The Launch of the *Data Science Journal* in 2002

FRANCIS J. SMITH

ABSTRACT

The Committee on Data for Science and Technology of the International Council of Scientific Unions (CODATA) decided in February, 2001 to publish a new on-line peer reviewed journal, freely available to all. It would be called the Data Science Journal. The subsequent steps taken to launch this journal are described, leading to the publication of the first issue of 10 papers in April 2002 and of the first volume of 19 papers by December 2002. Some necessary corrections and improvements before the publication of a second volume of 19 papers in 2003 are also described.

CORRESPONDING AUTHOR:

Francis J. Smith

Queen's University Belfast, IE

fjsmith181@gmail.com

BACKGROUND

CODATA was formed in 1966, and amongst other data activities, it provided a forum for data scientists to exchange knowledge on their discipline. To this end, it hosted international workshops, meetings and seminars on special data topics and published contributions from the participants, often in book form. It also held biannual international conferences, beginning in 1968, which particularly provided publication opportunities for young scientists and those new to the field. Like most other conferences, papers were lightly refereed to ensure a reasonable standard for an international conference. The standard of these papers was generally high—some were very high—but they were not viewed as having the standard required in a quality peer-reviewed journal. Data scientists lacked such a quality journal in which they could publish and get deserved recognition for the quality of their papers. So did their readers. I expect that these factors were in the minds of the far-sighted members of the CODATA Executive when they decided to start a quality Journal in Data Science in 2001.

EXECUTIVE DECISIONS 2001

The welcome decision to publish a CODATA journal on scientific data was taken by an Executive CODATA meeting in February 2001. This came as no surprise, as it had already been discussed at the two previous general assemblies. In the period just before 2001, much had been done in preparation, so many of the decisions concerning the nature and management of the journal could be taken at this meeting of the Executive, giving the journal its shape and management structure. The role of the Executive as the future policy maker would be aided by an executive working group chaired by Dr Steve Roussouw, of which Prof Gaude-Escard is a member, and the journal would be overseen by an Editorial Board, which we will discuss later.

Most importantly, it was agreed that the journal would be a quality journal, with all papers refereed to maintain a high standard. It was also agreed, after some investigation of the alternatives, that the journal would be an electronic journal, reducing costs and making it available to as wide a readership as possible. It was therefore agreed to fund the journal fully, enabling the journal to be made freely available to scientists and engineers in all countries. Whole papers as well as abstracts would be available for free.

The prospectus and scope of the journal were also agreed at this Executive meeting. These were based on the aims and scope of CODATA itself and were not controversial. They are appended in Appendices A and B.

It was further agreed that the journal should begin publication as soon as possible for an interim period, until the end of December 2002, when it would be reviewed. It was subsequently extended after 2002.

EDITOR-IN-CHIEF

I was honoured by being appointed as the first editor-in-chief. To aid me during the development stage, the journal would initially be hosted on a web server at my university, Queen's University, Belfast. A contract was also drawn up for the appointment of a secretary as an assistant editor for two days per week at Queen's University. Following this decision, Ms Maria Husin was soon appointed assistant editor, and a small editorial office was established at the university.

WEBSITE

Ms Husin's first task was to build a management system and website for the journal. For the website, she studied other journal sites on the web and built the Data Science site to be similar to the others. She worked with test papers in the first months before actual papers arrived, with little help from me apart from advice and encouragement. The website aimed to be simple but effective by the standards of that time. It was ready and needed few modifications when the first refereed papers arrived.

EDITORIAL BOARD

As previously mentioned, the CODATA Executive had agreed that the journal would be overseen by an Editorial Board. A list of members for this board was also selected by the Executive from a wide range of countries. Letters signed by the president of CODATA were then sent out to this list in April 2001, inviting participation on the board. A list of the resulting first Editorial Board is attached in Appendix C.

TITLE AND DOMAIN NAMES

After much discussion of several titles for the journal, a vote was taken by the members of the CODATA publications committee, and a clear majority voted for the title *Data Science Journal*. It had been pointed out that this title might be misunderstood. However, the majority felt that it was up to CODATA to ensure that it became understood. In time, the journal would itself define what is meant by 'Data Science', and indeed in time, it did just this.

Once the title was agreed upon, CODATA purchased rights to the following two internet domain names: http://www.datasciencejournal.org and http://www.datasciencejournal.net. The first of these was linked to the website. Unfortunately, it was too late to purchase the equivalent internet title ending in .com or the three shorter titles http://www.datascience.org, .com and .net, as they had already been registered.

Acquiring the domain name made it possible to begin the publication of the journal on the internet, where it joined a few other internet-only journals at that time.

PUBLIC ANNOUNCEMENT

Once the title *Data Science Journal* and the domain name had been agreed upon, the journal was first announced to the scientific public in the *CODATA Newsletter 82* in September 2001.

INSTRUCTIONS FOR AUTHORS/REFEREES

One of the first tasks before publication was to write manuscript preparation instructions for authors and guidelines for referees. For manuscript format, it was proposed that the journal adopt the format used in the published proceedings of the CODATA conferences, as it was already known by many CODATA members. Drafts of these instructions and format were sent to the Editorial Board and eventually approved by the majority of the board, although some board members thought the CODATA conference format was too complicated. In retrospect, it was indeed complicated and should have been simplified; but it would have held us up from the aim of the Executive to begin the publication as soon as possible. Also, we did not judge a paper by its format, which is easily fixed, but only by its content. So, following approval by the Editorial Board, the manuscript preparation instructions and the referee guidelines were posted on the new website.

REFEREEING

Each paper required the appointment of two referees, which was not always easy to do. To appoint the referees, the first task of an editor is to read a submitted paper and understand it. Then two referees who worked in the field of the paper had to be chosen. We had four sources of referees or advice: (1) members of the Editorial Board, (2) scientists referenced in the submitted paper, (3) scientists in a file we created of the authors and titles of papers published in the proceedings of the three previous CODATA conferences, and (4) scientists we knew. If a chosen scientist could not referee a paper, they were asked to suggest someone else, who in turn might suggest another person. Usually, however, one of the first scientists approached agreed to act as referee. Most referees sent in their referees' reports on the papers quickly, but some had to be chased up more than once. Eventually, when two referee reports had been received, the author(s) were told the outcome and, if positive, asked to resubmit if changes were needed. When the two referees disagreed, a third was appointed. We tried not to ask any referee more than once per year.

Our experience of refereeing demonstrated two things: (1) the length of time for refereeing was much longer than expected, and (2) the positive worth of the referee reports to the quality of the papers. Delays with referees, and by authors to make changes, were common. Some referees had to be reminded once or twice, some did not reply, and some eventually replied that they could not referee the paper, in which case another referee had to then be found and sometimes yet another. However, most referees took their task seriously, and the attention given by some referees and their detailed advice added considerably to the quality of the papers.

EDITING

Once a paper had been accepted as being of the required standard for the journal, the assistant editor then checked that the paper was in the correct format and made minor changes when necessary. Some papers—not only those by foreign authors—also needed the English to be corrected. If major changes were needed, the paper was returned to the author(s) with an explanation of the changes required.

FIRST MEETING OF THE EDITORIAL BOARD IN MONTREAL

The first meeting of the Editorial Board took place at the Montreal Conference in 2003. The minutes of this important meeting are given in Appendix D. The board had already been able to view the electronic publications of the *Data Science Journal* in the first year. The board rightly praised Ms Maria Husin, the assistant editor. But several changes were needed, although none was the fault of Ms Husin.

First, a critical omission was pointed out by the board that there was nothing on the website about copyright. It was recommended that copyright should be assigned to the author(s), with a licence given to CODATA to publish. Following the board's advice, the journal's copyright policy was defined with the support of the CODATA staff in Paris, and legal advice and a web page outlining the terms and conditions associated with the journal's licence to publish went online. Thereafter, authors were asked to agree to the terms and conditions of the licence on the website before submitting a paper.

Apart from copyright, several other comments or changes were recommended concerning the operation of the journal. These were as follows:

1. Confirmation was given that all papers should be refereed by at least two referees.

2. Referees should be given six weeks to respond before being reminded.

3. Special sections should be encouraged.

4. 'Issues' (used only in the first volume) should be discontinued. Papers should be published as soon as they are refereed.

5. Inclusion of complex data types (moving images, programs, etc.) should be encouraged.

6. A mirror website was needed for security.

7. Statistics on access should be collected.

These changes were implemented in the following months. To improve security, CODATA in Paris was sent regular copies of the journal website in case of a problem in Belfast until a mirror site was created. It turned out that statistics on access were already being collected.

INVITATION TO SUBMIT PAPERS FOR THE FIRST VOLUME

At first, almost no one outside the Executive knew about the journal. So, to obtain papers for the first volume, it was necessary to inform potential authors and at the same time invite them to submit quality papers to the journal. To do this, the CODATA president, Dr John Rumble, and I together drew up an invitation list of potential authors.

First, we selected approximately 90 quality papers given at the CODATA conference in 2000. Letters were then sent out inviting the first author on each paper to submit their paper to the new journal. Later, papers at the two previous CODATA conferences were examined, and a

smaller number of papers which had not been superseded by later papers were selected; again, the first authors were invited to submit papers to the journal. Later, a range of other papers and authors obtained from a variety of sources, including the Editorial Board, were also identified, and invitations were sent out. In total, about 200 invitations were sent to potential authors. We were disappointed that only 48 positive replies were received. But, in retrospect, we should not have expected more.

Of the 48 positive responses received to the invitation, 22 papers were submitted to the journal quite quickly. Referees were chosen, and 2 positive referee reports were eventually received for 16 of these papers; 4 others were delayed, and 2 were rejected by the referees.

THE FIRST PUBLISHED PAPERS

The first 10 of these refereed papers, in the order received by the assistant editor, made up the first issue of 10 papers of the *Data Science Journal*, which was published on the internet at the end of April 2002. A list of the authors and titles of these first 10 papers can be found in Appendix E. A second issue of another 9 papers was published by the end of 2002, making a first volume of 19 papers. A second volume of 19 papers was published in 2003. All these papers were doubly refereed by their peers.

So, with much effort by many people, particularly by the far-sighted members of the CODATA Executive, the Editorial Board, the CODATA staff in Paris and the first authors and referees, the peer-reviewed *Data Science Journal* was launched in April 2002. We believe that it was the first published journal on data science.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Appendix A.** The Prospectus in 2001. DOI: https://doi.org/10.5334/dsj-2023-11.s1
- **Appendix B.** Scope of *Data Science Journal* in 2001. DOI: https://doi.org/10.5334/dsj-2023-11.s2
- **Appendix C.** Editorial Board. DOI: https://doi.org/10.5334/dsj-2023-11.s3
- **Appendix D.** Meeting of the Editorial Board, Montreal, 30 September 2002. DOI: https://doi.org/10.5334/dsj-2023-11.s4
- **Appendix E.** The First Issue. DOI: https://doi.org/10.5334/dsj-2023-11.s5

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATION

Francis J. Smith
Queen's University Belfast, IE

]u[