Computerization of the IUCr Editorial Office, Chester: a Review of 1991

The Technical Editor's Office, International Union of Crystallography, 5 *Abbey Square, Chester CH*1 2*HU, England*

March 4, 1992

Abstract

A review is given of technical innovation and progress in the Editorial Office during 1991. Major developments include the connection of the Chester office to the world-wide data transfer networks *via* the UK JANET system, checking of *all* structural papers, and the introduction of procedures to generate structural papers from the IUCr's standard Crystallographic Information File.

I. Introduction

This report updates the document *Computerization of the IUCr Editorial Office, Chester: a Progress Report*, which was prepared by the Editorial Office and circulated among Co-editors and members of the Executive and Finance Committees in late 1990. Such an update is timely, as the specification of the Crystallographic Information File (CIF) has been published in the November 1991 issue of *Acta Crystallographica*, Section C, together with instructions on submitting machine-readable structural papers to the Editorial Office as CIFs. The major developments during 1991 have been: connection of the Chester Office to national and international data networks; acquisition of further software for checking and development of staff expertise in this field; collaboration with CIF developers in defining the new standard and development of software to format CIFs in publishable form; and continuing acquisition and expansion of computer hardware and software to develop a sound base for future automation. Each of these topics will be reviewed in turn.

II. Wide-area networking

Connection to the UK academic network JANET was established in November 1990, allowing some interaction with UK sites. However, administrative delays with registering our e-mail address prevented us from receiving mail until early 1991. Since then, electronic mail has come to be a very useful medium for exchanging information with people working on the CIF project, for collaboration on *International Tables* volumes, for exchange of checking and other software, and for communicating with Co-editors. There is now also a physical connection to the Executive Secretary's building to allow full e-mail communications facilities to the Executive Secretariat.

The JANET network is a point-to-point X.25 service connecting universities, polytechnics, research institutes and some commercial sites within the UK. Most participating entities are funded from central government, so that the participation of the IUCr as a 'learned society' had to be negotiated as a special case. The IUCr must pay a fixed annual fee for use of JANET services, which renders this initially a very expensive means of communication. However, as volume usage increases, with no attendant increase in charges, the cost per message will fall.

The X.25 protocol is comparatively unusual among national academic networks, so that interaction with other networks must be through gateways; however, the UK is well supplied with mail gateways, and increasingly with file-transfer and interactive gateways, so that UK-based services can be made available to the international community increasingly easily. Some details and implications of this are discussed below.

E-mail

The use of electronic mail is the most important benefit of our JANET connection. Mail from most large international networks is handled reasonably efficiently by gateway machines in the UK, and we have successfully communicated with Bitnet and Internet addresses, together with users in Japan and Australasia. We are presently not able to handle international uucp (Usenet) mail, simply because some additional charges are thereby incurred; but this lacuna should be easily resolvable if there is found to be a real need for this facility. It is our understanding that many sites in the Soviet Union, for example, still rely on this route. Mail users who wish to send electronic mail to the UK must, of course, be aware that domain addresses in the UK are set up in the opposite order to that used elsewhere (that is, the IUCr address is given within JANET as uk.ac.iucr, with most significant entity first; mailers in other countries may need to recast this as iucr.ac.uk if their relay machines rely on the more usual most-significant-entity-last arrangement). It is envisaged that CIF submissions to *Acta* C will come increasingly via e-mail.

File transfer

Files may be transferred direct from machines in the UK and (*via* a relay) on the Internet. This has proved useful in acquiring public-domain software for communications, graphics display and text formatting. It also offers a means of transferring large calculational programs and batches of CIFs (such as to databases). However, unless the data transferred includes non-printing characters or is in machine binary form, electronic mail may also be used in most cases. File transfer increases the security risks of access to the computers offering this service, and so the IUCr is not at present acting as a source of files for transfer. Files which are available for public distribution (the CIF Dictionary, source code for *QUASAR*, *CYCLOPS etc.*) will be made available through a mail server, rather than *via* file transfer, at least for the foreseeable future.

Interactive login

JANET supports interactive login *via* X.29 protocol to UK sites. We regularly use this facility to access the crystallographic databases at Daresbury Laboratories, and visitors have used it to demonstrate programs running on their home computers. It also makes available certain public information facilities within the UK, such as the commercial telephone directory known as 'Electronic Yellow Pages', the National Information on Software and Services (NISS) Bulletin Board [and its related service ASK based in Karlsruhe for the German academic community], and various library catalogues. Visitors have also been able to log in *via* a relay to Internet computers in the United States and Australia. Again, for security reasons, the IUCr does not currently support X.29 access, but it is feasible to provide login facilities for external users in the future, if this is deemed desirable. One current development which may be relevant to such considerations is the pilot project currently in operation on JANET to allow Internet protocols to be carried over the X.25 connections currently in place, thus permitting direct login to Internet sites. It is believed that our current communications hardware and software could be configured to participate in this project at no additional cost. This might, again, be worth considering if in future it is desired to allow login access to IUCr data from sites world wide.

At present there is a single connection to JANET operating at 9600 bps. This has proved reliable, and is expected to be adequate for our needs for some time.

III. Centralized checking

By late 1991, all Coeditors had been invited to send structural papers on receipt to Chester for checking. In 1991 some 780 papers were checked (about 950 structures). From the beginning of 1992, all structural papers for publication in *Acta* C are being sent direct by authors to the Technical Editor's office. The basic procedures involved in the checking were outlined in our last report; here we update the list of programs run.

Computer programs in use at the Chester office

Crystallographic packages for structure solution and refinement

*Xtal*3.0 (Hall & Stewart, 1990). The *Xtal* program system includes the geometry checking program *BONDLA* and the interactive graphics program *PIG*. It has a fully functioning CIF input/output routine, and contains many useful crystallographic tools, including a version of *ORTEP* (Johnson, 1976).

NRCVAX (Gabe, Le Page, Charland, Lee & White, 1989; Le Page, 1982, 1988). This is another fully-featured package, containing the powerful symmetry checking tool *MISSYM*, cell reduction routines, geometry programs and interfaces to the graphics programs *PLUTO* (Motherwell & Clegg, 1978) and *ORTEP*, together with some useful home-grown graphics utilities.

Database evaluation and checking programs

Cambridge Crystallographic Data Centre *UNIMOL* (Allen *et al*., 1974). This extremely powerful package is used for comparing the predicted geometry of a structure with that described by the authors; bond distances are checked automatically and, where discrepancies are found, alternative positional coordinates for the errant atoms can sometimes be deduced.

National Institute of Standards and Technology *NIST*LATTICE* (Mighell, Hubbard, & Stalick, 1981; Himes & Mighell, 1987; Karen & Mighell, 1991). We have been given several of the programs used by NIST for evaluation of structures to identify the lattice type from the Niggli matrix and from converse-transformation analysis (Karen & Mighell, 1991). We are also able to check reduced cell parameters against those stored in the NIST Crystal Data database.

Geometry analysis programs

STRUMO (Brown & Altermatt, 1989). This is an inorganic structure modelling program which assesses atomic positions from a valence sum map. It is currently run when required as an interactive tool; Dr Brown has indicated his willingness to implement modifications which would allow it to run in a batch mode. Although this program is helpful with many inorganic structures, we still sometimes run up against the general problem of efficiently identifying and checking all the bond distances and angles in high-symmetry space groups.

PARST (Nardelli, 1983). The Parma structural package is still heavily used as a comprehensive generator of geometry in structures which are described by a coordinate set comprising a connected asymmetric unit.

PLATON (Spek, 1990). This is a very powerful package which generates exhaustive output describing a crystallographic structure. It incorporates the symmetry checking program *MISSYM* (Le Page, 1988), together with a thermal-ellipsoid plotting routine and other graphical facilities. Our only complaint with this program is that it produces excessive detail for our usual needs, and so it is generally run to investigate in greater detail problems indicated by other programs.

Symmetry checking programs

LEPAGE (Spek, 1988). This is a DOS implementation of the *CREDUC* algorithm of Le Page (1982). It has been used routinely as a swift check of papers in hand that have not been checked, but will be used less often as checking is extended to all new papers.

NEWLAT (Mugnoli, 1985). A program which suggests alternative lattices consistent with some tolerance around the given cell parameters.

DELOS (Burzlaff & Zimmermann, 1985). Evaluates the probable Bravais lattice following the method of Delaunay.

TRACER (Lawton, & Jacobson, 1965). A lattice transformation/cell reduction program.

Graphics programs

PLUTON (Spek, 1991). A versatile graphics program, descendant of *PLUTO* (Motherwell & Clegg, 1978), which can generate various styles of molecular representation on a variety of output devices.

MOLDRAW (Ugliengo, Borzani & Viterbo, 1988). A very nice interactive graphics program for use on a PC. Unfortunately, PostScript hardcopy output is not supported.

MolDraw/MolView (Cense, 1990). Useful molecular graphics programs that run on Apple Mac computers. We are unable to use *MolView*, because of its requirements for a colour monitor.

MOLSCRIPT (Kraulis, 1991). A beautiful macromolecular graphics package that is heavily reliant on PostScript procedures for hidden-line removal, and so produces only PostScript output as hardcopy or as interpreted by workstation-based PostScript previewers.

GX Interactive ORTEP (Mallinson & Muir, 1985). A GKS-based user-friendly front end to *ORTEP* (Johnson, 1976).

IV. The Crystallographic Information File

Throughout 1991 the Chester Office collaborated with the authors of the CIF specification (Hall, Allen & Brown, 1991) and developed software and techniques for handling CIFs. The main contribution in this area has been the development of a prototype CIF→TEX translator, which allows a CIF to be formatted according to typographic rules that may be defined whenever the program is run. This allows different typographic conventions to be used for CIFs according to their different purposes. An early application of this was in the formatting of the CIF Dictionary included within the paper of Hall, Allen & Brown (1991) from the master electronic file **cifdic.C91**. The power of this approach has also been demonstrated in trials involving the *World Directory of Crystallographers*, and this may offer an attractive way of preparing the forthcoming Ninth Edition of that Directory.

An early version of that software was demonstrated at the American Crystallographic Association Meeting in Toledo in July 1991, and this generated a substantial amount of interest. Following this meeting, and after the publication in *Acta* of the CIF description paper and new *Notes for Authors*, several crystallographers contacted the Editorial Office for information and help, and already several authors have submitted trial papers in CIF format.

It may be useful to summarise the ways in which the Editorial Office currently handles CIFs.

A new submission received by e-mail may be piped from within the mail reading program to a procedure that checks the logical integrity of the file. If there are syntactic errors in the CIF, an error listing is printed out, and the file transferred to a holding directory. A member of the Editorial staff will examine the file; if it can easily be repaired, the edited CIF is mailed back into the system (so that all CIFs are treated uniformly within the mail program interface). If the file departs substantially from the syntactical rules of CIF, it can be returned to the author with information on its deficiencies.

If the incoming CIF is free from structural error (or if it has been corrected by Editorial staff), it is assigned a Co-editor number, and automatically a transmittal form, checking form and copy of the included letter of submission (if present as `_publ_contact_letter`) are printed out. Checks are also run automatically on missing mandatory data items and to determine whether any data names appear that are not given in the CIF Dictionary. The file is then operated upon by the program *QUASAR* to extract the data required for publication in the desired order. The output from *QUASAR* is itself in CIF format; this is translated to a TEX file by the program *ciftex*, and the resulting file is operated upon by the typesetting program TEX. A manuscript is produced in 'preprint' form for use by the Co-editor and referees. The CIF is then re-processed in a similar manner to produce the supplementary publication material (all atom coordinates, anisotropic displacement parameters and all geometry) that may be in the file. Finally, a spelling check is run on the manuscript for the benefit of the Editorial staff. The entire process usually takes only a minute or two of real time. A copy of the CIF is then placed in an appropriate directory for checking. In principle this could also run automatically, but recovery from errors at this stage is still not handled gracefully, so at present the checking process is run separately by an Editorial Assistant.

Following the review process, minor amendments may be made to the CIF by the Editorial staff. Major changes will be undertaken by the author, who would send a revised CIF for re-processing in the same manner. A second pass through *ciftex* will now produce a 'proof' of the paper, which should require only minor editing before it is ready for transmission to a phototypesetting agency. The typeset bromide prints will be cut and pasted into pages by the printer, together with artwork, which at present must still be supplied as conventional hard copy. We hope to be able to handle PostScript artwork electronically in due course.

The above description applies to a 'full' CIF, submitted electronically by an author. In the short term, we shall also handle papers submitted on paper in a similar way, entering the structural data required for checking into a CIF, and using that CIF for checking and typesetting of the experimental section and tables.

The CIFs in the checking directory are then processed in the way described in §III of our previous report.

The way in which a CIF is transformed to a publishable paper is not yet fully documented, and is moderately complex. In order to demonstrate to authors how their CIFs will be treated (and, consequently, how they can best compose their CIFs to ensure smooth processing), a derivative of the *ciftex* program, called *cifms*, has been developed. This will give a 'preview' of the paper in ASCII-only format, using simple codes to indicate subscripts and special characters. Its advantage is that it does not require an author to have access to the specific typesetting program (TEX) that we use.

Several prospective authors have indicated that they must submit preprints of work for publication to their employer or other supervisory body. Use of *cifms* would allow them to generate such a preprint for this purpose, and it is therefore planned to make *cifms* generally available for these reasons. The current version is, however, Unix-specific. It is hoped that in time versions can be produced for use on IBM PCs and Apple Mac computers, and possibly also on VAXes.

V. Equipment

An additional graphics workstation was purchased in 1991. This is a Sun SPARCstation 2, a higher-performance version of the machines already in use. Because of its speed and memory, it has already become established as the workhorse of the checking and CIF handling procedures.

Peripheral equipment attached to this machine includes a CD-ROM drive and a SPARCprinter. These supply networked services for the entire Sun cluster. The printer is driven by PostScript-compatible software

within the workstation. However, it has become apparent that this system is not fully PostScript compatible, and so it is still possible to transmit problem files for printing to the PostScript printer attached to the 3B2 computer.

An optical disk drive with read/write capability has also been acquired for attachment to the Sun fileserver. It is intended that this will act as a long-term storage medium for archiving CIFs. The optical disks have a large capacity (after formatting about 260 Mbyte is available for data storage), they have better long-term stability than magnetic tapes, and they are accessible as mountable filesystems (that is, data may be read from or written to disk at any location, unlike the linear storage format of a magnetic tape).

The other workstations have been given a memory upgrade to 16 Mbyte (from 12) to improve performance and allow the simultaneous running of more processes.

Ethernet cards have been installed in the office Macs and PCs to allow easier data transfer between these machines and the Sun cluster.

All the equipment continues to function reliably and to give satisfactory performance.

References

Allen, F.H., Kennard, O., Motherwell, W.D.S., Town, W.G., Watson, D.G., Scott, T.J. & Larson, A.C. (1974). *J. Appl. Cryst.* **7**, 73–78.

Brown, I.D. & Altermatt, U.D. (1989). *J. Chem. Inf. Comput. Sci.* **29**, 266

Burzlaff, H. & Zimmermann, H. (1985). *Z. Kristallogr.* **170**, 241–246.

Cense, J.M. (1990). *Acta Cryst.* A**46**, C-69.

Gabe, E.J., Le Page, Y., Charland, J.-P., Lee, F.L. & White, P.S. (1989). *J. Appl. Cryst.* **22**, 384–387.

Hall, S.R. & Stewart, J.M. (1990). Eds. *Xtal*3.0 *Reference Manual*. Univs. of Western Australia and Maryland.

Himes, V.L. & Mighell, A.D. (1987). *Acta Cryst.* A**43**, 375–384.

Johnson, C. K. (1971). *ORTEP*II. Report ORNL-3794, revised. Oak Ridge National Laboratory, Tennessee, USA.

Karen, V.L. & Mighell, A.D. (1991). *NIST*LATTICE: A Program to Analyze Lattice Relationships.*

Karen, V.L. & Mighell, A.D. (1991). *J. Appl. Cryst.* **24**, 1076-1078.

Kraulis, P.J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lawton, S.L. & Jacobson, R.A. (1965). *The Reduced Cell and its Crystallographic Applications.* USAEC RD Report IS-1141.

Le Page, Y. (1982). *J. Appl. Cryst.* **15**, 255–259.

Le Page, Y. (1988). *J. Appl. Cryst.* **21**, 983–984.

Mallinson, P.R. & Muir, K.W.(1985). *J. Appl. Cryst.* **18**, 51.

Mighell, A.D., Hubbard, C.R. & Stalick, J.K. (1981). *NBS*AIDS80: A FORTRAN Program for Crystallographic Data Evaluation.* NBS Tech. Note 1141.

Motherwell, W. D. S. & Clegg, W. (1978). *PLUTO.* Program for plotting molecular and crystal structures. Univ. of Cambridge, England.

Mugnoli, A. (1985). *J. Appl. Cryst.* **18**, 183–184

Nardelli, M. (1983). *Comput. Chem.* **7**, 95–98

Spek, A.L. (1988). *J. Appl. Cryst.* **21**, 578–579.

Spek, A.L. (1990). *Acta Cryst.* A**46**, C-34.

Spek, A.L. (1991). *PLUTON.* Program for the display and analysis of crystal and molecular structures. Univ. of Utrecht, The Netherlands.

Ugliengo, P., Borzani, Y. & Viterbo, D. (1988). *J. Appl. Cryst.* **21**, 75.