

Crystallographic publishing in the electronic age

P. R. Strickland and B. McMahon*

International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.
Correspondence e-mail: bm@iucr.org

The journals of the International Union of Crystallography have grown in size and number over the past 60 years to match developments in scientific practice and technique. High quality of publication has always been at the forefront of editorial policy and ways in which this has been achieved are described. In particular, the development of standard exchange and archive formats for crystallographic data has allowed the editorial office to conduct automated analyses of structural data supporting articles submitted for publication and these analyses assist the scientific editors in careful and critical peer review. The new information technologies of the Internet age have allowed the IUCr journals to flourish and to provide a wide range of powerful services to authors, editors and readers alike. The integration of literature and supporting structural data is of particular importance. The new technologies have also brought fresh economic and cultural challenges, and offer completely new opportunities to disseminate the results of scientific research. The journals continue to respond to these challenges and take advantage of new opportunities in innovative ways.

© 2008 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

The publication of crystallographic information, particularly the results of primary scientific research, has always been central to the mission of the International Union of Crystallography (IUCr). Over the years, its journals have evolved, from issues assembled by practising scientists, through professionally edited publications, to today's online hyper-linked pages.

In recent times, there have been retrospective articles on the development of the journals by Editors and distinguished contributors (Abrahams & Matula, 1988; Cruickshank, 1998). Now, to mark the 60th anniversary of *Acta Crystallographica* and the 40th of *Journal of Applied Crystallography*, we have been invited to contribute our perspective as members of the professional team that edits and publishes the journals in times of great technical change.

We reflect on the history of these publications but we also indulge in some predictions of the future. The last 15 years have demonstrated how scholarly communication can change in unimaginable ways with startling speed; we aim, therefore, to do no more than scope out some ways in which the journals can become even more vigorous in serving the interests of this most vigorous of sciences.

2. History

2.1. The early years

Following the demise of *Zeitschrift für Kristallographie* during the Second World War, there was no natural home for crystallographers to publish their results. Kamminga (1989)

and Cruickshank (1998) describe how the formation of an international scientific union for crystallography came about as a direct consequence of the need for a new journal, open to scientists of all nations.

Although several possible new publishing structures were considered (*e.g.* separate journals published in the UK, USA and USSR), there was an overwhelming desire for a single journal owned and managed by the crystallographic community. In July 1946, a Provisional International Crystallographic Committee approved the launch of a journal, to be called *Acta Crystallographica*, and recommended the establishment of an International Union of Crystallography charged with: ownership of the journal; publication of other crystallographic works; and standardization of methods and units. This Union would be the focus for international collaboration and would represent the interests of crystallography alongside the other sciences.

The IUCr was admitted to the International Council of Scientific Unions in 1947, and the first issue of *Acta Crystallographica* appeared in April 1948 under the editorship of Paul Ewald. The journal was an immediate success, rapidly exceeding the capacity of the publishing house charged with its initial production, and – despite initial funding from UNESCO and benefactors in the UK and USA – precipitating a financial crisis within a few years of its launch. Following a transfer to the publishing house of Ejnar Munksgaard in 1952 (a relationship maintained over 55 years through Munksgaard's integration within Blackwell Publishing) and a significant subscription price rise, the journal returned to financial stability and continued to grow in number of pages and in number of subscribers.

Table 1

Milestones in IUCr journal publishing.

A list of past and serving members of the Commission on Journals is maintained at the Web address <http://www.iucr.org/iucr-top/comm/cj/cjlist.html>.

1947	IUCr admitted to the International Council of Scientific Unions
1948	<i>Acta Crystallographica</i> launched; Editor P. Ewald
1960	A. J. C. Wilson succeeds as Editor of <i>Acta</i>
1968	Journal splits into <i>Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography</i> and <i>Section B: Structural Crystallography and Crystal Chemistry</i>
1968	<i>Journal of Applied Crystallography (JAC)</i> launched; Editor A. Guinier
1970	R. A. Young appointed Editor of <i>JAC</i>
1978	S. C. Abrahams appointed Editor of <i>Acta</i> ; M. Hart Editor of <i>JAC</i>
1983	<i>Acta</i> reconstituted as <i>Section A: Foundations of Crystallography</i> ; <i>Section B: Structural Science</i> ; <i>Section C: Crystal Structure Communications</i>
1984	M. Schlenker appointed Editor of <i>JAC</i>
1987	C. E. Bugg appointed Editor of <i>Acta</i>
1990	A. M. Glazer appointed Editor of <i>JAC</i>
1993	Separate Section Editors appointed: A. Authier (<i>Section A</i>); F. H. Allen (<i>Section B</i>); S. R. Hall (<i>Section C</i>); C. E. Bugg becomes Editor-in-Chief
1993	Launch of <i>Acta Crystallographica Section D: Biological Crystallography</i> ; Editor J. P. Glusker
1994	<i>Journal of Synchrotron Radiation (JSR)</i> launched; Editors S. S. Hasnain, J. R. Helliwell and H. Kamitsubo
1996	J. R. Helliwell appointed Editor-in-Chief
1999	Launch of Crystallography Journals Online
1999	G. Ferguson appointed Editor of <i>Acta C</i> ; G. Kostorz Editor of <i>JAC</i>
2000	D. M. Mills appointed Editor of <i>JSR</i>
2001	Launch of <i>Acta Crystallographica Section E: Structure Reports Online</i> ; Editors W. Clegg and D. G. Watson
2002	D. Schwarzenbach appointed Editor of <i>Acta A</i> ; C. P. Brock Editor of <i>Acta B</i> ; Å. Kvik and T. Ohta Editors of <i>JSR</i>
2003	E. N. Baker and Z. Dauter appointed Editors of <i>Acta D</i>
2005	Launch of <i>Acta Crystallographica Section F: Structural Biology and Crystallization Communications</i> ; Editors H. M. Einspahr and J. M. Guss
2006	G. Kostorz appointed Editor-in-Chief
2007	A. Pyzalla appointed Editor of <i>JAC</i>

In 1962, a professional technical editor, Stephen Bryant, took over the handling of manuscripts, which were now too numerous for a voluntary scientific editor to cope with. Bryant's home was in Chester, in north-west England. So began the growth of the professional editorial office on which the IUCr now depends. The staff has grown to a current complement of 17, a modest number to handle the eight journals, *International Tables for Crystallography*, Web and other publications.

2.2. Financial considerations

In the earliest days, the major concern was to generate enough cash flow to allow the journal to continue, while selling it at a subscription price that could be met in post-war depressed economies. Once financial stability had been established, the question arose of the extent to which income from journal subscriptions might be used to fund other activities of the IUCr.

It has been a general principle, certainly in recent decades, that the journals should cover their costs and operate at a modest surplus. The surplus is used to build up the IUCr's

funds generally, to develop its publishing programmes, and to support other educational and scientific activities.

A typical target has been to obtain revenues of 10% over the costs of producing and distributing the journals. These targets are not always met (they can be susceptible to currency fluctuations and are sometimes compromised by large investments in novel projects). Nevertheless, journal publishing provides a major component of the IUCr's overall income and succeeds in the objective of supporting many non-publishing activities. In particular, this financial strategy has allowed subscription costs for Adhering Bodies to be maintained at roughly constant levels over extended periods of time. This in itself has provided greater stability to the IUCr than to some other Scientific Unions or inter-Union bodies.

In spite of this, there is a general sense that IUCr journals constitute good value for money. Abrahams & Matula (1988) provided evidence for this in comparison with many other learned society and commercial publishers and discussed their criteria for this judgement. More recently, the network-theory-derived criteria adopted by the Eigenfactor.org research project rank IUCr journals highly in terms of cost-effectiveness (<http://www.eigenfactor.org/pricesearch.php>).

2.3. Growth and diversification

By the end of the 1960s, the number of reports of crystal structure determinations was so high that a split into two sections was necessary. One section concentrated on structural papers; the other retained a more physical character. The partitioning into *Sections A* and *B* of *Acta* coincided with the launch of a companion *Journal of Applied Crystallography* (Table 1).

A further partition occurred in 1983, when crystal structure reports were split off into a separate section, and *Sections A* and *B* were reformed to address respectively the foundations of crystallography and structural science (the latter referring to advances in understanding chemical or physical systems arising from structural properties of the solid state).

The new *Acta Crystallographica Section C* was subtitled *Crystal Structure Communications* and replaced the structure reports journal of the same name initiated by Mario Nardelli at the University of Parma in 1972. Nardelli could no longer handle the growing number of structure reports as experimental techniques continued to improve and became easier to use. Even within the growing editorial framework of the IUCr journals, the continuing explosion of structure reports became – and remains – a challenge to traditional procedures for dealing with the scholarly literature.

However, the same advances in computing which were driving scientific progress were producing new technologies in publishing; and increased efficiencies through automation, described further in the sections below, have allowed *Section C* to be sustained as an economically viable publication.

By 1993, the need had arisen to create another section specializing in the rapidly growing field of biological crystallography. With the appointment of Jenny Glusker as Editor of the new *Section D*, the Commission on Journals was reformed

Table 2Articles in IUCr journals that have been discussed in *Citation Classic Commentaries* between 1977 and 1992.

Hanson, Lea & Skillman	HFS atomic scattering factors	<i>Acta Cryst.</i> (1963), 16 , 1040–1044
Karle & Karle	The symbolic addition procedure for phase determination for centrosymmetric and noncentrosymmetric crystals	<i>Acta Cryst.</i> (1966), 21 , 849–859
Karle & Karle	An application of a new phase determination procedure to the structure of <i>cyclo</i> (hexaglycyl) hemihydrate	<i>Acta Cryst.</i> (1963), 16 , 969–975
Ibers & Hamilton	Dispersion corrections and crystal structure refinements	<i>Acta Cryst.</i> (1964), 17 , 781–782
Schomaker & Trueblood	On the rigid-body motion of molecules in crystals	<i>Acta Cryst.</i> (1968), B24 , 63–76
Doyle & Turner	Relativistic Hartree–Fock X-ray and electron scattering factors	<i>Acta Cryst.</i> (1968), A24 , 390–397
Shannon & Prewitt	Effective ionic radii in oxides and fluorides	<i>Acta Cryst.</i> (1969), B25 , 925–946
McWeeny	X-ray scattering by aggregates of bonded atoms. I. Analytical approximations in single-atom scattering	<i>Acta Cryst.</i> (1951), 4 , 513–519
Berghuis, Haanappel, Potters, Loopstra, MacGillavry & Veenendaal	New calculations of atomic scattering factors	<i>Acta Cryst.</i> (1955), 8 , 478–483
Blow & Crick	The treatment of errors in the isomorphous replacement method	<i>Acta Cryst.</i> (1959), 12 , 794–802
North, Phillips & Mathews	A semi-empirical method of absorption correction	<i>Acta Cryst.</i> (1968), A24 , 351–359

with separate specialist Editors handling the individual titles, and an Editor-in-Chief chairing the Commission and overseeing strategic development.

In 1994, the IUCr launched another new title, *Journal of Synchrotron Radiation*. This both acknowledged the leading role played by crystallographic applications of synchrotron radiation and signalled a willingness by the IUCr to accommodate related sciences in an interdisciplinary publication. This journal has gained respect across the different scientific fields it represents and currently enjoys a high citation ranking.

2.4. The electronic era

The impact of the new technologies was also changing journal publishing in two profound ways, which we discuss at greater length in subsequent sections. The development of a standard exchange format for crystallographic data allowed semi-automated checking of reported structures; and the advent of Internet communications protocols, particularly the http of the World Wide Web, allowed the online journal to become a reality. The IUCr's own online journals platform was established in time for the 18th IUCr Congress in Glasgow in 1999.

As submissions of structure reports continued unabated, the IUCr launched in 2001 an online-only title carrying brief accounts of novel structure determinations, *Acta Crystallographica Section E: Structure Reports Online*. This journal took advantage of automated workflow to create a very rapid turnaround in peer review and production. Once more, the volume of submissions exceeded expectations and *Section E* rapidly became the largest publication in terms of numbers of articles and of pages (see the annual reports of the Executive Committee at <http://www.iucr.org/iucr-top/cexec/index.html>). Despite efficient production methods, this journal entered the market at a time when library budgets were under intense pressure and, given its very rapid growth, commercial considerations suggested that it would not survive if sold at a realistic subscription price. At the same time, the demands of the open-access movement (see §10) suggested that the time

was right to test a new financial model and *Section E* enters 2008 as a fully open access publication, supported directly by author fees.

Another online-only journal was launched in 2005; *Section F* concentrated on articles in structural biology and crystallization. Again, the community has welcomed rapid publication in the online medium and the journal has become successful in its short lifetime as a natural home for crystallization papers and structural communications. It publishes a growing number of protein structures, many resulting from structural genomics programmes, and is actively working on streamlining protein structure deposition and reporting in ways parallel to the small-unit-cell structures described in *Sections C* and *E*.

3. Significant publications

Bibliometric and citation analyses over several years show the continuing importance of the IUCr journals as core journals in the field (Hawkins, 1980; Garfield, 1990*a,b*; Behrens & Luksch, 2006). They remain central to the crystallographic community but are also well cited in influential journals across many areas of science, demonstrating the interdisciplinarity of the subject. Earlier studies emphasize the strength of the links between crystallography and chemistry and physics journals, but, as late as 1990, Garfield (1990*a*) reported an almost total absence of links to life sciences journals in a 1988 citation analysis. The strong growth of *Sections D* and *F* of *Acta Crystallographica* would suggest this is no longer likely to be the case.

Many papers in the journals of the IUCr have been very influential, or have played a central role in the development of theories, procedures, methodologies or software implementations. In a broad survey such as this, it is neither desirable nor appropriate for us to single out individual contributions. Cruickshank (1998) mentions a few favourites of his own. We present instead in Table 2 a list of some older articles featured in *Citation Classic Commentaries* (part of the *Current Contents* literature awareness series). Such commentaries indicate

Table 3

Journal impact factors calculated over 2, 5 and 23 years for core crystallography journals.

Source: Thomson ISI: http://www.in-cites.com/research/2004/september_27_2004-2.html

Rank	2003 Impact Factor	Impact 1999–2003	Impact 1981–2003
1	<i>Acta Cryst. B</i> (3.64)	<i>Acta Cryst. B</i> (5.32)	<i>Acta Cryst. A</i> (28.19)
2	<i>J. Mol. Graphics Model.</i> (2.93)	<i>J. Appl. Cryst.</i> (5.05)	<i>J. Appl. Cryst.</i> (20.08)
3	<i>Cryst. Growth Des.</i> (2.74)	<i>Acta Cryst. D</i> (4.56)	<i>J. Mol. Graphics Model.</i> (18.11)
4	<i>CrystEngComm</i> (2.73)	<i>J. Mol. Graphics Model.</i> (4.10)	<i>Acta Cryst. B</i> (14.72)
5	<i>J. Appl. Cryst.</i> (2.32)	<i>Acta Cryst. A</i> (3.88)	<i>Acta Cryst. D</i> (10.12)
6	<i>Acta Cryst. D</i> (2.21)	<i>J. Inclusion Phenom. Mol.</i> (3.83)	<i>Liquid Crystals</i> (9.57)
7	<i>Polyhedron</i> (1.56)	<i>Liquid Crystals</i> (3.18)	<i>J. Crystal Growth</i> (8.61)
8	<i>Acta Cryst. A</i> (1.21)	<i>Polyhedron</i> (2.83)	<i>Polyhedron</i> (8.53)
9	<i>J. Crystal Growth</i> (1.41)	<i>J. Crystal Growth</i> (2.63)	<i>J. Inclusion Phenom. Mol.</i> (7.79)
10	<i>Z. Kristallogr.</i> (1.27)	<i>Z. Kristallogr.</i> (2.37)	<i>Mol. Cryst. Liq. Cryst.</i> (6.93)

Table 4

Some recent articles listed in the Thomson ISI 'Hot Papers' database.

Brünger, Adams, Clore, DeLano, Gros, Grosse-Kunstleve, Jiang, Kuszewski, Nilges, Pannu, Read, Rice, Simonson & Warren	<i>Crystallography & NMR system: a new software suite for macromolecular structure determination</i>	<i>Acta Cryst.</i> (1998), D54 , 905–921
Terwilliger & Berendzen	Automated MAD and MIR structure solution	<i>Acta Cryst.</i> (1999), D55 , 849–861
Altomare, Burla, Camalli, Cascarano, Giacovazzo, Guagliardi, Moliterni, Polidori & Spagna	<i>SIR97: a new tool for crystal structure determination and refinement</i>	<i>J. Appl. Cryst.</i> (1999), 32 , 115–119
Allen	The Cambridge Structural Database: a quarter of a million crystal structures and rising	<i>Acta Cryst.</i> (2002), B58 , 380–388
Spek	Single-crystal structure validation with the program <i>PLATON</i>	<i>J. Appl. Cryst.</i> (2003), 36 , 7–13
Emsley & Cowtan	<i>COOT: model-building tools for molecular graphics</i>	<i>Acta Cryst.</i> (2004), D60 , 2126–2132

works that have been highly cited not merely incidentally but because they have provided a recognized and enduring contribution to the science – and can rightly be considered classics. Such a criterion is, of course, not exhaustive but is useful in illustrating the variety of topics that have attained this status.

We also note that this selection includes articles published years, or even decades, before the commentary was commissioned, an indication of the durability of seminal articles. Indeed, it is instructive to compare a journal's 'impact factor' over different time spans. The current impact-factor metric published by the Thomson ISI *Journal Citation Reports* calculates impact factors over a two-year moving window. The results for crystallography journals over 2-year, 5-year and 23-year periods are listed in Table 3.

There are dangers in placing too much emphasis on individual metrics such as impact factor. Short sampling windows may reflect passing fashions, journal characteristics may be too highly dependent on individual 'maverick' articles *etc.* Nevertheless, the strong performance of IUCr journals against many such metrics, together with continuing publication of articles highly cited over a short time scale (Table 4), reinforces the picture of successful and influential titles that still retain the respect of crystallographers and the wider community of science.

In recent years, the main Editors of the IUCr journals have highlighted articles of particular note for inclusion in a regular feature in the *IUCr Newsletter*, thus bringing them to the notice of a wider public and showcasing some of the most interesting articles published in the previous few months.

4. Editorial policy initiatives

From the outset, there have been agreements with commercial publishing houses and financial pressures on the journals have reflected general trends in the global publishing market. However, the IUCr has always closely retained ownership, editorial control, and strategic development of its journals and other scholarly works.

Editorial policy is determined by the Commission on Journals (originally the Commission on *Acta Crystallographica*), formally established at the First General Assembly of the IUCr at Harvard University. It acts as the Editorial Board for the journals, each of which has one or more nominated Main Editors (or Section Editors for the separate sections of *Acta*), and a set of Co-editors who receive submitted manuscripts and are responsible for overseeing the peer-review process.

The Commission meets formally once every three years, in association with an IUCr Congress and General Assembly. At that time, there are also meetings of the individual Editorial Boards (these are sometimes combined for related titles). Informal meetings of members of the Commission on Journals also happen regularly at the major regional crystallographic meetings, allowing Co-editors to be kept informed of developments and to raise any topics of concern.

The Commission has always worked closely with the Commissions on Crystallographic Nomenclature and on *International Tables*, which were also established at the First General Assembly. Indeed, the Commission on Crystallographic Nomenclature is comprised largely of members

associated *ex officio* with the Commissions on Journals and *International Tables*. The journals' *Notes for Authors* frequently reference reports and recommendations of the Commission on Crystallographic Nomenclature.

Among the most fruitful collaborations has been that with the Commission on Crystallographic Data, which has had a profound impact on attitudes to presenting and preserving primary crystallographic structural data. Forty years ago, Kennard *et al.* (1967) prescribed a set of items that should be reported in a single-crystal structure determination. This set matches closely the content of articles published today. An emphasis on proper scientific analysis ('optimistic guesses at the accuracy attained are to be shunned') accompanied a recognition that printed tables of structure factors were less than optimal, although still preferable to complete loss of the data.

Twenty years later, the 'sub-optimal' nature of printed data was causing increasing problems for the small-unit-cell databases, especially as many chemistry journals were consigning most of the crystallographic information to a supplementary document for deposition, rather than primary journal publication (Bergerhoff *et al.*, 1986). Among the recommendations of that article, the call for a standard interchange format was influential in the IUCr's subsequent decision to develop the Crystallographic Information File (CIF) that was to have such a profound effect on its own publishing activities (Hall *et al.*, 1991).

Bergerhoff *et al.* (1986) commend the practice of the Protein Data Bank (PDB), already well established, of requiring machine-readable deposition of data. Two important consequences of the PDB's early emphasis on machine-readable deposition were that it became established as the acknowledged archive for experimental data relating to biological macromolecular structures, and therefore funding bodies could be lobbied to require deposition as a condition of funding (Berendsen, 2003); and it had the capacity to archive structure factors as well as the coordinate data, even if deposition of structure factors with the PDB is not yet universal (Wlodawer, 2007). Structure factors for small-molecule and inorganic structures are still not routinely archived by central database organizations, nor do they invariably find their way in machine-readable form to the depositories of chemical journals. This should give cause for concern in an age when computer storage facilities are widespread and relatively inexpensive. IUCr journals, of course, do require the deposition of structural data, and make such data sets openly and easily available on the Web.

The Commission on Journals has always encouraged the publication of the best attainable scientific description of crystal structures. Numerous quality assurance criteria, formal and informal, have guided Co-editors and referees in their acceptance or rejection of publications. The adoption of guideline criteria for assessing quality (such as 'target' *R* factors) was sometimes contentious, since the particular circumstances of any experiment might affect what was attainable; there was also the risk that authors might introduce biased procedures into their data analysis in a misguided

attempt to satisfy such criteria. Nevertheless, continuing interest by the Co-editors in issues of quality assurance has been central in highlighting what should be regarded as best practice, and has subsequently informed the standardized machine-based checking of results that became possible after the adoption of CIF (§6).

5. Early developments in automation

The material in this section draws from a series of internal reports written by the editorial office staff to describe and develop early work in introducing a computerized workflow. These reports have been deposited as supplementary materials.¹

5.1. Archaeology

By the late 1980s, with several sections of *Acta Crystallographica* and the *Journal of Applied Crystallography* in production on a monthly or bimonthly schedule, the staff complement in the editorial offices was growing and a computer had been introduced to help with managing workflow and typesetting indexes and contents listings.

The first computer, a Systems Group 2900 Series machine, was purchased in 1984 with a generous gift from the Japanese crystallography community (and was housed in an attractively carpentered wooden cabinet). It had 512 kB memory and 34 MB hard disk storage. A system of programs in Basic was commissioned to maintain records of manuscripts being processed, prepare and sort cumulative indexes, store formulae of structures reported in *Acta* and search for duplicate submissions. The indexes were created in a proprietary typesetting language and passed to a typesetting house for printing of annual and five- or ten-year index booklets.

In 1989, a Rank Xerox Documenter desktop publishing workstation was used to produce the camera-ready copy for the new *Fast Communications* section of *Acta*. It had a user-friendly graphical interface but was slow, relatively expensive, and had limited abilities to inter-operate with other systems. It was retired after a few years in favour of more general solutions.

In 1988, a DOS-based PC was also acquired to run checking software, to begin to offload from Co-editors the burden of checking structural results. A pilot scheme was run on this machine for three years using the program *PARST* (Nardelli, 1983), developed for such a purpose by Mario Nardelli when Editor of *Crystal Structure Communications* (see §2.3).

Editorial staff keyboarded the essential data fields required by the checking software. When the draft specifications for the Crystallographic Information File were circulated to the IUCr office, it seemed sensible to store these data in CIF format. Software was written to transform the standard CIF data to the other formats required by *PARST* and other programs that we were starting to use. In this way, the evolving CIF

¹ Supplementary data for this paper are available from the IUCr electronic archives (Reference: SC5008). Services for accessing these data are described at the back of the journal.

specifications were tested against hundreds of real data sets before being released to the wider world. This process was an important step in tailoring CIFs to the real requirements of the journals.

Also in 1988, the IUCr acquired its most significant new computer hardware, a Unix machine (AT&T 3B2/500, 4 MB RAM, 147 MB disk). The existing software on other computers was gradually ported across, and the machine acquired Fortran and C compilers, PostScript (the page description language which was becoming the standard interchange format for printing), and the open-source \TeX typesetting package. With this equipment, the editorial office was capable of implementing emerging new technical standards and undertaking innovative in-house development.

5.2. A database of high-quality bibliographic metadata

An early function of the new Unix machine was the improved generation of journal indexes. Existing procedures for generating and checking proprietary files for external typesetting were complex, slow and prone to error. New procedures were developed to create PostScript files that could be readily proofed and modified. This immediately made production of printed indexes faster, more efficient and cheaper.

It also made it possible to create online searchable plain-text index files, and all the existing index files accumulated over the previous decade were converted in this way. A simultaneous project to print a complete index for *Journal of Applied Crystallography* was also adapted to generate such electronic files. Since they were intended for creating printed indexes for publication, the transformed files were subjected to close editorial scrutiny and proof-reading. The result was a high-quality collection of tables of contents and indexes that was easily tagged to produce machine-readable files suitable for loading databases with reliable bibliographic metadata.

Already in 1992, we were beginning to make these tables of contents and indexes available over the Internet, using the fledgling *gopher* and *WAIS* wide-area information protocols. When the http protocol was introduced and the World Wide Web became a reality, these contents pages were translated again, this time to HTML, and a project was undertaken to scan remaining printed contents listings, convert them to readable text by optical character recognition, and thus create for the Web a complete searchable contents list – in effect, a bibliographic metadata database – for the complete run of the journals. This was achieved by 1998, and allowed high-quality Web searches on authors and titles of all published research articles.

When the decision was taken in 1999 to scan the entire set of printed volumes of the IUCr journals for Web distribution, the existence of such a complete and reliable metadata database was crucial in setting up an automated workflow to drive this process. From the metadata files, subsequently to be converted to a true relational database known as the Published Items Database (PIDB), an inventory of journal articles with their extent was easily created, and worksheets

generated for the outsourced scanning operation, overseen by the UK's Higher Education Digitization Service. The completion of the digitization project during 2001 saw the availability of the complete run of IUCr journals from 1948, an achievement significantly in advance of other scientific publishers' initiatives in the same spirit.

The PIDB has also allowed us to create metadata feeds that nowadays populate the databases of secondary abstracting and indexing services, and provide the information needed for dynamic hyperlinking to articles through the inter-publisher CrossRef service (<http://www.crossref.org>).

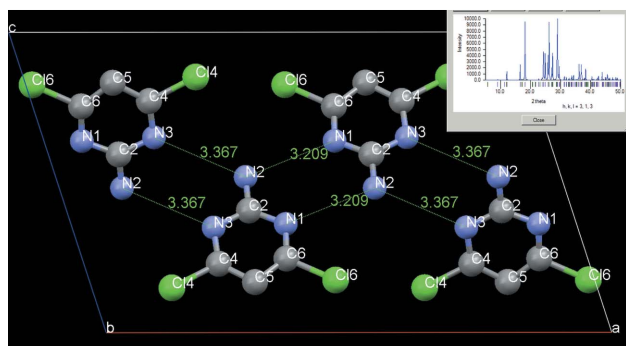
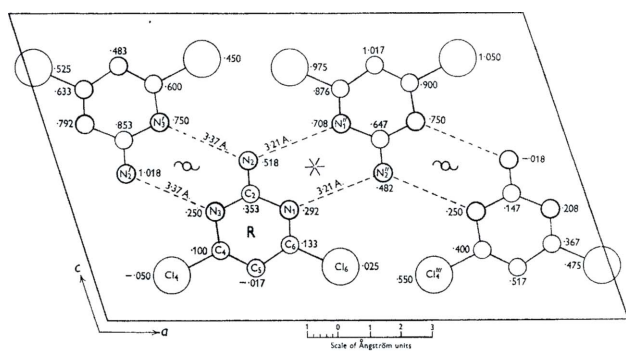
6. Data validation and online submission of structure reports

Until the arrival of CIF as a standard interchange format, many referees undertook manual checking of experimental and modelled data in the paper under consideration. Section 5 describes how the editorial office began to test centralized checking on behalf of referees, using CIF as an interchange source to produce input files for different checking programs.

For a number of years, the Chester editorial staff supervised a semi-automated procedure in-house that created a synoptic validation report based on output from a number of programs (McMahon, 1996). The report was often lengthy, containing the tabulations and calculations output by software that had been designed to create structural models rather than to evaluate them critically. Editorial staff, assisted by rudimentary machine intelligence that highlighted words or phrases indicating potential errors or anomalous values, provided some help to the referees in interpreting these reports. However, there was still a need for more concise and targeted evaluation.

During the latter half of the 1990s, under the guidance of Syd Hall and then George Ferguson as Editors of *Acta C*, clearer algorithmic measures of quality and self-consistency of structural reports were developed. In 1996, CIF was mandated as the only submission format for *Acta C*. Once again, in requiring electronic submission, the IUCr was anticipating developments in electronic publishing that would not become widespread for several years. In 1997, data validation criteria were published, relating to specific CIF data items. To resist financial pressures arising from a rapid growth in the number of printed pages, shorter articles were made available in electronic format only, presaging the launch of a purely online journal, *Acta E*, in 2001.

By the time *Section E* was launched, the checking procedures were much more clearly defined, thanks largely to the development of a specialized checking module in the program *PLATON* (Spek, 2003). *PLATON* became the crystallography validation engine, while other routines written in-house became the publication component of a public *checkCIF* service that authors could use prior to submission (and that formed the basis of the evaluation reports received by the journal referees) (Strickland *et al.*, 2005).



```

Alert level A
ATOM007_ALERT_1_A _atom_site_aniso_label is missing
Unique label identifying the atom site.
GEO0001_ALERT_1_A _geom_bond_atom_site_label_1 is missing
Label identifying the atom site 1.
GEO0003_ALERT_1_A _geom_bond_distance is missing
Distance between atom sites 1 and 2.
GEO0006_ALERT_1_A _geom_angle_atom_site_label_2 is missing
Label identifying the atom site 2.
PLAT027_ALERT_3_A _diffn_reflns_theta_full (too) Low ..... 0.00 Deg.
PLAT029_ALERT_3_A _diffn_measured_fraction_theta_full Low ..... 0.00
PLAT049_ALERT_1_A Check Reported Molecular Weight ..... 0.00
PLAT049_ALERT_1_A Check Reported Molecular Weight ..... ?
PLAT150_ALERT_1_A Volume as Calculated Differs from that Given ..... ?
PLAT197_ALERT_1_A Missing_cell_measurement_temperature ..... ?

Alert level C
PLAT040_ALERT_1_C No H-atoms in this Carbon Containing Compound .. P21/a ?
PLAT122_ALERT_4_C Non-standard setting of Space group P21/a ..... ?
PLAT151_ALERT_1_C No su (esd) Given on Volume ..... ?
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... N1
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... N2
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... N3
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... C1
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... C2
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... C3
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... C4
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... C5
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... C6
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... Cl1
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... Cl2
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... Cl3
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... Cl4
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... Cl5
PLAT161_ALERT_4_C Missing or Zero su (esd) on x-coordinate for ... Cl6
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... N1
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... N2
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... N3
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... C1
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... C2
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... C3
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... C4
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... C5
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... C6
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... Cl1
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... Cl2
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... Cl3
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... Cl4
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... Cl5
PLAT162_ALERT_4_C Missing or Zero su (esd) on y-coordinate for ... Cl6
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... N1
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... N2
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... N3
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... C1
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... C2
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... C3
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... C4
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... C5
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... C6
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... Cl1
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... Cl2
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... Cl3
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... Cl4
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... Cl5
PLAT163_ALERT_4_C Missing or Zero su (esd) on z-coordinate for ... Cl6

10 ALERT level A = In general: serious problem
0 ALERT level B = Potentially serious problem
30 ALERT level C = Check and explain
0 ALERT level G = General alerts: check

10 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
0 ALERT type 2 Indicator that the structure model may be wrong or deficient
2 ALERT type 3 Indicator that the structure quality may be low
28 ALERT type 4 Improvement, methodology, query or suggestion
0 ALERT type 5 Informative message, check
    
```

Figure 1
The first crystal structure reported in *Acta Crystallographica* (Clews & Cochran, 1948). (a) Selected intermolecular distances as given in Fig. 5 of the original paper. (b) The same information shown graphically with modern visualization software (Macrae *et al.*, 2006). A simulated powder diffraction pattern calculated by the same program is also shown. (c) *checkCIF* report on the data provided by the authors. Although the original paper omits many items that would today be regarded as important, there is nothing to suggest significant error in the model.

There was some initial resistance on the part of authors to the idea of increased automation in the data reviewing process, but the launch of *Section E* provided an opportunity to demonstrate how such validation allowed the objective characterization of the reliability of numeric values judged against agreed criteria (Fig. 1). The *checkCIF* reports for every structure published in *Section E* are made available as supplementary documents from the online table of contents pages; and include *key indicators*, a small set of data items extracted from the CIF that are felt to characterize the overall nature and quality of the study.

As mentioned in §4, ‘poor-quality’ structure determinations (as indicated by various validation metrics) do not automatically exclude an article from publication. Nevertheless, as experimental procedures continue to improve, structures that do generate significant validation alerts probably reflect a problem with the crystal quality, the performance of the experiment, or unusual chemistry – all of which need to be discussed in the descriptive text of the article.

The submission system therefore checks each CIF *as it is submitted* and will not accept it if severe validation alerts (those classed as ‘category A’ in the *checkCIF* classification scheme) are generated, *unless there is a formal statement within the CIF accounting for the presence of such alerts*. This formal statement is supplied as a private CIF data field; the submission system returns to the submitting author a proforma CIF fragment (a ‘validation reply form’) that may be completed and embedded in the CIF to allow the submission process to continue.

This type of automated filtering efficiently alerts referees to those cases where particular judgement must be exercised over acceptance criteria on the grounds of data quality. It greatly assists the review process by highlighting potential problems but does not usurp the considered judgement that still underpins and provides trust in scholarly reviewing practice.

Despite arguments on points of detail, the Commission on Journals has over the years shared a broad consensus of appropriate metrics for the quality of small-unit-cell structure determinations. The same is not yet uniformly true of biological macromolecules. The field is itself younger and the availability of high-resolution structural models rather more recent. There is also more scope to debate appropriate indicators for the large and complex structures of protein and nucleic acid molecules. Nevertheless, structure analysis programs such as *WHAT IF* (Vriend, 1990) and *PROCHECK* (Laskowski *et al.*, 1993) offer similar analyses and reports to those originally generated for small-unit-cell structures. It may be that the community will move towards identifying particular validation indicators based on individual mmCIF data items as a basis for future characterization and filtering of protein structure reports. Steps in this direction have been taken recently by Howard Einspahr and Mitchell Guss in establishing recommended data items to describe structural and crystallization communications for proteins and nucleic acids in *Acta F*.

7. Design considerations for Crystallography Journals Online

By 1999, the transition to a computer-based production system had made great progress and we were ready to consider the launch of an online journals service.

7.1. On shoulders of giants

As with most innovations, the experience of others was important in helping us to clarify our objectives, select our methods, and learn from failed or unfulfilled experiments. Many of our early investigations were guided by the advice, encouragement and support of Ted Maslen and Howard Flack, the first two Chairs of the IUCr's Committee on Electronic Publication, Dissemination and Storage of Information, who took a close scientific interest in developments in the publishing and information technology industries.

Electronic typesetting systems had been developed during the 1980s [*t_roff* for Unix, T_EX (Knuth, 1986) for mathematical articles] that allowed publishers to design and develop more lightweight and flexible workflows for composition, printing and possible electronic delivery. The adoption of word-processor packages on affordable personal computers was beginning to allow authors to submit their work electronically. Publishers began to understand the need for structured markup of electronic documents (to allow them to be distributed in different media and increase their functionality in online database searching and linking applications). They helped to develop Standard Generalized Markup Language (SGML), which became established as the industry-standard markup protocol [and led eventually to the HTML (Hypertext Markup Language) of the Web and the XML (Extensible Markup Language) of current structured document handling].

Many learned-society publishers, especially those in the mathematical and physical sciences, developed T_EX-based publication systems for their journals. Learning how to use T_EX required substantial investment of effort by their authors, but the reward of high-quality typesetting of complex mathematics won the support of these academic communities. Structured versions of T_EX, such as L^AT_EX (Lamport, 1986), were developed and started to be adopted by some commercial publishers of mathematics-rich scientific journals. The use of structured markup was important for automated workflow and to allow integration with SGML workflows used for other journals.

In anticipation of the need to handle electronic journals in academic libraries, several pilot projects were initiated with participation from commercial publishers, academic libraries, and government funding bodies. Among the best known were the Red Sage Project (University of California, San Francisco, in association with AT&T Bell Labs, Springer Verlag, and several commercial and learned-society publishers), CORE (Cornell University), TULIP (Elsevier Science) and CEDARS (Consortium of University Research Libraries, UK). We were particularly influenced by the results of the UK SuperJournal pilot, undertaken as a response to the Follett Report of 1993 (<http://www.ukoln.ac.uk/services/papers/follett/report/>).

Although we were not active participants in SuperJournal, we did liaise with the project coordinator, David Penfold, who had been Technical Editor of the IUCr journals from 1976 to 1985.

A valuable review of electronic publishing in science at this time is the 1996 Proceedings of an ICSU Press/UNESCO Expert Conference, at which the IUCr was represented by André Authier (IUCr President 1990–1993), Syd Hall and editorial staff members (Shaw & Moore, 1996).

Until the mid-1990s, we remained interested observers of all these developments; but the likely development costs of full-scale investment in the new technologies were worryingly high. By the middle of the decade, however, lower-cost technologies were becoming established that seemed promising for long-term development. They included PostScript and PDF page-description languages, T_EX as a typesetting system, powerful image-editing tools and, most importantly, HTML as a lightweight but powerful dissemination protocol. Coupled with the widespread availability of personal computers, graphics workstations and increasing bandwidth in academic research networks, they encouraged us to start developing the necessary skills and resources to take the next step.

7.2. Preparation for electronic publication

Starting in 1995, a document type definition (DTD) was developed to allow IUCr journal articles to be marked up as electronic files in SGML. The philosophy of SGML (see previous section) was to embed markup codes within the text of an electronic document that represented the structure or purpose of pieces of text, rather than specifying purely typographic or ornamental aspects of rendering the text. Typography would be overlaid through style files that associated particular rendering styles with the different structural elements. The IUCr DTD was informed by the requirements of ISO Standard 12093 for scientific articles and was similar to existing examples used in the publishing industry.

The acquisition of software to edit SGML files and of composition software to allow the creation of formatted pages allowed us to begin serious in-house composition, although initially the tasks were shared with external contractors. SGML was designed to allow reuse of documents in contexts other than printing and so the same source files could be used to create HTML pages for the Web.

In 1999, the World Wide Web was already in its eighth year of operation outside its birthplace at CERN and was becoming established as a serious platform for information delivery. Many science publishers were facing the challenges of providing services over the Web (challenges that were no longer technical but social and financial) and the IUCr was investigating the use of a third-party service provider to share some of the risks and develop the potential of the new medium.

7.3. Design goals

However, the design goals that the IUCr identified to serve the crystallography community (Table 5) could not all be met

Table 5

Design goals for the IUCr online journal project.

These goals were specified in an internal report presented by the authors to the IUCr Executive Committee in August 1999 (see supplementary material).

1	The full contents of the printed journals should be available over the Internet
2	High-quality page images should be available
3	Full-text articles should be suitable for Web browsers
4	Access to individual articles should be by subscription
5	The contents [lists] of individual issues should be freely browsable
6	The contents should be searchable by author and by title, and in due course also by keyword
7	Full bibliographic data should be available for all articles
8	Supporting (supplementary) materials should be freely available for all articles
9	Access should be provided to articles already published
10	The IUCr should have full control over the look and feel of the online journal
11	The online journals should be integrable with the existing IUCr pan-crystallographic Web services

by the major platforms then under development. With the experience we had already gained in running a Web server since 1994, we believed that we could meet all the goals ourselves and so a pilot service hosted on a server in the Chester office was introduced at the Glasgow Congress, becoming a fully functional service with paid-for access early in the following year.

7.4. Supporting materials

Of the design goals, the one that was felt to be most poorly addressed by other potential service providers was goal 8, the free (and easy) provision of supplementary material to all readers. With the journals' long-established requirements for deposition of structural data, and because of the deposition of CIF data sets since 1991, we had a large collection of structural data whose importance had repeatedly been emphasized (Kennard *et al.*, 1967; Bergerhoff *et al.*, 1986). There were also other supplementary documents deposited in support of many articles. In the proposed **Crystallography Journals Online** service, the supplementary documents would be as easily accessible as the primary articles themselves.

This feature remains an important component of the online service today and is one, we suggest, that is still poorly implemented on other online journal platforms.

Access to supporting data is very important in allowing readers to interact with the science they are reading by visualizing or processing the data directly. We host movies and animations as supplementary files, demonstrating such things as protein conformation changes in lobster colourization (Chayen *et al.*, 2003), time-lapse improvement in the diffraction limit of protein crystals under dehydration (Bowler *et al.*, 2006), mosaicity changes through cryogenic cooling (Lovelace *et al.*, 2006) and microgravity crystallization in space (Snell *et al.*, 1997).

The CIF archive is, of course, the most powerful supplementary data resource. Since the launch of the online service, we have provided a means to associate CIFs with browser

helper applications so that a reader may launch any suitable visualization tool of choice. Indeed, the mechanism, which relies on chemical MIME types (Rzepa & Murray-Rust, 1998), allows the association of any software application, not just a visualization program, with a downloaded data file. More recently, '3D view' links from papers with associated crystal data have been augmented to show rotating interactive views of the structures reported, using the popular Java plug-in *Jmol* (<http://www.jmol.org>).

We believe that continued support for associating helper applications with particular MIME types will be an important factor in broadening the usefulness of files of varied type disseminated from publication platforms, and urge developers of new-generation Web browsers not to remove altogether or needlessly obscure such functionality in the drive to make their software more 'user-friendly'.

7.5. Other considerations

Design goal 9 (Table 5) was directed at the proposed subscription model – access would be provided for the duration of a subscription to *all* articles available from the journal subscribed to, rather than guaranteeing long-term access to the content published during a particular year. However, it also acted as an incentive to proceed with the suggestion to digitize the existing back catalogue of printed journals, and thus led directly to the availability of all the IUCr's legacy journal content online.

In addition to the tabulated design goals, the internal report outlining the prospects for an electronic journal called upon the Executive Committee to review any opportunities that the new medium afforded for providing increased access to the crystallography literature to developing countries. This subsequently bore fruit with the launch of a Journals Grant Fund. It also informs the IUCr's response to the Global Information Commons for Science Initiative launched in 2006 under the aegis of a number of scientific and development organizations (see <http://www.iucr.org/iucr-top/iucr/gicsi/positionpaper.html>).

Interestingly, the design goals do not specify rich hyper-linking within articles, between articles or from articles to other Web resources, such as other journals, databases or search engines. It is clear from other statements within that report that such functionality was taken for granted, as part of the very reason for placing the scientific literature online.

Considerable thought and effort went into optimizing the navigational tools within the online journals site and the basic elements of the original design are still in use.

In recent years, the journals have also been made available through the Blackwell Synergy platform, operated by our commercial publishing partner, to allow greater access to library communities.

7.6. Pace of developments

It is interesting to compare what we have achieved – and how rapidly we have achieved it – with other learned-society and commercial publishers.

We explained in §7.1 how our vision of an electronic journal and our efforts to produce one were influenced by the early experiments of many publishing organizations. In one sense, we started late; as we explained, the high development costs of the novel technologies were initially prohibitive. Large commercial publishers had the capital to fund early research and to invest in the specialized and expensive hardware that was initially required. Some large learned-society publishers had sufficient turnover to operate in the same way; the IUCr was not one of these.

Society publishers for some communities (mathematics, physics, astronomy *etc.*) benefited from the free availability of T_EX, as well as the commitment of their authors to use this new software. We were able to take limited advantage of T_EX. Not all our journals are mathematics-rich – although we are a small organization, we serve a broad range of scientific communities, with different publishing requirements.

And so our move into electronic publishing was relatively late. However, once the necessary technologies came within reach, we were able to move swiftly. We had the benefit over the earlier adopters of not needing to retain experimental, but already outdated, work practices. We were able rapidly to build the **Crystallography Journals Online** platform as a service that is comparable to those of many larger publishing houses.

We were, as described above, early in disseminating full lists of contents. We were also early in handling data, largely on account of the IUCr's longstanding commitment to the provision of data in support of scientific research findings; and we believe that we are still at the forefront in that area.

Now, of course, we have an established workflow, and a need to support it and its necessary computing infrastructure. Radical revisions of the electronic journals service are increasingly unlikely; future developments will most likely consist of additional novel services that can be put in place without disturbing the existing systems. Nevertheless, we confidently expect many more developments in the years ahead, even if the pace is slower than we have become accustomed to in the last decade.

8. Electronic journals and the World Wide Web

The revolution brought about by the World Wide Web has so changed our habits that it is difficult to remember how learned journals operated just a few years ago. Yet the changes have been remarkable. In his 40th-anniversary review of the IUCr's publication history and prospects, the then Editor Sidney Abrahams (Abrahams & Matula, 1988) was able to look ahead to journal distribution in media other than paper, inspired by the novel technologies of CD-ROM and digital video cassette. Now, a generation later, the existence of CDs and video cassettes even as entertainment media is threatened by the Internet, and scientists are close to realizing the dream of having all the world's learned publications immediately accessible from the desktop.

Many of the technologies of the Web that we have already mentioned combine to provide an incredibly powerful digital

library. Comprehensive databases of bibliographic metadata facilitate hyperlinks from reference lists direct to the cited articles. Powerful full-text search engines index all accessible publications on the Web, whether the informal or commercial public Web pages that Google covers so comprehensively or more selective indexes of scholarly publications, refereed or not (*e.g.* Google Scholar, Scirus, Microsoft Windows Live Academic Search). 'Forward citations' list articles that cite the current one. Reference manager software (and citation-sharing Web sites) facilitate the exchange and critical collection of literature databases. And, increasingly, journals can provide access to the supporting scientific data and allow the reader to interact with the science reported by an author (Helliwell *et al.*, 2006).

9. Publishing technologies

9.1. Markup languages

The key to electronic publication – and, more generally, information dissemination – is structured document markup. We explained in §7 how SGML allowed the same source files to be used for printing and for Web publication. We have shown how our earliest efforts in online publication had succeeded because of CIF, a structured information markup language. All our journals are now created from SGML files (CIF content for publication is also converted to SGML).

SGML remains important but is now complemented (and, for some publishers, replaced) by XML, the markup language of choice for new information applications. XML is a particular form of SGML, used to markup data sets as well as textual documents. As such it may, in future crystallographic publishing applications, marry the best aspects of SGML, primarily a text markup language, and CIF, primarily a data markup language.

It is likely also to form the basis for future 'datuments', a term coined by Murray-Rust & Rzepa (2004) to describe 'a hyperdocument for transmitting and preserving the complete content of a piece of scientific work', and that typically will include interactive online content with text and data.

XML is well specified but has almost unbounded freedom in what the markup should represent. Understanding the content of an XML file depends on DTDs, schemas or other formal specifications of what the markup represents. To be useful for interoperability or as standard exchange mechanisms, these specifications must be in the public domain.

Of the current submission formats used by authors, Microsoft Word, a proprietary word-processing application, offers least in the way of structured markup, and is therefore most difficult to convert automatically to SGML. We address this problem by asking authors to use a document template we have prepared that has concealed structural cues within its typographic styles.

T_EX (Knuth, 1986), designed as a typesetting language for mathematics-rich documents, is itself rather unstructured but has powerful macro programming features allowing typographic directions to be embedded in structural markup tags.

A particular implementation of such macros is L^AT_EX (Lamport, 1986), common in the physical and mathematical fields, and particularly useful as a submission format because of its ease of conversion to SGML. As with Word, we have provided a L^AT_EX template to authors that provides additional structure helpful for automated conversion to SGML.

CIF, as mentioned above, is also structured and lends itself well to format conversion, though it has limited ability to encode detailed markup within text fields.

9.2. Handling special text, mathematics, chemistry and graphics

There are still many challenges in the effective Web presentation of complex technical material. The Unicode standard allows better handling of accented and multilingual text but there are still considerable problems with mathematics. The XML-based markup language MathML, coupled with new scientific fonts designed for Web publications, will help to address these for much straightforward mathematical work, but MathML has limitations in marking up the very complex mathematical expressions that appear in some of our articles.

The portable and information-preserving handling of chemical formulae and schemes could also be improved. Initiatives such as chemical markup language (CML) and the International Chemical Identifier (InChI), which encode aspects of chemical connectivity, stereochemistry, valence *etc.*, have much to offer in this respect. The InChI also has the ability to act as a unique and exact identifier of chemical species, and therefore will be pivotal in providing hyperlinks to other Web resources referring to a given chemical species.

Graphics also represents an area where much remains to be done. Raster images for publication are reasonably well handled by a number of standard formats (*e.g.* TIFF, JPEG, PNG). However, graphs and line illustrations often contain much information which it would be better to encode in a scale-independent form, ideally one that was able to carry along the semantic information that the figure was aiming to represent. Scalable vector graphics (SVG) offers some potential here but seems to be slow in development.

In the future, it may also be possible to publish experimental scientific image data, for which fewer standards exist. In crystallography, the imgCIF standard is likely to be important for this purpose.

Also beginning to appear in online journals are natural-language-processing technologies, where software can process suitably marked electronic text and layer hyperlinks to glossaries or other relevant online resources, generate contextual indexes or search strategies, or analyse and insert semantic tags ('argument', 'citation', 'evidence') within sentences that are couched in the formal language of scientific communication.

These latter steps begin to take us towards a realization of the 'semantic Web', the use of the World Wide Web not simply as a dissemination medium but as a knowledge base.

10. The future of the learned journal: policy

So it is clear that early in the 21st century continuing advances in information and computing science are poised to drive forward the evolution of the electronic journal. We shall return to this technical potential in the following section.

However, the policy evolution of scholarly journals may exert even greater influence. In analysing and trying to respond to policy issues, the IUCr benefits from its contacts and collaborations with other organizations. We greatly value the advice, discussions and considered analyses from scientific interdisciplinary bodies such as the International Council for Scientific and Technical Information (ICSTI) and ICSU's Committee on Data for Science and Technology (CODATA); from trade organizations such as the Association of Learned and Professional Society Publishers (ALPSP); from partnership with commercial publishers such as Wiley–Blackwell, Springer and Oxford University Press; and from academic community bodies like the Joint Information Systems Committee (JISC) and the Research Information Network (RIN) in the UK.

10.1. Open-access publication

There is growing pressure for 'open access' to the scientific literature. The phrase is used in many ways, but all postulate that the end user may read scientific articles at no direct cost, and without access restrictions. In itself, this is a goal which the IUCr as a learned society considers very desirable. 'Openness', in its many aspects, is a central part of the scientist's personality and culture. Another major driver is the increasing requirement of funding bodies [*e.g.* the National Institutes of Health (USA) and The Wellcome Trust] that the findings of research paid for out of public or charitable monies should be freely available. However, 60 years of experience in publishing scientific literature to the highest quality have demonstrated to the IUCr the very real costs necessary to administer, organize, review, assess, collect, edit and distribute scholarly communications. A proper open-access economy must ensure that these real costs are met in some way.

In our own open-access initiatives (first the voluntary open dissemination of individual papers, and now the new arrangements for *Acta E*), we have relied upon the 'author-pays' model, perhaps the most common approach taken by open-access publishers. This has advantages of tying costs to individual contributions, allowing a natural scaling, but it challenges existing funding strategies, places financial burdens on poorer authors or institutions, and in the worst scenario might introduce potential conflict of interest between quality assurance and the need to maintain an appropriate cash flow.

10.2. Economic factors

10.2.1. Peer review. Of course, real cost savings can be achieved – and have been – through electronic processing and automation. However, the relative magnitude of these savings is often overestimated. A significant part of the real cost of journal production remains in the peer review process, which is still very highly regarded amongst scientists. Suggestions are

sometimes made to replace the traditional process by Web-based discussion forums, open to all. An experiment by *Nature* (2006) suggested that the scientific community is not yet ready to embrace this approach, although the journal believes that the problem is more one of timing than of inherent shortcomings in the approach. Certainly, the growing use of the Web for social networking, informal communication through blogs, and community discussion forums demonstrates that it is becoming a much more important channel for active communication, and it may well be that a younger generation of scientists will come to demand more use of such mechanisms. Our own opinion, however, is that such media invite immediate response, and may not encourage the considered and thoughtful criticism that is common in existing peer-review practices. Of course, we must remain responsive to genuine cultural changes and accommodate them – but there is no real evidence that this will materially affect the amount of effort (and therefore real cost) of peer review for some considerable time to come.

10.2.2. Other production costs. Other potential cost savings related to quality assurance seem to lie with the automated generation of checking reports of crystal structures that we already provide to referees. However, experience again shows that the computer cannot make suitable judgments; considerable investment of time is required by referees and Co-editors to interpret the significance of the *checkCIF* diagnostics in relation to the scientific findings described in an article.

Traditional copy-editing skills are also an expensive, but necessary, component of overall publishing costs. Many enthusiasts claim that scientists can and do write clear and comprehensible articles unaided; but such enthusiasts are usually themselves talented writers. Science is an international endeavour, and a growing proportion of the world's scientists are writing in a language that is not their mother tongue. Even native language speakers often benefit from a professional editorial eye.

A contemporary publishing house invests considerably in research and development, to stay abreast (or, ideally, ahead) of the rapidly changing technologies of the information age. There are also costs in collating and distributing metadata to secondary publishers, to abstracting and indexing services, search engines and libraries. Perhaps of greatest importance is the need also to lay down infrastructure for the long-term preservation of electronic publications, and continued access to the literature as the technology of the Internet and of individual computer workstations evolves rapidly.

Certainly a low-cost pilot or prototype publishing operation can be established with relatively little effort; the actual costs of converting a pilot scheme to a trusted, reliable and permanent service remain substantial.

10.2.3. External market drivers. Other changes in the publishing environment have potential impact on the future evolution of scientific journal publishing. Increased focus on citation metrics as a basis for research funding or career development is bound to influence the pattern of publishing. Other quantitative measures of the usage of journal literature

are now possible: access statistics provide data (though not necessarily insight) into the number of readers of individual articles. This may reflect perceived scientific value; equally, it may lead to decisions by libraries to cancel subscriptions depending on perceived volume of use. Or it may contribute towards a move away from subscriptions and towards individual-article-purchase mechanisms.

Library budgets are also placed under pressure by funding caps and by arrangements with large publishers to acquire many additional titles at a greatly discounted rate in so-called 'big deal' agreements, with a consequent detrimental effect on subscriptions to small society publications.

At this point, the slow downward spiral of traditional library subscriptions seems likely to continue, with consequent challenges to financial stability. Publishers, ourselves included, attempt to counter this with annual subscription price increases, but this squeezes static library budgets further, and risks further loss of subscriptions. However, at least in the short term, consortial deals help to counter this. They often require libraries to maintain existing subscriptions, so that subscriptions held within such a consortium are safeguarded. In the case of consortia that did not have previous subscriptions, such deals provide a small additional revenue. For IUCr journals, the readers served by these consortia now have access to crystallography journals, and the increased visibility of our titles is a great benefit.

This may, however, just provide a breathing space. It is possible that new subscription models will emerge, based for example on access to individual articles rather than to journal collections. It is clear that the IUCr will need to remain alert to any such changes, although they will of course affect all scientific publishing businesses.

'Author-pays' open-access publication is becoming more popular (see §10.1) and raises new challenges in providing value for money to authors or their funding bodies. At this time, it is far from clear to what extent open access might in future guarantee publishing income. We are gaining some experience through our experiments with a hybrid model and with the introduction of full open access for *Acta E*. However, the future will depend on many factors outside our control – how different academic communities respond to the new publishing model, how funding bodies react to the new economics; and how well the traditional peer-reviewed journal survives in the face of further technical developments.

Preprint servers, informal peer review and community circulation of unrefereed articles may come to challenge traditional publisher roles. The response of scientific communities to these new practices differs from one discipline to another. Since crystallography is by its nature an interdisciplinary science, each of these developments is likely to impinge upon the future directions of one or other of our journals.

10.2.4. Data publication at source. An example of a possible real change in publishing patterns is the growing practice of service crystallographers or structural chemistry laboratories depositing data sets in local repositories, open to the Web through various open-access technologies. These data

sets may or may not be associated with subsequent publications but are available for harvesting by structural databases and can be cited using persistent digital object identifiers (DOIs). In many cases (e.g. eBank-UK <http://www.ukoln.ac.uk/projects/ebank-uk/>) *checkCIF* reports accompany the data sets as an indicator of their quality.

As structure determination becomes increasingly routine, a growing number of publications have little novel to say about individual structures and published structure reports become less distinguishable from well annotated data sets. It may be that the continuum of detail in description of new structures, ranging from novel scientific principles or properties in *Acta B*, through interesting characteristics of individual compounds or families of compounds in *Acta C* to routine but novel structures reported in *Acta E*, may extend into unrefereed data 'publications' in such repositories. We hope that quality assurances born in the journal publishing process will inform such novel dissemination channels and ensure that such public data sets are prepared to the highest reasonably attainable quality.

The structure repositories also in some cases offer links to diffraction images and other raw data sets archived in high-capacity data stores. This practice offers valuable insight into ways to federate data storage facilities, and ultimately to arrange for the long-term archiving of a full collection of experimental data. It is possible that future policies may lean towards mandatory deposition of diffraction data images.

11. The future of the learned journal: technical

We described in §9 a number of technologies central to the continued development of electronic journal publishing. We might add to those the procedures for creating and maintaining blogs, RSS feeds, podcasts, Wikis, social bookmarking, folksonomies and the social networking platforms, together with the harvesting and overlay of data from different sources to create 'mash-ups' that collectively form the 'Web 2.0' view of networked information dissemination (O'Reilly, 2005).

In fact, we believe that this proliferation of applications certainly indicates great vigour in Web development, but does not represent a fundamental change in how the Web is used. Our own opinion is that major changes to the way the Web functions will arise from novel protocols rather than distinct applications. Two that we would identify as significant for future directions on the Web are OAI-PMH and openURL (see below), although these probably also represent evolutionary developments rather than the basis for a new paradigm shift.

OAI-PMH, the protocol for metadata harvesting of the Open Archives Initiative, allows automated cataloguing and indexing of Web resources, including time-stamping and version identification. Coupled with suitable server discovery mechanisms, it opens the prospect of collecting not only metadata but also (with suitable access rights) complete resources, and thus allows the establishment of centralized archives, indexes and search engines; or the possibility of federated entities providing these services. OpenURL is a

protocol allowing the standard exchange of queries with a Web resource identified by a URL. Although the original specification for openURL assists bibliographic discovery (it is a mechanism by which libraries can associate their local holdings with publisher Web sites), it is extensible to other areas. We see it, coupled with the unique identification of digital objects by the DOI mechanism, as a potential vehicle to cite and address subcomponents both of complex documents and of data sets.

We close with one example of how current technical developments may have a profound impact on the scientific literature in future. Fig. 2 shows an oligonucleotide structure (Creze *et al.*, 2007). Already the abstract cartoon representation and colour selection provide an information-rich view of the molecule. However, the online version of the supplementary figure, which shows (initially) the same view, is interactive; the reader may manipulate the structure in three dimensions. The supplementary figure also possesses the facility to run author-created scripts that allow animations or selections emphasizing specific features of interest.

The content that is driving this representation in an interactive applet is the structural data set itself, but the combination of critical annotations by the author with the reader's ability to interact directly with the data maximizes the information content of the presentation. The reader may use features of the applet to interrogate the data directly for bond distances or angles, or may select individual features for closer inspection. It is also possible to overlay representations of other structures and compare them visually.

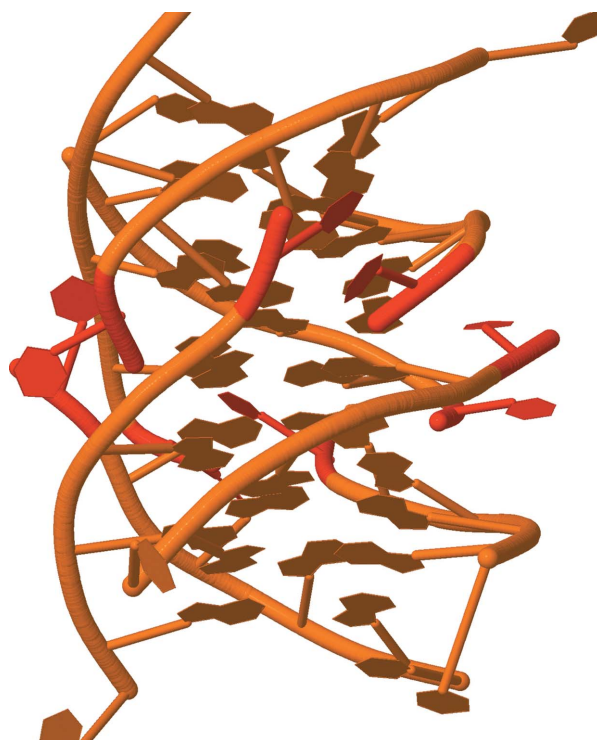


Figure 2
A d(TGGGGT) oligonucleotide shown in cartoon representation.

The annotations (in the form of author-supplied scripts) have been created in this case with a toolkit that we provide for this purpose. In practice, CIF already contains data items that could be used to embed such annotations within the data file itself. Given that CIFs may also contain the text of a paper (and, *via* imgCIF, other graphical material), it is certainly feasible that future scholarly communication could consist of a single file suitable not merely for reading but for full interaction with the reader.

Indeed, we have already seen, as long ago as 1999, an impressive application that used annotation fields of a type supported by mmCIF for just this purpose [this program was an early development version of *BioEditor* (Yang *et al.*, 2003)]. We suspect that one reason this software has not already found its way into the scholarly publishing environment is that it is for most prospective *authors* too difficult to use, although the results are very helpful to the reader. The difficulty lies not in the manipulation of the software user interface but in the deeper problem of telling a coherent story in an unfamiliar medium. And yet, in a communications age that relies increasingly on visual media, we suspect that the next generation of scientists will find this much less of a difficulty.

The challenges that we mentioned in §10 – the economics of open access, changes in the status of peer review, the trend to publish more articles rather than fewer grand syntheses – will undoubtedly have some impact on journals in the future. But we believe these will be less important, in the long run, than the dwindling in importance of ‘literature’ as the preferred means of communicating science. While crystallographers as *authors* have created the journals of the IUCr for the last 60 years, future generations will increasingly tell the scientific story using the talents of animators, of movie directors – and even of designers of interactive videogames.

References

- Abrahams, S. C. & Matula, R. A. (1988). *Acta Cryst.* **A44**, 401–410.
- Behrens, H. & Luksch, P. (2006). *Acta Cryst.* **B62**, 993–1001.
- Berendsen, H. J. C. (2003). *Acta Cryst.* **D59**, 777–782.
- Bergerhoff, G., Allen, F. H., Bellard, S. A. & Lucas, C. V. (1986). *Acta Cryst.* **C42**, 1671–1675.
- Bowler, M. W., Montgomery, M. G., Leslie, A. G. W. & Walker, J. E. (2006). *Acta Cryst.* **D62**, 991–995.
- Chayen, N. E., Cianci, M., Grossmann, J. G., Habash, J., Helliwell, J. R., Nneji, G. A., Raftery, J., Rizkallah, P. J. & Zagalsky, P. F. (2003). *Acta Cryst.* **D59**, 2072–2082.
- Clews, C. J. B. & Cochran, W. (1948). *Acta Cryst.* **1**, 4–11.
- Creze, C., Rinaldi, B., Haser, R., Bouvet, P. & Gouet, P. (2007). *Acta Cryst.* **D63**, 682–688.
- Cruickshank, D. W. J. (1998). *Acta Cryst.* **A54**, 687–696.
- Garfield, E. (1990a). *Curr. Contents*, Issue 36, pp. 5–14.
- Garfield, E. (1990b). *Curr. Contents*, Issue 37, pp. 3–11.
- Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* **A47**, 655–685.
- Hawkins, D. T. (1980). *Acta Cryst.* **A36**, 475–482.
- Helliwell, J. R., Strickland, P. R. & McMahon, B. (2006). *Inf. Serv. Use*, **26**, 45–55.
- Kamminga, H. (1989). *Acta Cryst.* **A45**, 581–601.
- Kennard, O., Speakman, J. C. & Donnay, J. D. H. (1967). *Acta Cryst.* **22**, 445–449.
- Knuth, D. E. (1986). *The TeX Book*, Vol. A of *Computers and Typesetting*. Reading: Addison-Wesley.
- Lamport, L. (1986). *LaTeX. A Document Preparation System*. Reading: Addison-Wesley.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Lovelace, J. J., Murphy, C. R., Pahl, R., Brister, K. & Borgstahl, G. E. O. (2006). *J. Appl. Cryst.* **39**, 425–432.
- McMahon, B. (1996). *J. Res. Natl Inst. Stand. Technol.* **101**, 347–355.
- Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M. & van de Streek, J. (2006). *J. Appl. Cryst.* **39**, 453–457.
- Murray-Rust, P. & Rzepa, H. S. (2004). *J. Digital Inf.* **5**, Article No. 248.
- Nardelli, M. (1983). *Comput. Chem.* **7**, 95–98.
- Nature (2006). *Overview: Nature's peer review trial*, <http://www.nature.com/nature/peerreview/debate/nature05535.html>.
- O'Reilly, T. (2005). *What is web 2.0: Design patterns and business models for the next generation of software*, <http://www.oreillynet.com/lpt/a/6228>.
- Rzepa, H. S. & Murray-Rust, P. (1998). *J. Chem. Inf. Comput. Sci.* **38**, 976–982.
- Shaw, D. & Moore, H. (1996). *Electronic Publishing in Science. Proceedings of the Joint ICSU Press/UNESCO Expert Conference, February 1996*. Paris: ICSU Press/UNESCO.
- Snell, E. H., Boggon, T. J., Helliwell, J. R., Moskowitz, M. E. & Nadarajah, A. (1997). *Acta Cryst.* **D53**, 747–755.
- Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.
- Strickland, P. R., Hoyland, M. A. & McMahon, B. (2005). *International Tables for Crystallography*, Vol. G. *Definition and Exchange of Crystallographic Data*, edited by S. R. Hall & B. McMahon, pp. 557–569. Dordrecht: Springer.
- Vriend, G. (1990). *J. Mol. Graphics*, **8**, 52–56.
- Wlodawer, A. (2007). *Acta Cryst.* **D63**, 421–423.
- Yang, P., Craig, P., Goodsell, D. & Bourne, P. (2003). *Bioinformatics*, **19**, 897–898.