MKM

A New Interdisciplinary Field of Research*

William M. Farmer†
McMaster University

June 9, 2004

Abstract

Mathematical Knowledge Management (MKM) is a new interdisciplinary field of research in the intersection of mathematics, computer science, library science, and scientific publishing. Its objective is to develop new and better ways of managing mathematical knowledge using sophisticated software tools. Its grand challenge is to create a *universal digital mathematics library* (UDML), accessible via the World-Wide Web, that contains essentially all mathematical knowledge (intended for the public). The challenges facing MKM are daunting, but a UDML, even just partially constructed, would transform how mathematics is learned and practiced.

1 Introduction

Prior to the Information Age mathematical knowledge was managed for several centuries in a simple way. Motivated by problems in science and technology as well as in pure mathematics, mathematicians defined mathematical concepts and then explored them by stating and proving conjectures. Their results, usually in the form of theorems, were presented in mathematical journals and textbooks. Scientists and engineers then read the results and applied them to their problems.

The new technology of the Information Age—computers, the Internet, and the World-Wide Web—is transforming how mathematics is practiced

*This paper appeared in the *ACM SIGSAM Bulletin*, 38:47–52, 2004.

†Address: Department of Computing and Software, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. E-mail: `wmfarmer@mcmaster.ca`.

and, as a result, is also transforming what mathematical knowledge is and how it is produced, communicated, and applied. The change taking place today in the nature of mathematical knowledge is striking. In the past the body of mathematical knowledge grew gradually and mainly reflected the interest and work of mathematicians. However, today mathematical knowledge is produced at a prodigious rate by many nonmathematicians as well as mathematicians. It is becoming increasingly difficult to ascertain what results are known and how they are related to each other. Also, nonmathematicians working on applications are producing new kinds of mathematical knowledge in which the mathematical content is quite different from that of traditional mathematical knowledge.

The best example of this new kind of mathematical knowledge comes from software development. Software systems implement algorithms, which embody mathematical knowledge in procedural form. In addition, the knowledge that is used to design, implement, and analyze software systems is largely mathematical. The mathematical knowledge arising from software development is not something that excites many mathematicians, but it is absolutely vital to the development of safe and useful software.

Today many scientists and engineers as well as mathematicians are producing mathematical knowledge with the help of computers and the Internet. Much of this knowledge concerns specific applications and is expressed procedurally instead of declaratively. It is no surprise that very little of it makes its way to traditional journals and textbooks—which are indeed inadequate for communicating this kind of mathematical knowledge. Although much of it is stored in electronic form, it is usually not widely accessible and not easily searched when it is accessible. As a consequence, it is likely that mathematics practitioners are wastefully solving the same mathematical problems over and over again because they do have the means to effectively transmit and receive the mathematical knowledge they need for their applications.

This situation must be addressed. Science and technology is vital to our society, and mathematical knowledge is vital to science and technology. But the traditional way of managing mathematical knowledge is no longer adequate, and current computer and communication technology does not provide an immediate solution. We need new ways of managing mathematical knowledge based on new technology and new theory.

2 What is MKM?

Mathematical Knowledge Management (MKM) is a new interdisciplinary field of research in the intersection of mathematics, computer science, library science, and scientific publishing. The objective of MKM is to develop new and better ways of managing mathematical knowledge using sophisticated software tools. MKM is expected to serve mathematicians, scientists, and engineers who produce and use mathematical knowledge; educators and students who teach and learn mathematics; publishers who offer mathematical textbooks and disseminate new mathematical results; and librarians and mathematicians who catalog and organize mathematical knowledge.

The management of mathematical knowledge can be divided into four activities: articulation, organization, dissemination, and access. We will discuss each of these activities individually and will state several challenge questions for MKM that are relevant to these activities.

2.1 Articulation

Mathematical knowledge cannot be communicated unless it is *articulated*. An articulated piece of mathematical knowledge has a *language* in which it is expressed, a *context* within which it is understood, and a *representation* by which it is conveyed.

Language. All mathematical knowledge is expressed in some language. The language may be *informal* and based on natural language (such as the languages used in most mathematics textbooks). Or the language may be *formal* with a precise syntax (such as the language of a typical computer algebra system) and possibly also with a precise semantics (such as Zermelo-Fraenkel set theory). It is usually easier and less costly to express mathematical ideas in an informal language, but mathematical knowledge expressed in a formal language can be read, analyzed, and presented with the help of sophisticated software tools.

Context. A piece of mathematical knowledge is understood within a context of background definitions and assumptions. An understanding of the context is needed to understand the mathematical knowledge itself. In traditional mathematics, the context is largely implicit; it is not precisely described but is indicated by conventions known to the mathematically literate. (This tradition is reflected in the MathML [17] and OpenMath [2] programs to put mathematical knowledge on the Web; expressing the context axiomatically, or in some other precise way, is not a goal of these programs.) In some cases, the context is assumed to be any set of definitions and assumptions in which the mathematical knowledge makes sense. In order to effectively understand and process mathematical knowledge, software tools need direct access to the context in which the knowledge resides. Software tools that do not have adequate access to the context cannot reliably process mathematical knowledge in sophisticated ways.

Representation. How a body of mathematical knowledge is conveyed is determined by its representation. It can be represented *declaratively* as an explicit set of statements, the set of logical consequences of a mathematical theory, or the set of theorems of a proof system. It can be represented *procedurally* as the knowledge that is embodied in a computation system consisting of a collection of data structures and algorithms. It can be represented *visually* by diagrams and animations. A body of knowledge can also be represented by a combination of declarative, procedural, and visual means.

Challenge questions for MKM:

1. What kind of software support is needed to convert an informal articulation of mathematical knowledge into a formal articulation?

2. When do the benefits of formalizing mathematical knowledge outweigh the costs?

3. How should the context of mathematical knowledge be expressed?

4. How can the declarative representations of mathematical knowledge offered by computer theorem proving systems be integrated with the procedural representations offered by computer algebra systems?

2.2 Organization

The world of mathematical knowledge is unimaginably immense. It can even be argued that it is inherently infinite and thus possibly even bigger than the physical world. In addition, mathematical knowledge is extraordinarily interconnected; the same piece of knowledge may appear in many different places and in many different forms. Articulated mathematical knowledge needs to be carefully organized to avoid redundancy and to capture connections. This requires identifying and abstracting common structure and then formalizing it as an axiomatic or algorithmic theory.

Challenge questions for MKM:

1. How should axiomatic and algorithmically theories be linked to avoid redundancy and to capture connections?

2. How should the role of mathematicians differ from the role of librarians in the task of organizing mathematical knowledge?

2.3 Dissemination

After mathematical knowledge is articulated and organized, it needs to be disseminated. It can be distributed as text in traditional journals and textbooks, it can be digitally stored and provided on the Web, and it can be incorporated into mathematical software systems such as computer theorem proving systems and computer algebra systems.

Challenge questions for MKM:

1. What role should universities, governments, professional societies, and publishers play in disseminating mathematical knowledge?

2. Who should own and administer mathematical knowledge?

3. How should disseminated mathematical knowledge be certified?

2.4 Access

People need software tools for finding the mathematical knowledge they require in a body of knowledge that has been disseminated. Tools are needed for doing searches and making queries, for performing deductions and computations with mathematical software systems, and for understanding how the knowledge has been articulated and organized. These software tools need to be much more sophisticated and easier to use than current tools. For example, search engines must understand the semantics of mathematical languages and, for example, when syntactically distinct expressions such as $(A \cup B^c)$ and $(A^c \cap B)^c$ are semantically equivalent.

Challenge questions for MKM:

1. What new software tools are needed?

2. What mechanism should be used to standardize and integrate software tools?

3 The Grand Challenge

The grand challenge of MKM is to develop a *universal digital mathematics library* (UDML). Composed of many heterogeneous, intercommunicating systems, it would be easily accessible via the World-Wide Web. It would be constructed in an open, cooperative fashion in the same way that the Internet was constructed. Never finished, it would continuously grow and in time would contain essentially all mathematical knowledge (intended for the public). It would also be continuously reorganized and consolidated as new connections and discoveries were made.

A UDML would contain a highly structured and interconnected mixture of axiomatic, algorithmic, diagrammatic, and other kinds of mathematical knowledge. Each piece of mathematical knowledge in it would carry a certification of its correctness (relative to a specified set of assumptions). It would also include an integrated collection of tools for exploring its contents. It is important to note that a UDML would be a library and not an archive. That is, its primarily purpose would be to make mathematical knowledge widely accessible, not just to store and catalog mathematical knowledge.

Creating a UDML will be a herculean project requiring the development of many new kinds of technology. Some of this technology is being developed now on current formal mathematics library projects including the NIST Digital Library of Mathematical Functions (DLMF) [4], the Formal Digital Library (FDL) [5], Hypatheon [3], Logosphere [7], Mizar [10], and the Wolfram Functions Site [18].

Challenge questions for MKM:

1. What is the best way to start designing and implementing a UDML?
2. Who should administer a UDML?
3. How should bodies of mathematical knowledge based on different foundations be integrated within a UDML?

4 Computer Algebra and MKM

Contemporary computer algebra systems embody an awesome amount of procedurally represented mathematical knowledge. They have had a huge impact on the way people practice mathematics. However, they are not good managers of the knowledge they contain. They are basically black boxes in which most of the knowledge within them is inaccessible to the ordinary user. For instance, the user does not have direct access to how computations are performed and to the context of definitions and assumptions that is being employed. As a result, it can be difficult to properly interpret the meaning of a computation's input/output relationship.

The future successors of computer algebra systems will be central components in a UDML. They will combine symbolic computation with visualization and formal deduction. Both the context in which a computation is performed and the algorithm by which it is performed will be accessible to the user. Strongly integrated with the other components of a UDML, they will produce trustworthy, well-understood results that be transported to other systems. And like the Axiom system [6], they will cover abstract mathematics as well as the mathematics of the complex numbers and its subsystems.

5 The MKM Consortium

As a new field of research, MKM was launched by the *First International Workshop on MKM* (MKM 2001) [11] in September 2001 at Hagenberg, Austria. Organized by Bruno Buchberger and Olga Caprotti, MKM 2001 lead to the founding of the *MKM Consortium* in December 2001 under the leadership of Michiel Hazewinkel and to a special issue [1] of the *Annals of Mathematics and Artificial Intelligence* dedicated to MKM.

The MKM Consortium is an international group of researchers dedicated to the promotion of research and interest in MKM. It has organized two subsequent MKM conferences: The *Second International Conference on MKM* (MKM 2003) [12] was held in February 2003 at Bertinoro, Italy, and the *Third International Conference on MKM* (MKM 2004) [13] will be held September 19-21, 2004 in Bialowieza, Poland.

The MKM Consortium currently consists of a European Chapter and a North American Chapter. The European Chapter obtained funding from the European Union in 2002–03 for a large, short-term exploratory project named the *Mathematical Knowledge Management Network* [8]. It also organized the *Mathematical Knowledge Management Symposium* [9] in November 2003 at Heriot-Watt University in Edinburgh, Scotland.

The North American Chapter [14] has organized two MKM workshops: *A North American Workshop on Mathematical Knowledge Management* (NA-MKM 2002) [15] held in June 2002 at McMaster University in Hamilton, Ontario and the *Second North American Workshop on Mathematical Knowledge Management* (NA-MKM 2004) [16] held in January 2004 at the *Joint Mathematics Meetings* in Phoenix, Arizona.

For more information about the issues and challenges of MKM, see the Web sites mentioned above.

6 Conclusion

The challenges facing MKM are daunting. In particular, it is not clear whether it is possible to construct a UDML with the attributes we have described. However, a UDML with sophisticated tools for exploring its contents, even if it is just partially constructed, would transform how mathematics is learned and practiced.

To be a success, MKM needs expertise and input from mathematicians, scientists, engineers, educators, librarians, publishers, computer scientists, and software developers. The perspective and understanding of mathematicians is especially needed. If mathematicians ignore MKM, they may find MKM producing a line of misguided technology that cannot be easily halted.

References

[1] B. Buchberger, G. Gonnet, and M. Hazewinkel, editors. *Mathematical Knowledge Management*, 2003. Special issue of *Annals of Mathematics and Artificial Intelligence*, 38:1–232.

[2] S. Dalmas, M. Gaëtano, and S. M. Watt. An OpenMath 1.0 implementation. In *International Symposium on Symbolic and Algebraic Computation (ISSAC-97)*, pages 241–248. ACM Press, 1997.

[3] B. L. Di Vito. Hypatheon: A mathematical database for PVS users. Available on the Web at `http://imps.mcmaster.ca/na-mkm-2004/proceedings/pdfs/divito.pdf.`, 2004.

[4] Digital Library of Mathematical Functions (DLMF). Web site at `http://dlmf.nist.gov/`.

[5] FDL project. Web site at `http://www.nuprl.org/FDLproject/`.

[6] R. D. Jenks and R. S. Sutor. *Axiom : The Scientific Computation System*. Springer-Verlag, 1992.

[7] Logosphere: A formal digital library. Web site at `http://www.logosphere.org/`.

[8] Mathematical Knowledge Management Network. Web site at `http://monet.nag.co.uk/mkm/index.html`.

[9] Mathematical Knowledge Management Symposium. Web site at `http://www.macs.hw.ac.uk/~fairouz/mkm-symposium03/`.

[10] The Mizar system. Web site at http://mizar.org/.

[11] MKM 2001. Web site at http://www.risc.uni-linz.ac.at/institute/conferences/MKM2001/.

[12] MKM 2003. Web site at http://www.cs.unibo.it/MKM03/.

[13] MKM 2004. Web site at http://mizar.org/MKM2004/.

[14] NA-MKM. Web site at http://imps.mcmaster.ca/na-mkm/.

[15] NA-MKM 2002. Web site at http://imps.mcmaster.ca/na-mkm-2002/.

[16] NA-MKM 2004. Web site at http://imps.mcmaster.ca/na-mkm-2004/.

[17] W3C Math Home. Web site at http://www.w3.org/Math/.

[18] Wolfram Functions Site. Web site at http://functions.wolfram.com/.