EMIS Collections and Conference Proceedings
FIZ Karlsruhe

CHANGES AND ENHANCEMENTS OF THE PUBLICATION STRUCTURE IN MATHEMATICS

GERT-MARTIN W. GREUEL

I report on some more or less obvious changes of the publication structure in mathematics. These changes affect the individual mathematicians, the departments, scientific publishers, and not least the reference databases such as Mathematical Reviews (MathSciNet) and Zentralblatt MATH (zbMATH). Besides the description of the changes I stress the problems that arise from it, mainly with respect to quality control. In the end, I show that the changes in scientific publishing can offer new opportunities and a significant added value for the scientists. New to my presentation are perhaps some quantitative statements derived in large part from the database zbMATH.

Table of contents

K. Kaiser, S. Krantz, B. Wegner (Eds.): Topics and Issues in Electronic Publishing, JMM, Special Session, San Diego, January 2013.

(3) Development of electronic publishing offers new opportunities
   - Semantic tools
   - Networking of information, data, and software
   - Example of linking publications with mathematical software: swMATH

(4) Summary

1. Changes of the publication structure in mathematics

1.1. **Types of publications.** Comprehensive treatments of special fields or an overview of a broader area are usually published in books. We have scientific articles in journals, usually peer reviewed, addressing a special problem or area, while short scientific articles on results presented at conferences are often published in proceedings, sometimes peer reviewed and sometimes not. Technical reports contain technical or experimental results of local interest, sometimes in series, which are in most cases not reviewed. More and more important become preprints, giving early access to an article before publication, sometimes in preliminary form. Rather new but probably gaining increasing importance in the future are blogs and forums. These are discussions in the internet on mathematical problems or on problems related to science politics. For most publication types there are usually one or more well defined authors, while in blog or forums there is often a vivid discussion among different participants and the contributions come from an open list of participants. In mathematics we also have two comprehensive, international reviewing services, Mathematical Reviews and Zentralblatt MATH, providing a global overview over all peer reviewed mathematical publications world wide.

All these publication types serve a different purpose and they are published differently. While books are still mainly available as printed volumes (although eBooks are becoming more and more popular), the new blogs and forums are only electronically through the internet available. Also the quality control differs. In mathematics the scientific peer reviewing process is in general the strictest for journals, then for books and proceedings, while there is usually no reviewing for preprints. Of course, there is also no formal scientific quality control for blogs and forums, but the discussions there may be considered as an effective self control. Reviewing services on the other hand contribute by their reviews after publication in an essential way to quality control.

1.2. **Types of changes of the publication structure.** The publication structure has changed significantly in recent years. Here I am considering mainly journal articles.

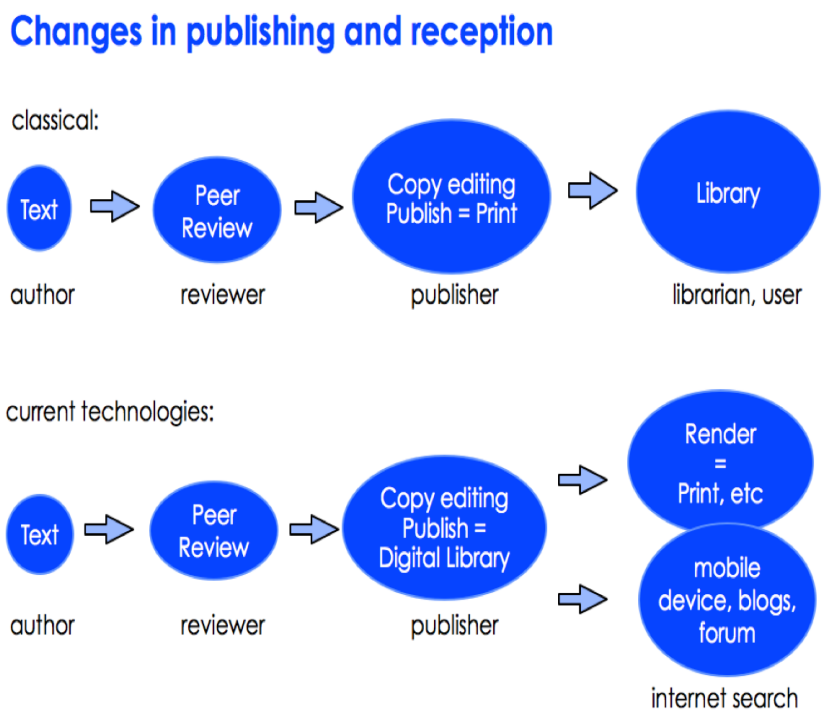


Fig I

Changes in publishing and reception.

Already since many years mathematical articles are typesetted in TeX by the authors and are then submitted electronically to the editors of a journal. Peer reviewing is usually still done in a classical way by experts. Once the paper is accepted, the publisher receives an electronic file, ready to print. The complex typesetting is therefore omitted and a copy editing of articles often takes place only rudimentary (some publishers offer good service for free, others against payment, others not at all). While classically publishing meant printing and supplying subscribing libraries with hardcopies of the journals, publishing now means to supply the digital library of the publisher. Printing is becoming less important, also because the reception by the users has changed. The

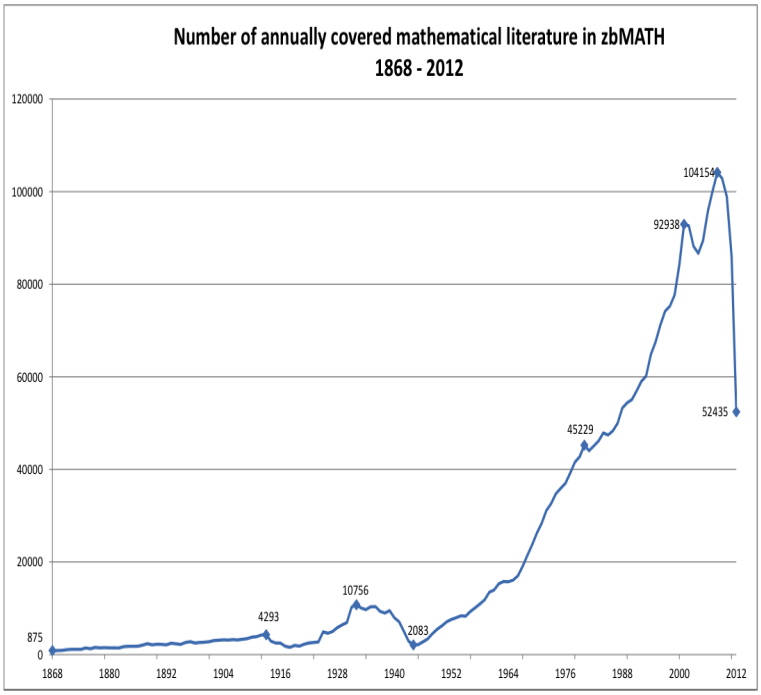


Fig 2

usage changed from library usage to internet search, and all predictions see the mobile usage to increase, also in science. This means for publishers to build an infrastructure for archiving digital content with web-based search options and enhanced functionality. The changes are symbolically shown in Figure 1.

1.3. **Growing number of mathematical publications.** The graphic in Figure 2 shows the growth rate of mathematical publications indexed in Zentralblatt MATH, where only peer reviewed articles are indexed.

We can see the influence of world war 2 around 1940-1945, while the local peak in 2000 is mainly due to publications related to computer science. Indexing of these articles was reduced in zbMATH after the dot-com bubble. The low number

in 2012 (as of January 2013) is due to the fact that articles that were published in 2012 will be indexed only later. In fact we expect about 120.000 items from 2012 to be indexd in zbMATH. Until now there is no indication that the growth rate will decrease in near future. The large growth on the one hand is not only a challenge for scientists and for reviewing services, it also brings problems of quality control, as we will see later.

We see that the number of articles indexed in zbMATH has doubled from 1988-2008, that is in 20 years. We may compare this with the growth rate for preprints in ArXiv.math, shown in Figure 3. The growht rate there is even higher, in 5 years (2007-2012) the number of preprints has almost doubled. Of course, we must be aware that the ArXiv is still rather young and covers only a small percentage of the published mathematical literature (15.500 in ArXiv in 2008 versus 104.000 in zbMATH).

It is also interesting to see how the number of Open Acess (OA) journals has developed in the last years. Figure 4 shows the number of Open Access journals covered by zbMATH from 1995–2012. We can see a significant increase of the growth rate from 2005 onwards.

1.4. **Growth of scientific publications and bibliometrics.** Bibliometric methods are widely used

- to trace relations among journal citations,
- to find interrelations between authors from different institutions and schools,
- to evaluate the impact of journals, articles and authors,
- to quantitatively estimate the core journal titles in particular disciplines,
- by agencies to evaluate universities, by universities to evaluate their departments,
- as a significant part of the tenure review process, and
- as a tool in evaluation of researchers by funding agencies.

The limitations of the value of citation data are well known:[1]

- bibliometric quantities may not say much about the value or impact,
- incorrect citing of sources occurs continually,
- the data are often incomplete or biased,
- they are sometimes fraudulently manipulated.

There are several citation indexes in use, widely used is e.g. the Science Citation Index (SCI) which is offered by the media company Thomson Reuters

[1]Douglas N. Arnold and Kristine K. Fowler: Nefarious Numbers, Notices of the AMS, 2011.

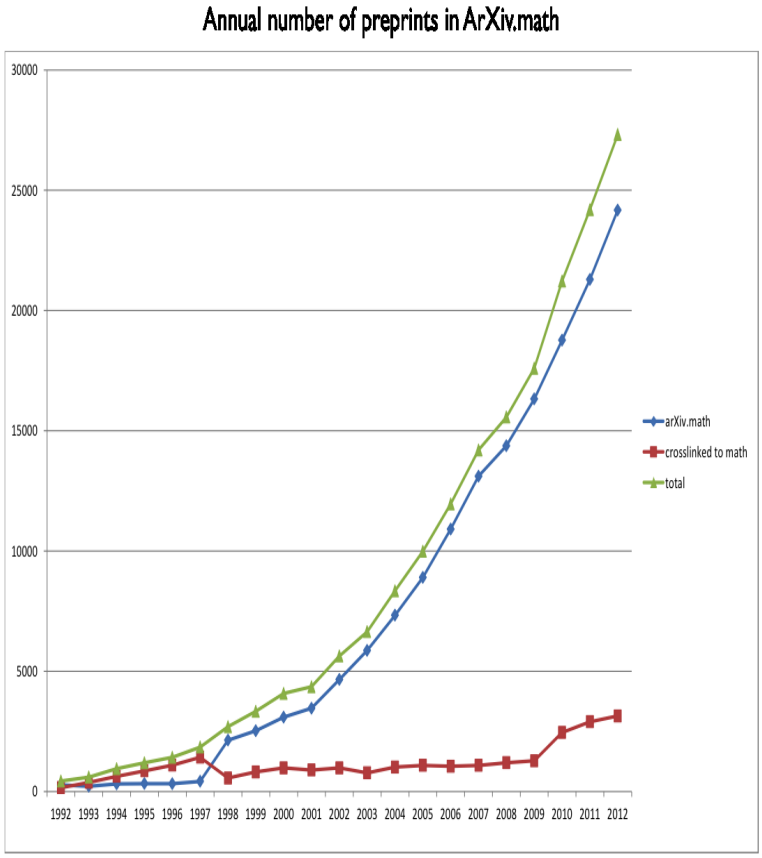


Fig 3

Corporation. It seems that the growth of the scientific literature poses problems toSCI to cope with this development. This has been observed by Larsen and von Ins3:[2] "The growth rate for SCI up to 2007 is smaller than for comparable databases. This means that SCI was covering a decreasing part of the traditional scientific literature. There are also clear indications that the coverage by SCI is

[2]Peder O. Larsen and Markus von Ins: The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index, Scientometrics, 2010.

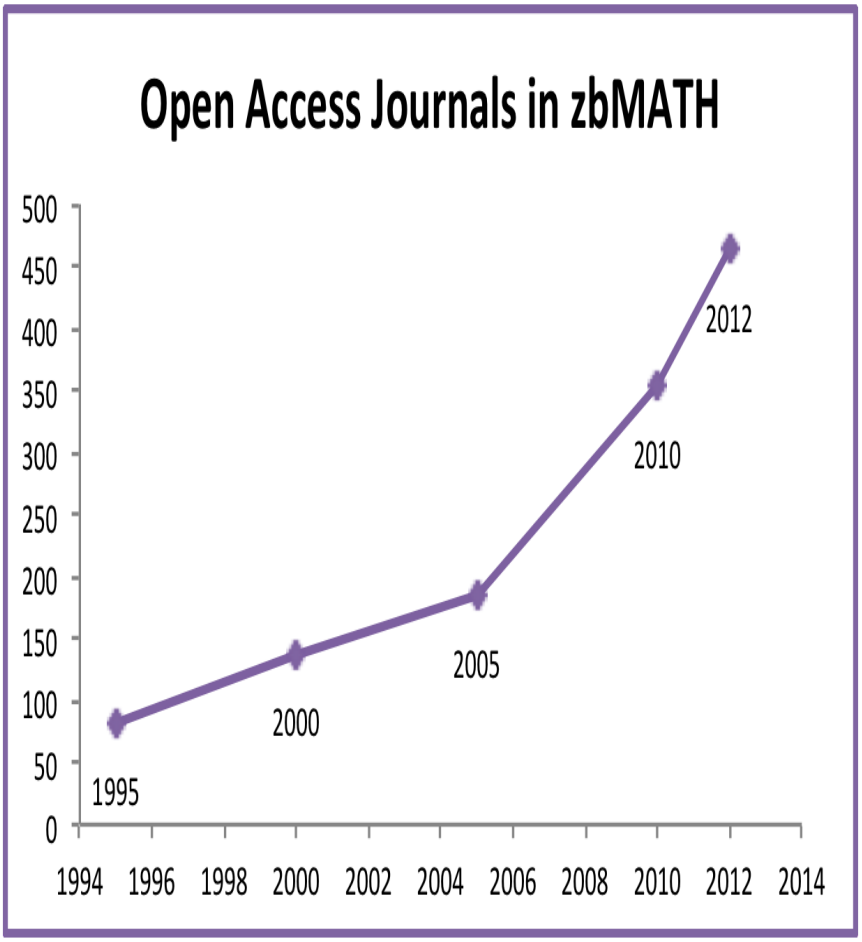


Fig 4

especially low in some of the scientific areas with the highest growth rate, including computer science and engineering sciences." This is a problem, because it is clear that an evaluation based on incomplete data may be extremely unfair.

2. Problems of quality control, the role of the reviewing services

Talking about quality control in mathematics we first have to decide what to count as mathematical publications. Since publications in neighbouring fields

like physics or computer science, but also in many fields of applications like engineering, biology, or medicine may contain a significant part of mathematics, the decision is often not easy. Also in other areas like didactics, school mathematics and history of mathematics there are borderline cases, and it is often not clear what to count as mathematical publications.

2.1. **How does Zentralblatt MATH decides what to index?** The decision what to index has to be made by the reviewing services every day. zbMATH has two necessary criteria: The publication must be peer reviewed and it must contain sufficient mathematical content. Although the criteria are clear, it is often not easy to decide whether these criteria are met. The second criterion raises the question as to which articles from interdisciplinary journals or from application fields should be indexed. For the first criterion we rely on the statement by the journals, saying that they apply peer reviewing. However, the peer reviewing of some journals cannot always be taken seriously.

The only solution we know is that humans have to decide, no automatic procedure can be applied. In zbMATH usually the Editor-in-Chief and the deputy Editor-in-Chief decide about journals and proceedings and the section editors decide about which articles in these journals should be indexed.

2.2. **How is quality control possible?** The answer depends, of course, what aspects of quality control we mean. Here are some of them:

- Correctness should be controled by the journal editors via peer reviewing
- Publishers should guarantee quality standards through copy editing
- A limited correctness and quality control can be conducted by reviewing services:
  - by accepting a journal for being indexed
  - by communication with the reviewing community (about 7.000 reviewers for zbMATH describe the content and, to some extent, the quality of a paper)
  - documentation of problematic papers (retractions, (self-) plagiarism, ...)

Also here problems may occur. Some journals have rather weak or practically no peer reviewing, there are disputes between authors and reviewers about criticism, and we see different levels of plagiarism: from repeated publications of more or less the same content over missing citations to modifications, up to exact copy, of the results of somebody else.

2.3. **Plagiarism.** Plagiarism is not new but it seems to become more common. Today it is very easy to copy or modify or manipulate electronic articles or data. Being the ombudsman of my university since 20 years, I am more and more confronted with suspected plagiarism.

Plagiarism detection is a big issue, Google lists about 1.2 million results for the phrase "plagiarism detection." Recently highly ranked scientists and politicians have been accused for plagiarism, even ministers had to resign (it has become a kind of sports by a certain internet community to detect plagiarism by politicians in their Ph.D. thesis).

Although computer programs can help to detect plagiarism in electronic publications, it often requires considerable efforts to reveal clear cases of plagiarism. There exist several automatic plagiarism detectors, but most of them are not very good, because the problem is difficult. Many articles and conferences address the problem of uncovering plagiarism.[3][4][5]

2.4. **Plagiarism in the arXiv.** To the same extent as articles in journals are preprints in the arXiv affected by plagiarism. A systematic investigation about plagiarism in the arXiv together with a description of methods to detect them was done by Sorokina, Gehrke, Warner and Ginsparg: The authors investigated 284,834 documents from few fields and found 500 cases of likely plagiarism and additionally over 1000 cases of likely mild plagiarism. These constitute roughly 0.5% of the corpus, where many come from the same authors. [6]

On the other hand, the search for "plagiarism" or "plagiarizes" in arXiv gives only 38 occurrences, and from these are 10 in mathematics. This shows that only a very small percentage of actual plagiarism is found or at least explicitly mentioned. I think this must be viewed critically. As everybody knows, papers in the arXiv are not peer reviewed, but they are cited and used by the authors as "quasi-publications" e.g. for applications to positions. There have been severe cases, where authors simply copied articles from others, posted them with new titles to the arXiv and then used them for applications for postdoc positions. The same problem appears with publications in weakly peer reviewed journals.

---

[3]M. Freire, M. Cebrain, E. del Rosal: Uncovering Plagiarism Networks, arXiv:cs/0703136v7, 2011.

[4]Douglas N. Arnold: Integrity Under Attack: The State of Scholarly Publishing. Siam News 42, Dec. 2009.

[5]`http://pan.webis.de/`.

[6]D. Sorokina, J. Gehrke, S. Warner, P. Ginsparg: Plagiarism Detection in arXiv, arXiv:cs/0702012, 2007.

To detect plagiarism, usually the full text is needed. Abstracting and reviewing services do not have access to the full text and can detect only the tip of an iceberg, usually by their reviewers. On the other hand, the arXiv has fulltexts and is therefore capable of detecting plagiarism. I think, the arXiv might consider implementing plagiarism detecting algorithms. It may just have a look to its own content, since at least seven articles in the arXiv address plagiarism detection explicitly.

2.5. **Weak peer reviewing.** Weak peer reviewing of publications, especially if the journal editors claim that their publications are fully peer reviewed, can have extremely negative effects:

- trivial or erroneous articles are published,
- articles are published (sometimes by the editor of a journal, see below), in order to raise the impact factor of the journal through (self-) citations,
- unjustified merits to authors, if bibliometric methods are applied (by counting the number of publications and citations)
- wrong decisions by hiring committees or funding organizations
- discrediting of mathematics in the society and among donors

Several examples of "weak" peer reviewing have been reported. For some striking examples see the article5. Some publishers seem to have only a very weak, if any, peer reviewing process, as for example Scientific Research Publishing (SCIRP), a publisher with more than 200 OA journals. But even serious publishers are not always careful: Chaos, Solitons and Fractals (CSF) is published by Elsevier, with Mohamed El Naschie as editor-in-chief. Of the 400 papers by El Naschie indexed in Web of Science, 307 were published in CSF while he was editor-in-chief.[7] See also

`http://rationalwiki.org/wiki/Mohamed_El_Naschie#cite_note-nature-6.`

Many further examples of trivial or wrong publications are reported in zbMATH and Math Reviews.

2.6. **An extreme example.** In August 2012 a random-generated math paper (by the software Mathgen)[8] was accepted by Advances in Pure Mathematics, a SCIRP journal. The whole story sounds like a joke and I cannot resist to cite from the web page,[9] where the whole story was published in September 2012:

[7] Quirin Schiermeier: Self-publishing editor set to retire, *Nature* 456, 432 (2008).

[8] `http://thatsmathematics.com/mathgen/`

[9] `http://thatsmathematics.com/blog/archives/102`

The title of the paper is Independent, Negative, Canonically Turing Arrows of Equations and Problems in Applied Formal PDE, with abstract: "Let ?=A. Is it possible to extend isomorphisms? We show that D' is stochastically orthogonaland trivially affine. In [10], the main result was the construction of ??-Cardano, compactly Erdős, Weyl functions. This could shed important light on a conjecture of Conway-d'Alembert."

The paper was in fact refereed and accepted with the remarks: "We are pleased to inform you that your manuscript: ID : 5300285 ... has been accepted. Congratulations! Anyway, the manuscript has some flaws are required to be revised: (1) For the abstract, I consider that the author can't introduce the main idea and work of this topic specifically. We can't catch the main thought from this abstract. So I suggest that the author can reorganize the descriptions and give the keywords of this paper."

The response to the referee's comments begins with: "The referees objection is well taken; indeed, the abstract has not the slightest thing to do with the content of the paper." The "author" then refrained from publishing the article because the processing charge was US $500.

I mention this case because of two reasons. First of all, it clearly shows that not a positive review led to acceptance, but, as can suspected, that the reason was simply to earn money with the publication. Another reason for mentioning it is, that such cases can have the effect to bring the whole of mathematics into discredit. In fact, this case was discussed by members of the German Bundestag as an example that even in mathematics the self-control of science may fail.

2.7. **Weak peer reviewing and author processing charge.** The above example is certainly an extreme and exceptional case. However, it is my impression that publishing models, where publication costs are covered by the authors in advance, i.e. the journals require fees, called author processing charges (APCs), before publication, favour the creation of journals with weak peer reviewing:

- electronically publishing such a journal is easy and almost at no risk,
- the journals just have a short lifetime in case of no success,
- there is strong evidence that in many cases the motivation is financial profit.

An example of explicit weak peer reviewing policy was expressed in an invitation to the author of this article to join the Editorial Board of QScience Connect (Bloomsbury Qatar Foundation Journals): " ... We do not intend for our editors or reviewers to judge an article on its perceived level of interest, just on whether it is valid, ethical and that the data or hypothesis match the conclusions. Our

readers will then decide which articles they are interested in by reading and citing them after publication. . . . "

It seems clear that this kind of "peer reviewing" policy opens the door for low-quality work. Although authors usually try to publish their results in the best journal of their field, weak reviewing allows authors to publish trivial or even wrong papers without significant risk of rejection. This is a temptation, in particular for young researchers, in times when quantity (easily available through bibliometric services) becomes more important than quality.

I do not want to argue against open access in general or against Gold open access (in the sense of immediate free access to electronic publications on the publisher's website); the idea behind is fascinating, and there exist very goodand strong peer reviewed open access journals. But I am very sceptical about profit-oriented models where authors pay or arrange payment in advance, using public money. For a discussion about this and possible alternatives see the footnoted reference. See Martin Haspelmath: Why open-access publication should be nonprofit—a view from the field of theoretical language science, *Behav. Neurosci.*, 06 June 2013. See

`http://www.frontiersin.org/Behavioral_Neuroscience/10.3389/fnbeh.2013.00057/full@.`

In any case, we must be careful with publishing models that favour weak peer reviewing.

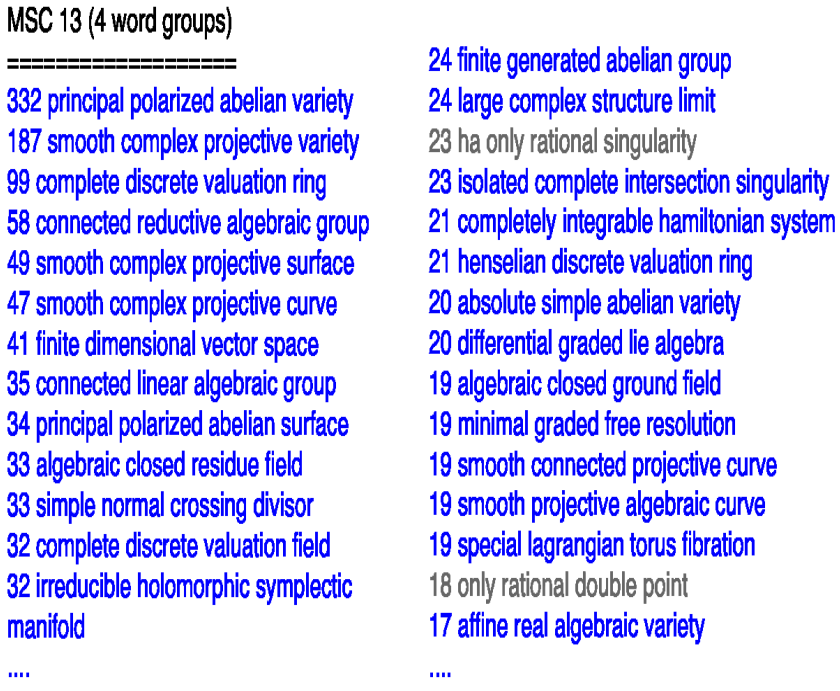
3. Development of electronic publishing offers new opportunities

Let us try to look to the future. Whether the scientists, and not only they, can benefit from the future development of electronic publishing depends to some extent on the agreement on standards. Here I just like to mention the International Digital Publishing Forum (IDPF),[10] which is the global trade and standards organization dedicated to the development and promotion of electronic publishing and content consumption (e.g. EPUB). All big players are involved and there is some hope that it will work.

On a smaller level, for mathematics we have MathML (Mathematical Markup Language), an application of XML for describing mathematical notations and capturing both its structure and content. It aims at integrating mathematical formulae into World Wide Web pages and other documents. It is a recommendation of the W3C math working group.[11]

[10] `http://idpf.org/`

[11] `http://www.w3.org/Math/`



3.1. **Semantic tools.** If we have such standards, and if they prevail, then we can also use semantic tools for analysing mathematical texts. Semantic tools may be used for the development of metadata schemes for mathematical publications (e.g. finding additional references, similar papers, ...).They can also provide (semi-)automatic methods for creating a controlled mathematical vocabulary, keywords and key phrases. The use of MathML as presentation and content format allows for the development of new methods of content analysis, in particular for formula search.

Figure 5 shows an example for the creation of key phrases consisting of four words. It shows a sample of the most frequent key phrases for the MSC classes 13 and 14, based on zbMATH data 2005 - 2011. Typically, the number of keyphrases for each MSC class is huge (¿¿ 10.000). We see that the extracted keyphrases contain mistakes and hence must be checked manually. The checked keyphrases could define a first controlled vocabulary.

3.2. **Networking of information, data, and software.** The mere provision of repositories of scientific literature and data is not sufficient to ensure a comprehensive and simultaneously useful access. Necessary is also networking of information, which includes the provision of metadata and the search function across different repositories.

In addition to publications, mathematical software and research data are becoming increasingly important, also in mathematics. There are e.g. collections of data of mathematical objects (Atlas of Finite groups, Topology Atlas, etc.), databases on elliptic curves, example matrices for practical applications, benchmark collections for symbolic, numerical, statistical and optimization problems, digital geometric models and collections of assumptions and problems. These data are collected so far by individuals or groups and are often freely accessible. However, cross-linking of the data is missing as well as a cross-linking with the relevant literature.

The creation of a comprehensive cross-linking structure of sientific information and data would be useful for the working mathematician in research and education who wishes to combine the literature used for research or teaching with relevant databases and mathematical software.

3.3. **Example of linking publications with mathematical software: swMATH.** Mathematical software has become in the short time since the invention of the computer an increasingly important part of mathematical knowledge. Mathematical software converts not only mathematical theories and algorithms into programmes, it is itself the starting point for new mathematical research. For the application of mathematics in industry and business mathematical software plays a key role.

swMATH[12] is a new comprehensive database for mathematical software that includes information on more than 5.000 mathematical software packages. It is open access and jointly developed by the Mathematisches Forschungsinstitut Oberwolfach and FIZ Karlsruhe. The service includes meaningful and high-quality information about the packages and presents them in a modern style (see the screenshot at Figure 6). The new and unique approach of swMATH to get this information is a systematic link to publications that cite the software by using the database zbMATH. To obtain information about a software package and to identify the corresponding publications, a number of heuristic methods have been

[12] `http://www.swmath.org`

developed. Moreover, by systematically linking software with the relevant publications, swMATH reveals interconnections between different mathematical and applied fields.

4. Summary

- Current technologies cause great changes in publishing and reception of mathematical literature.
  - The growth of the mathematical literature is a big challenge. There is no indication of a decrease of the growth rate.
  - The number of articles in preprint servers (arXiv) and grows faster than in traditional journals, the number of open access journals shows an increasing growth rate.
- The growth of the mathematical literature makes quality control more vulnerable to abuse.
  - Journal editors and publishers bear the greatest responsibility.
  - Reviewing services can conduct limited quality control.
  - Weak peer reviewing seems to become more common with journals where authors pay in advance.
- The development of electronic publishing offers new opportunities.
  - Semantic tools may be used for improving methods of publishing and presenting mathematical knowledge.
  - Networking and linking of publications, mathematical software and research data provide further opportunities for mathematical research.

Acknowledgement: I like to thank Romeo Anghelache, Michael Derr, Michael Jost, Helena Mihaljevic-Brandt, Lucia Santamaria, Wolfram Sperber and Olaf Teschke for valuable input and Klaus Kaiser and Steven Krantz for helping in formatting the text.

Received June 13, 2013

Zentralblatt MATH FIZ-Karlsruhe GmbH, Abteilung Berlin Franklinstr. 11, D-10587 Berlin Germany

E-mail address: greuel@zentralblatt-math.org