDF reader deals with the text, in particular with respect to indented text, special characters, and paragraph styles. Using Adobe Acrobat on a computer running the Mac OS X operating system, for instance, right justification in the linefeeds of text is interpreted as a paragraph end marker at the end of each justified line, both in the abstracts and in the main text. For instance, the first paragraph of the main text appears as

```
Recent investigations of both terrestrial and
marine paleoclimate archives have led to a
concerted debate regarding the nature of Late
Cenozoic environmental changes in East Africa
and their influence on mammalian and hominin
evolution (1-3).
```

In this case, the paragraph end markers have to be removed either manually or using search and replace, prior to making further modifications of the text. Making the same experiment using Apple Preview avoids this problem, resulting in

```
Recent investigations of both terrestrial and marine paleoclimate
archives have led to a concerted debate regarding the nature of
Late Cenozoic environmental changes in East Africa and their
influence on mammalian and hominin evolution (1–3).
```

Another problem when extracting text fragments from an article is the incorrect transfer of subscripts, superscripts, and other paragraph styles. For instance, the term *^{40}Ar/^{39}Ar age calibration* from the second paragraph of the main text becomes

```
(...)
with intercalated volcaniclastic deposits that
permit high-precision 40Ar/39Ar age calibration
of lake-level highstands (5, 6) (Fig. 2).
(...)
```

in which the numbers *40* and *39* are transferred as a smaller font but not as superscripts, with both Adobe Acrobat and Apple Preview. Copying this paragraph from the PDF viewed with Apple Preview again interprets the linefeeds correctly but puts hyphens and a space within words at linefeeds, such as *calibra- tion* in the text below:

```
Although much smaller than the lakes in the western branch and
often subaerially exposed, these basins host a rich sedimentary
record, with intercalated volcaniclastic deposits that permit
high-precision 40Ar/39Ar age calibra- tion of lake-level
highstands (5, 6) (Fig. 2).
```

These errors again need to be corrected either manually or by using the search and replace option in text processing software.

In the next experiment we try to import the two figures included in the article by Trauth et al. (2005). We use the import feature of vector graphics software such as the open source *Inkscape* software, or the commercial *Adobe Illustrator* software. Opening the first page of the PDF file reveals that Figure 1, which shows a map of East Africa, is indeed in a vector format. After importing the figure we can click on individual elements such as polygons or text, and modify colors, line thicknesses and font sizes (see Chapter 8). However, importing the second page of the article, which includes Figure 2, reveals that this figure is in a pixel format, preventing any further modification. We suspect that the publisher may include data graphs, such as the compilation of lake records presented in Figure 2, in a pixel format in order to prohibit unwanted modification or manipulation of the author's original graphs. This figure, however, could be imported into pixel software such as *Gimp* or *Adobe Photoshop* and modified with the tools available for editing photos and other pixel graphics. Furthermore, the pixel graphics can be imported into vector graphics software and overlaid with vector elements such as polygons and text.

2.6 Organizing Literature in a Computer

Familiarization with a new topic at the beginning of a research project requires the collection and reading of a large number of articles, books, and reports, and their organization as PDF files on a computer. To organize the collection of references and electronic documents, we need literature or bibliographic management software. A large number of both open source and commercial software products are available for managing your literature. We use the open source *BibDesk* (http://bibdesk.sourceforge.net) software for Mac OS X as well as the commercial *EndNote* (http://www.endnote.com) and *Papers* (http://mekentosj.com/papers), which are representative examples of the typical design of literature management software products.

Bibliographic management software typically consists of a searchable reference data base, often linked with online data bases such as Google Scholar and Thomson Reuters ISI Web of Science, and a file directory service for managing the PDF files on a hard drive. Most software packages can be integrated into word processing software such as *Microsoft Word* or *Apple Pages*, and reference lists are then generated automatically in a specific journal's layout from references included in the manuscript text. Alternatively, software based on the *BibTeX* standard, such as *BibDesk* for the Unix-based Mac OS X, or *KBibTeX* for Linux or Unix, uses *LaTeX* to prepare formatted literature lists.

The most popular commercial software for organizing reference data bases is EndNote. The software was first released in 1988 and is therefore one of the oldest reference management tools, designed before Adobe had released the PDF. The software was therefore primarily created to manage citation libraries rather than to organize PDFs on a hard disk. In the early days, the user had to type references into the data base manually, using an input mask with fields for authors, title, year and other kinds of information related to the article. Subsequently, most literature search data bases allowed users to export the search results in a format that could be directly imported into EndNote. Today, EndNote can also be used to organize PDF files. The software can be integrated with word processors and automatically generates reference lists in the format required by a specific journal.

Most universities provide campus licenses that can be acquired at a reasonable price. EndNote is available for Microsoft Windows and Mac OS X operation systems.

Launching the EndNote software brings up a single multi-panel user interface that includes a panel for browsing the library (left panel), a list of all references in the data base (upper right panel) and a panel with tabs for *Preview*, *Search*, and *Quick Edit* (lower right panel). Since EndNote is linked to Thomson Reuters ISI Web of Science, we can again start a search for the keywords *Younger Dryas East Africa*. We select *Web of Science* from *Online Search* in the *My Library* panel, select Title as the criterion for the search in the *Online Search* tab, type the keywords *younger dryas east africa* in the search field, and then press *Search*. This yields a list of several references including the article by N. Roberts, which is the earliest article listed. Double-clicking the article in the list opens a new window displaying the data sheet for this particular reference with the categories *Author*, *Year*, *Journal*, *Volume*, *Issue*, *Pages*, and so forth. The data sheet also includes the abstract, number of citations, and the author's address. The *Reference* menu of the software includes a *Find full text* feature that automatically includes the PDF of the article if the user has an active subscription to Nature. Having included the reference and the file of the article in the data base, we can export the reference in a format of the user's choice for inclusion in a manuscript requiring a specific format. As an example, the reference to the article by N. Roberts can be previewed and exported in the Nature format:

1 Roberts, N. *et al.* Timing of the Younger Dryas Event in East-Africa from Lake-Level Changes. *Nature* **366**, 146–148 (1993).

Switching to the format of Elsevier's Earth and Planetary Science Letters journal creates the reference

Roberts, N., Taieb, M., Barker, P., Damnati, B., Icole, M., Williamson, D., 1993. Timing of the Younger Dryas Event in East-Africa from Lake-Level Changes. *Nature* **366**, 146–148.

The format for the Proceedings of the National Academy of Sciences of the United States of America (PNAS) is as follows:

1. Roberts N, *et al.* (1993) Timing of the Younger Dryas Event in East-Africa from Lake-Level Changes. *Nature* **366**(6451):146–148.

The user interface and features of the free BibDesk software are very similar to those of EndNote. The open source software is a BibTeX front end that uses LaTeX to export formatted reference lists. BibDesk (http://bibdesk.sourceforge.net) is only available for Mac OS X; it was first released in 2002 and can be downloaded as a separate product, without LaTeX. A recommended alternative is to download the software as part of the *MacTeX* bundle (http://www.tug.org/mactex), which includes the official standard distribution of *TeX* for Macs. Installing MacTeX allows formatted reference lists to be prepared without requiring a separate LaTeX editor. We can select *Web of Science* from the *Searches* menu, causing a search panel to open in which entering the keywords *Younger Dryas East Africa* yields no results. Since there are no further options available for the search fields, we need to consult the ISI Web of Science help page for further instructions. We learn that we have to include the search terms manually and modify our search into *TS=Younger Dryas East Africa*, where *TS* stands for *Topic Search*. This again yields the papers by D. Williamson and N. Roberts as the earliest references on the list. The data sheet for the article by N. Roberts includes a link to the ISI Web of Science webpage but no direct link to the PDF of the article on the webpage of the journal. The *Preferences* menu provides numerous BibTeX styles with which to format the reference list for export. Choosing *apalike* as an example format and then clicking the *TeX Preview* button from the user interface results in the formatted output

ROBERTS, N., TAIEB, M., BARKER, P., DAMNATI, B., ICOLE, M., and WILLIAMSON, D. (1993). Timing of the younger dryas event in east-africa from lake-level changes. *Nature*, 366(6451):146–148.

which follows the capitalized format used in the ISI Web of Science data base for the authors' names. The commercial EndNote software clearly corrects this mistake and results in the output seen previously.

An award-winning commercial alternative to EndNote and BibDesk for Mac OS X is Papers by *Mekentosj* from The Netherlands (Fig. 2.5). The software was released in 2007 and is currently available as Version 2, which provides numerous additional features to those of the first version. In contrast to EndNote and its open source counterpart BibDesk, this software was designed from the start to maintain large libraries of PDF documents comprising journal articles, books, theses, and other types of scientific literature. As with the other tools, Papers offers repository searches using the ISI Web of Science and other data bases, editing of the file metadata using data sheets, full-screen reading, and commenting, as well as multiple import and export features. A great advantage of the software is the very straight-forward approach to renaming and reorganizing the PDF files on a hard drive. As soon as a PDF file has been imported into the software, it is renamed in a way defined by the user. In our example, for instance, the PDF file of the article

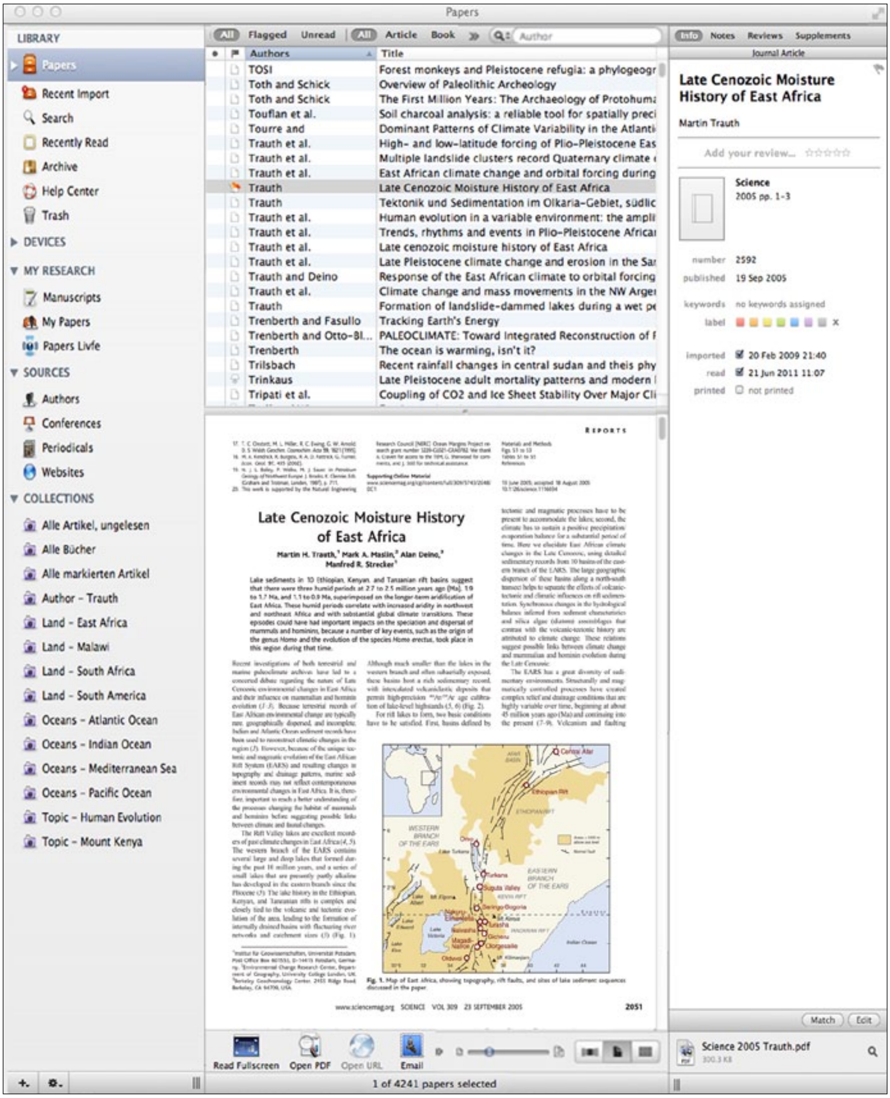

Fig. 2.5 Screenshot from *Papers 2.1* by *Mekentosj* (http://www.mekentosj.com), an award-winning bibliographic management software for organizing a collection of references and electronic documents.

by N. Roberts is renamed *Nature 1993 Roberts.pdf* and copied into the *1993* folder and a *Roberts* subfolder in the *Papers* directory, in the user's home directory. This contrasts with the way EndNote organizes its Library as it first creates a directory named *EndNote Library.Data*, then a subdirectory named *Library Trauth.Data*, in which the file named *Library Trauth.enl* is located. The directory *EndNote Library.Data* also contains a subdirectory, among others, named *PDF*, which in turn contains the folder *Roberts-1993-Timing of the Younge-2052588800*. This folder, with its long name, contains the PDF renamed as *Roberts-1993-Timing of the Younge.pdf.*

BibDesk, EndNote, and Papers are only three of the reference management tools that are available for different operating systems, but they are representative examples. All of these tools help in managing rapidly expanding reference data bases and PDF collections during the course of a typical earth sciences research project.

Recommended Reading

Bradley RS (1999) Paleoclimatology – Second Edition: Reconstructing Climates of the Quaternary. Academic Press, Burlington

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences 102:16569-16572

Langenberg H (2010) Editorial – Number crunch. Nature Geoscience 3:445.

Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D (1993) Timing of the Younger Dryas event in East Africa from lake-level changes. Nature 366:146–148

Ruddiman WF (2000) Earth's Climate – Past and Future. W.H. Freeman and Company, New York

Williamson D, Taieb M, Damnati B, Icole M, Thouveny M (1993) Equatorial extension of the Younger Dryas Event - Rock Magnetic Evidence From Lake Magadi (Kenya). Global and Planetary Change 7:235–242

http://www.springer.com/978-3-642-32543-4

MATLAB® and Design Recipes for Earth Sciences
How to Collect, Process and Present Geoscientific
Information
H. Trauth, M.; Sillmann, E.
2013, XII, 292 p. With online files/update., Hardcover
ISBN: 978-3-642-32543-4