The Formation of CrossRef: A Short History

YEARS OF COLLABORATION
1999–2009

CELEBRATING 10 YEARS OF COLLABORATION 1999–2009
The Formation of CrossRef: A Short History

1999

CrossRef is born.

2000
The Beginnings

With the emergence of the World Wide Web, it was inevitable that publishers would use its capacity to disseminate content, and that in time it would be possible to navigate from one publication to another through reference citation linking, greatly benefitting researchers and scholars working in the online environment and librarians facilitating access to online content. It was by no means inevitable, however, that broad-based linking would be achieved through the collaboration of an organization of publishers. The emergence of CrossRef as the linking service of Publishers International Linking Association, Inc. (PILA) is the story of two plotlines that converged dramatically at the 1999 Frankfurt Book Fair — one an initiative to develop a prototype reference linking service using the Digital Object Identifier (DOI), and the other an effort to assemble a coalition of publishers with the critical mass to launch, grow, and sustain such a system. The successful launch of CrossRef brought benefit for researchers, scholars, librarians, and publishers alike.

The DOI emerged from the work of the Enabling Technologies Committee of the Association of American Publishers (AAP), which in late 1994 was investigating how to develop both the infrastructure and the market for publishing on the Web in a way that would protect intellectual property. In connection with this, the Committee in March 1996 issued a Request for Proposal for a system that would, among other things, provide unique and persistent identifiers for online content items. They selected the Corporation for National Research Initiatives (CNRI), with its Handle System technology, to develop a prototype.

Craig Van Dyck of Springer-Verlag (later of Wiley), who by then was Chair of the Committee, described a 1996 meeting at the Manhattan office of Sun Microsystems. Among those also present were Bill Rosenblatt of Sun, Carol Risher of the AAP, Larry Lannom of CNRI, and David Sidman and Andy Stevens of Wiley. “The DOI was invented as an identifier for digital content, but at that point there wasn’t a specific application for it,” he said. “We were exploring possibilities, trying to ‘connect the dots,’ and the idea came up that there could be a service that captured metadata for some kind of syndication. As it turned out, the concept of a metadata database incorporating the DOI became the core of the CrossRef system.”

Carbon Fibre

Publishers began to collaborate in exploring the possibilities of the new online environment early on, as in the Red Sage Project, an experimental networked library of digital health sciences journals that ran from 1993 through 1996. Launched by AT&T Bell Laboratories, Springer-Verlag, and the University of California, San Francisco, Red Sage involved the participation of 18 publishers in addition to Springer.

Another such effort was the U.K.-based SuperJournal project, through which a consortium of 20 publishers worked with a group of universities and libraries to develop electronic journals and identify the key requirements for successful electronic publishing. Led by David Pullinger of Nature magazine, the SuperJournal project ran from December 1995 through 1998 and included as one of its initiatives a proposed linking scheme codenamed Carbon Fibre. Pullinger has explained that the name was intended as a metaphor for publisher collaboration, evoking a union of individual elements combining great strength with lightness, the lightness also suggesting that the load on each individual player would be slight. He noted that if publishers were to band together in this way, they would free themselves from potential dependence on the growing strength of secondary publishers as providers of access to online content — a theme that was to emerge distinctly in the creation of CrossRef.

Pullinger presented the Carbon Fibre concept at a meeting on October 1, 1996, at the Frankfurt Book Fair. Among those present were Pieter Bolman of Academic Press; Eric Swanson of Wiley; Dick Rudick of Wiley; Herman Spruijt of Elsevier; Norman Paskin of Elsevier; Sally Morris of Wiley; Jan Velterop of Academic Press; and Bob Campbell of Blackwell Publishing.

Eric Swanson recalled the immediate reaction to the presentation. “A number of us said, ‘This is very important — this must be done,’” he said. “As Pieter and I walked out of the room, we said to each other that if Nature didn’t achieve it, we had to find some other way to do it.”
Further Developments

ALTHOUGH THERE WAS A FOLLOW-UP Carbon Fibre meeting on December 6, 1996, Bolman and Swanson were left with the impression that there was no substantive activity toward realization of the concept. They had already discussed this possibility in a November 5th conversation in which Bolman voiced the strong opinion that the two of them should take the project into their own hands if need be.

At the time, however, publishers were primarily focused on making their journals available online and working out the business models that would make that feasible. There was considerable discussion of linking among journals, but no consensus on how that might be achieved. But by early 1998, several publishers had launched a significant number of their journals on the Web and interest in linking had intensified.

As Pullinger had recognized, one obvious possibility was that intermediaries such as Thomson ISI (Institute of Scientific Information), with its Web of Science service, as well as Wolters Kluwer’s Ovid, CAS (Chemical Abstracts Service, with its ChemPort technology), and others, might provide a solution. Bolman addressed this option in a presentation to the International Association of Scientific, Technical, & Medical Publishers (STM) in Heidelberg, Germany, in April 1998, but, like Pullinger, advocated direct linking among publishers as the better alternative. He proposed a DOI-enabled Link Processor, a centrally located “black box” that would process reference information to effect linking. This would allow what came to be known as “distributed aggregation”—broad access via linking to content that primary publishers would continue to hold, as opposed to the centralized aggregation of content by intermediaries that would then provide linking, or the provision of linking by intermediaries functioning as gatekeepers to publisher-held content. The Link Processor would provide maximum publisher autonomy, and equally important, scalability, in sharp contrast to the bilateral linking agreements that a number of publishers and other organizations were beginning to craft. “It is clear,” said Bolman at the time, “that ultimately all publishers need to participate.”

The concept that all content need not be hosted on the same platform to enable linking was a relatively new one at the time. When Bolman spoke of a Link Processor, however, he was not speaking of a theoretical construct, but rather of a so-named project conceived by Ken Metzner of Academic Press in late 1996 and initiated in June 1997 by Tony Hammond and Ed Pentz, who was to become Executive Director of CrossRef.

Around the same time, Craig Van Dyck, David Sidman, and Andy Stevens of Wiley developed a metadata database, originally intended for Wiley’s journals.

Monzu

WITH THE DISSOLUTION OF THE CARBON FIBRE PROJECT as a cautionary example, Bolman and Swanson decided to proceed in confidentiality to create a viable linking system, working with a core group of committed partners, before sharing their results with the world at large. “We had to develop something,” said Swanson, “and we had to develop it with a small group of publishers that had a high level of credibility.”

In August 1998, Wiley and Academic Press agreed formally to create a prototype Link Processor and develop a business plan for a linking service that would allow broad publisher participation. The Link Processor was developed by Hammond and Pentz of Academic Press in collaboration with Van Dyck and Stevens of Wiley; it harnessed the Wiley metadata database, in which DOIs and bibliographic information could be deposited, to a reference parser created by Hammond that could match references to DOIs using the metadata database. By May 1999, the prototype had been completed, demonstrating reference linking between Academic Press and Wiley content.

The process of assembling a core group of partners was more complicated. Bolman and Swanson initially sought to enlist the American Chemical Society (ACS) and the American Institute of Physics (AIP), two prestigious non-profit organizations, but
for a variety of reasons the AIP decided not to participate. Bob Bovenschulte of the ACS was favorably impressed, however, and the ACS became involved in the development of the Link Processor prototype. Bovenschulte also suggested that Bob Massie of CAS, which was a division of the ACS, be included in further discussions.

The four met for dinner in April 1999 at Monzu, an Italian restaurant then located at the corner of Mercer Street and Prince Street in New York’s Soho neighborhood. As it turned out, ACS only participated in the project through June, but the evening had given it a codename—thenceforth, Monzu.

At the same time, work proceeded on the creation of a business plan. At a May 21 meeting, Bolman, Swanson, and Bovenschulte called for an analysis of the specifications, size, and cost of a scaled-up linking service for participating publishers based on the Link Processor prototype. Van Dyck, Pentz, Stevens, and Lorrin Garson of the ACS completed the analysis in June, delivering estimates for staff and technical infrastructure requirements as well as startup and ongoing operational costs. Based on these results, Van Dyck and Pentz then produced a business plan for the “Central Facility,” as the service was then referred to.

With the growing recognition of the DOI’s importance, the International DOI Foundation (IDF) had been founded in March 1998 to develop and promote the DOI System as a common infrastructure for online content management, with Norman Paskin as Founding Director and Charles Ellis, formerly Wiley’s CEO, as Chairman. Swanson and Bolman both cited the appointment of Ellis, at the suggestion of Lex Lefebvre (then Secretary General of STM), as an important factor in creation of CrossRef. “Charles had served as Chairman of the AAP and Vice President of the IPA [International Publishers Association], and he had also chaired the joint STM/IPA Information Identifier Committee,” said Swanson. “He had tremendous credibility within the industry, and with his European experience with Pergamon Press and Elsevier, he brought a valuable international perspective to the table.” The IDF was well aware of, and in fact involved in, the Monzu initiative; Paskin participated in at least two meetings in August 1999, and Ellis in at least one.

In the months following the Monzu dinner, Bolman and Swanson began discussions with Stefan von Holtzbrinck of Nature Publishing Group, and then Mike Spinella, who at the time was with the American Association for the Advancement of Science (AAAS), the publisher of Science. Both agreed to participate. With the approach of the 1999 Frankfurt Book Fair, and with the completion of both the Link Processor prototype and the business plan, the four were nearing readiness to make an announcement. As Bolman had jotted in a note to himself, “go for quality, and the quantity will follow.”

**DOI-X**

**IN EARLY 1999, THE DOI SUBCOMMITTEE** of the AAP’s Enabling Technologies Committee, under the leadership of Howard Ratner, then of Springer-Verlag and later of Nature Publishing Group, decided to launch its own initiative to create a reference-linking prototype, named DOI-X. Unlike Monzu, the DOI-X project was announced publicly, under the joint sponsorship of the AAP, CNRI, and the IDF, and with the participation of Academic Press, AIP, Elsevier, ISI, Wiley, Kluwer Academic, the subscription agent RoweCom (formerly Dawson/Faxon), and Springer-Verlag.

Work began in July, with the goal of delivering results in time for the Frankfurt Book Fair in October. The project focused on defining metadata standards, deposit procedures, general rules, and aspects of the requisite technical architecture. A paper detailing the project appeared in the February 2000 issue of D-Lib Magazine, coauthored by Ratner, Helen Atkins of ISI, Catherine Lyons of Springer-Verlag, Carol Risher of the AAP, Chris Shillum of Elsevier, and David Sidman and Andy Stevens of Wiley.

Ratner delivered the DOI-X findings on Tuesday, October 12, at the STM Annual Frankfurt Conference, held as always the day before the opening of the Book Fair. A demonstration of the prototype was given during the fair’s general session.
The Wiley and Academic Press participants in DOI-X contributed technology developed in their joint venture, and the Monzu group anticipated that the two projects might ultimately merge to enrich each other. But while the other members of the DOI-X team were aware that Wiley and Academic Press had been working together, they had no knowledge of the full scope of Monzu, with its working prototype, business plan, and organizational vision.

**Frankfurt 1999**

**THE PRESENTATION OF THE DOI-X RESULTS** at the 1999 Frankfurt Book Fair may have been anticipated, but there were other events that were not.

At the Board meeting of the STM association, held the afternoon of Monday, October 11, before the fair’s Wednesday opening, discussion focused on an emerging U.S. National Library of Medicine (NLM) initiative called E-Biomed (later PubMed Central) that had been proposed by Harold Varmus of the National Institutes of Health in the spring of 1999. Varmus envisioned a digital archive of journals, accessible free of charge and with the added value of reference linking. “Our consensus was that publishers should be the ones doing the linking,” said Bob Campbell, who chaired the meeting. “Since we were ‘higher up the stream,’ so to speak, we should be able to link our articles ahead of the NLM as part of the process of producing them. Stefan von Holtzbrinck then set the ball rolling by offering to link Nature publications with anyone else’s. We decided to issue an announcement of a broad STM reference linking initiative. It was, of course, a strategic move only, since we had neither plan nor prototype.”

A small group led by Arnoud de Kemp of Springer-Verlag met in an adjacent room immediately following the Board meeting to draft the announcement, which was distributed to all attendees of the STM annual meeting the following day and published in an STM membership publication.

Campbell recalled running into Bolman and Swanson (neither of whom was then on the STM Board) in the hotel lobby immediately after the drafting of the announcement. Their astonishment at hearing what had just transpired was matched by Campbell’s own on learning what they had been working on. Like virtually every other STM Board member (with the obvious exception of von Holtzbrinck), Campbell knew nothing of Monzu.

Bolman and Swanson chose to seize the moment, and called an ad hoc meeting the following evening, Tuesday, October 12, to announce their venture and assemble a coalition of publishers to launch it. Also present at the meeting were Campbell, Derk Haank of Elsevier, Stefan von Holtzbrinck, Tim Ingoldsby and Marc Brodsky of AIP, Arnoud de Kemp, Ken Metzner, Ed Pentz, Howard Ratner, and Craig Van Dyck. Although there was some apprehension in the air that one faction or another might seek to sway the outcome, a collegial spirit of collaboration prevailed, with the central theme described by Bolman as, “What next, and how?” In the end, the participants in the meeting agreed to join the venture, with a number prepared to provide startup loans.

The potential benefit of the service that would become CrossRef was immediately apparent. Organizations such as AIP and IOP (Institute of Physics) had begun to link to each other’s publications, and the impossibility of replicating such one-off arrangements across the industry was obvious. As Tim Ingoldsby later put it, “All those linking agreements were going to kill us.”
Coming Together

THE “WHAT NEXT, AND HOW” were worked out over the next three months through a series of meetings, the first held on November 8 at Wiley’s New York offices. The meeting opened with a general statement that the initiative was designed to provide “a service demanded by our users, which will benefit science,” and included reports from the Technology, Legal, and Publicity Working Groups. Governance and funding were addressed, with Wiley, Academic Press, Elsevier, Blackwell, Springer-Verlag, and Nature Publishing Group committing to provide startup loans; Kluwer Academic Publishers and The Institute of Electrical and Electronics Engineers, Inc. (IEEE) joined the funding group soon after.

On November 16, a press release announcing the imminent launch of the as-yet unnamed service was issued by the 12 participating publishers, in cooperation with the IDF: AAAS, Academic Press, AIP, Association for Computing Machinery (ACM), Blackwell, Elsevier, IEEE, Kluwer Academic, Nature Publishing Group, Oxford University Press, Springer-Verlag, and Wiley. The release included quotes from representatives of most of the publishers.

Swanson described the November 8 meeting as key. “With everyone’s commitment to a public statement, and with the commitment to funding, it became very unlikely that we would not be able to proceed as planned.”

The Governance Task Force met at Wiley’s offices on November 29, establishing that the 12 original members would comprise the Board of Directors, that there would be a five-member Executive Committee, and addressed various other issues, including the creation of a code of practices for member publishers. A second press release was issued on December 9, announcing the CrossRef name and the inclusion of four additional participants—The University of Chicago Press, Institute of Physics Publishing, World Scientific Publishing, and Taylor & Francis.

Representatives of the 16 participating publishers, with the addition of the journal hosting platform Catchword and, by invitation, Norman Paskin of the IDF and Marco Bronckers of the law firm Stibbe Simont Monahan Duhot, met at the Macmillan (Nature) offices in London on December 13. This meeting signaled the formal establishment of the Board of Directors, with Swanson as President and Bolman as Treasurer, and with Roy Kaufman, a Wiley counsel, to serve as Secretary pending the appointment of an Executive Director. In addition to Swanson and Bolman, the Board consisted of Mike Spinella of AAAS, Marc Brodsky of AIP, John R. White of ACM, John Strange of Blackwell, John Regazzi of Elsevier, Anthony Durniak of IEEE, Jeffrey K. Smith of Kluwer Academic, Stefan von Holtzbrinck of Nature Publishing Group, Martin Richardson of Oxford University Press, and Ruediger Gebauer of Springer-Verlag.
Bob Campbell cited the composition of the Board as critical to CrossRef’s success. “Right from the start, we had very significant representation from key societies and not-for-profit organizations,” he said. “That was tremendously helpful politically, in gaining acceptance within the broader scholarly community.”

The final meeting prior to incorporation took place on January 13, 2000, again in the Wiley offices. In addition to addressing a range of legal, technical, financial, and administrative matters—including the distinction between PILA and CrossRef, and the design of the CrossRef logo—the meeting marked the appointment of Ed Pentz as Executive Director, effective February 1.

The Early Days

CROSSREF WAS INCORPORATED ON JANUARY 27, 2000, and began operating out of an office rented from Harcourt Brace & Company in Burlington, Massachusetts, in February, with Ed Pentz as the first employee, followed by Lisa Hart, now Senior Financial Manager, in April. The system went live in June 2000, a blend of elements from the DOI-X Project and the Monzu Link Processor.

First-year expenses had been estimated at $660,000, and the full cost of the first three years at $2.4 million. Actual first year expenses were $838,000, within a reasonable margin of the initial estimate. The actual three-year cost figure of $4.2 million was considerably higher than projected, but the system grew much more quickly than anyone had anticipated. By the end of 2003, CrossRef had 300 members with 12 million DOIs assigned, compared to the initial projection of 60 participating publishers and 3 million DOIs assigned.

The financial support of the lenders through two rounds of loans in 2000, backed by their willingness to restructure the repayments when necessary, was crucial to the early success of CrossRef. The total amount loaned was $1.2 million, comprising two payments of $600,000. The presence of commercial and non-profit publishers alike among the lenders reflected broad support for the venture, and their agreement to a second round of loans signaled their conviction that CrossRef’s strategy was sound and its operation well managed under Pentz’s leadership. Two additional rounds of loans totaling $750,000 were advanced early in 2001 and early in 2002 to fund the first system rewrite, with AAAS, AIP, and Oxford University Press joining the original eight lenders. The repayment of the principal and accrued interest on the loans began in 2004 and was completed in March 2007, well ahead of schedule.

Campbell has noted Pentz’s role as key to CrossRef’s growth and acceptance. “Ed brought to his position an unusual range of skills, combining technical expertise with exceptional sensitivity to people, to the different cultures and philosophies of CrossRef’s constituencies,” he said. “He and the rest of the staff did a remarkable job, creating a brand and a successful culture in that first year or two. It was no small challenge to start with an unproven concept and work with organizations not accustomed to collaborating in this way. The team Ed put together—with people like Lisa Hart, Amy Brand, Chuck Koscher, and now Geoff Bilder, representing a sort of ‘second generation’ CrossRef—has built on that foundation very effectively.”

Originally, CrossRef focused on linking primary content to primary content, enabling readers to access cited content with a single click. Soon after CrossRef was established, secondary publishers and other intermediaries were welcomed to participate on standard terms as affiliates. CrossRef did become, in fact, the “broad church” that Bolman and the other founders had envisioned.

Rapid Expansion

CROSSREF GREW MORE RAPIDLY on all fronts than anyone had expected, soon reaching a critical mass of content and participants that would ensure the organization’s long-term sustainability. Membership grew across the board, including publishers, affiliates, and libraries. Content registered in the CrossRef system expanded both in quantity and type to include journal articles, working papers, technical reports, book series, titles, and chapters, reference work entries, conference proceedings, components (such as figures, graphs, tables), dissertations, standards, databases, and database entries. With the massive backfile digitization projects undertaken by publishers, the CrossRef linking network now extends back to 1665 with the first articles published in Philosophical Transactions, placing the literature of four centuries at the fingertips of researchers and scholars.

CrossRef’s membership spans the globe and represents every type of organization, business model, and scholarly discipline, and the establishment of CrossRef’s Oxford office in 2004 has given the organization an international presence.
Strategic Direction, Forward Vision

IN 2003, CROSSREF DEVELOPED A FORMAL MISSION STATEMENT that provided the organization with a well-defined strategic focus. The statement was updated in 2007 to read:

CrossRef is a not-for-profit membership association whose mission is to enable easy identification and use of trustworthy electronic content by promoting the cooperative development and application of a sustainable infrastructure.

The effective governance of the Board of Directors, backed by the work of the CrossRef Committees and Working Groups and a dedicated professional staff, ensures that the organization’s initiatives and ongoing operation are aligned with the mission.

It is worth noting that the mission statement makes no mention of reference linking or DOIs, the organization's original focus. It instead addresses CrossRef's broader capacity to develop services collaboratively that would not be possible for publishers individually. Recent offerings include Cited-by Linking, which allows publishers to provide links to articles that cite a particular article, and CrossCheck, powered by iThenticate, which provides editors with a tool to help detect plagiarism. Currently in development, the CrossMark service will enable publishers to identify the version of record that they maintain with a logo and embedded metadata. CrossMark addresses the theme of trustworthiness that is central to the mission statement, and highlights the value publishers add both in the publishing process and as stewards of content after publication, through the issuance of corrections, retractions, and enhancements.

Taking Stock

THOSE CLOSE TO THE FORMATION OF CROSSREF had high expectations of its success. As Karen Hunter of Elsevier, a Founding Member of the IDF Board, put it, “There really wasn’t any question about it working, because all the big players were in at the start.” Similarly, there has been high estimation of CrossRef’s achievement, not only in achieving its primary goal of reference citation linking but in preventing a “digital divide” between large publishers with the resources to create platforms and smaller ones whose works could have been marginalized.

Perhaps the single most telling indicator of CrossRef’s success is its universal acceptance. “You don’t hear a bad word about CrossRef,” said Bob Campbell. “It’s well regarded amongst the library community, and well regarded amongst scientists and policy makers. That is really quite extraordinary.”

As CrossRef approaches its 10th anniversary, it offers services that would have been difficult to imagine at the outset. Future possibilities run the gamut from text mining to uniquely identifying authors and other contributors to the process of scholarly communication. It may not be possible to predict what CrossRef will achieve in the next 10 years—but whatever it is, it will be decided by consensus, and realized through collaboration.
CROSSREF AND THE ENVIRONMENT.

CrossRef is committed to the conservation of precious natural resources and the continued health of our planet. As a global citizen, we continually strive to reduce the environmental impact of the work we do.