ing. However, in a comprehensive collection of structural data sets, such as in a public structural database, it might be possible to identify particular data items that could be used for weighting individual data sets when the database is being 'mined' for particular patterns or characteristic values. It will be interesting to see whether a consensus emerges on what items would be suitable. It is clear that reliance on a single indicator will not be appropriate for sophisticated studies. The old idea that a structure could be classed as 'good' or 'bad' on the basis of its final residual *R* factor alone has long been abandoned, but it may be possible to stipulate criteria for a set of interrelated data items and use these to filter specific information from a database.

5.7.2.7. Submission and review

When an author has previewed and checked the contents of the CIF and has made the changes suggested by a careful study of the preprint and the *checkcif* report, the article may finally be submitted to *Acta Cryst. C* or *E* by file upload over the web. Other files completing or supporting the submission are also transferred to the editorial office at this time. These include structure-factor or powder profile listings for each structure, figures and chemical diagrams, and sometimes other supplementary documents. Structure-factor listings are supplied in CIF format. Figures may be in one of a number of standard graphics file formats, and at the moment have to be uploaded as separate files. Future extensions to CIF, perhaps following the imgCIF approach, may allow all the items needed to submit an article, including figures, to be prepared as a single file.

When all the files have arrived at the editorial office, a review document is generated that can be sent to the referees. This document contains: the text and tables of the article that will appear in the final publication, but laid out in a more open style suitable for annotation by hand; tables of atomic positions and geometry (containing all the data in the CIF, not just the subset that has been selected for displaying in the published article); certain fields from the CIF that are not normally printed but which may contain details of the way in which the experiment was carried out (these fields might have been completed manually or by the software controlling the experiment); the figures and other supplementary documents; and a print-out of the report from a final *checkcif* cycle, including a displacement-ellipsoid plot of the molecule in a minimal-overlap least-squares plane view. This composite document provides the information that a referee will typically want to consider in a compact and convenient form. Because the CIF is so highly structured, producing this review document is in most cases entirely automatic. The complete CIF as submitted by the author and the experimental data are also made available to the reviewer.

If revisions are requested, authors may upload modified files. The generation of revised versions of an article is also largely automatic.

5.7.2.8. Publication

When the final version of a CIF for *Acta Cryst. C* or *E* is approved, the article is ready for publication. Once more, the data fields required for the published article are extracted from the CIF and sorted. If the author has asked for additional items to be printed by using `_publ_manuscript_incl_extra_item`, these also are extracted. The result is transformed to a file suitable for processing by typesetting software. For *Acta Cryst. C* this was originally a TEX file; now a further transformation generates an SGML file that conforms to the document type definition (DTD) common to all IUCr journals. This allows not only typesetting and printing, but also the generation of the HTML for the navigable online version of the article, and the extraction of metadata for building online tables of contents and for supplying to bibliographic databases.

The conventional published article then appears in a monthly issue. Each article is still similar in style to the type of structure report published in journals for decades, although tables of atomic positions and geometric data are not usually displayed now, since these data are so readily available from the online article.

The online version of the journal, however, presents a much more information-rich version of the article. Each article is generally available in the form of a PDF file, suitable for downloading and offline printing. There is also an HTML version of the same text, and this version has rich internal links that make it easy to scroll back and forth through the article, jump to specific sections and see figures in low-resolution thumbnail or high-resolution views. The reference list contains links to the articles that are cited. There may also be links to related records in chemical or crystal structure databases. The reader may also download the experimental data and any supplementary documents associated with the article. As mentioned above, for *Acta Cryst. E* a summary of the check report is also available.

Finally, the structural data may be downloaded directly in CIF format. The CIF is presented in two ways. If a reader follows one link in a web browser, the file is interpreted simply as a text file and appears as a simple listing in the browser window, from which it may be printed or saved to disk. However, if the reader follows the other link, the CIF is transmitted to the browser with a header declaring its MIME type (Freed & Borenstein, 1996) as 'chemical/x-cif'. This is one of several MIME types registered for particular presentations of chemistry-related content by Rzepa & Murray-Rust (1998). The reader may then configure a web browser to respond in a specific way to content tagged with this MIME type; typically a helper application such as a molecular visualizer [*e.g. Mercury* (Bruno *et al.*, 2002)] will be launched that allows three-dimensional visualization and manipulation of the molecular or crystal structure.

When an article has been published in *Acta Cryst. C* or *E*, the CIF is transferred to the relevant public structural databases. Thus, the transcription errors that used to cause so many problems for data harvesters are completely avoided and one of the initial goals of the CIF project is achieved: uncorrupted data transfer from diffractometer, through publication, to a final repository.

Because *Acta Cryst. C* and *E* handle almost exclusively the publication of structure reports, the editorial workflow based on CIF lends itself to a very high level of automation and the journals are produced efficiently and on short timescales. Routine refereeing of structures is made very easy by the provision of checking reports, and the universal use of e-mail and web file transfer means that production times can be very fast.

5.7.3. CIF and other journals

Not every journal will be able to benefit to the same extent from the handling of CIFs. For many journals, structure reports will be secondary to the main purpose of most articles, and CIF data will more usually be deposited as supplementary or supporting documents, while only a summary (if anything) of the structure will be reported in the article body.

Nevertheless, the ability to extract data from CIFs automatically and the ability of much crystallographic software to read CIFs mean that even journals that do not specialize in crystallography can provide a production stream that includes careful checking of crystal structure data. The IUCr continues to develop *checkcif* as

a service which can be used by other publishers to enhance their checking of crystal structures, and there is considerable interest in this approach.

All journals publishing the results of crystal structure determinations may easily collect the supporting data in CIF format and transfer the files to public databases, improving the accuracy and efficiency of the database-building procedures.

5.7.3.1. Including CIF data in an article

For journals other than those specializing in full-scale structure reports, including CIF data in tables or reports of structures within general articles is rather more problematic. The translation of CIF data into XML seems to be a promising route to explore, as journals and reference volumes are increasingly being typeset from XML files. Traditionally, publishing has emphasized content markup that leads to a particular typographic representation. Modern trends are towards markup that tags the content by purpose, with the representation directed by external 'style files'. Consider Fig. 5.7.3.1, which shows the typeset representation of a set of data items in a CIF for a structural paper.

First, it can be seen that several CIF data items are omitted from the printed representation, such as the *International Tables* space-group number and the Hall symbol for the space group. For compactness, the printed data value does not have a legend or annotation if the meaning of an item is clear from the context; thus, the crystal system and Hermann–Mauguin space-group symbol are printed without any accompanying text. The journal may also omit information that is implicit given other data; thus the cell angles are not printed for an orthorhombic cell. On the other hand, units, which are implicit in the definition of a CIF data item, are printed. Related items are grouped together in a single expression, as in the case of the θ range or the crystal dimensions. In some cases, numerical values have been rounded to meet the journal's policy.

All of these transformations are matters of style, but it can be seen that they are not always trivial mappings to single data names. The style files determining the transformation from a detailed explicit data tabulation in the initial CIF may need to implement complex logical tests to suit the requirements of the journal.

Fig. 5.7.3.2 shows the same extract in TEX, the markup and typesetting language that was used for several years to produce *Acta Cryst. C*. It can be seen from this extract that the actual markup maps very closely to the initial CIF. All the cell parameters, including the cell angles, are present in the source file. The expansion of the macros (*e.g.* `\cellalpha`) executes the logic required to determine whether the value is to be printed and generates the additional text surrounding the value. Each data name is mapped to a distinct macro (even if the macros themselves have identical or near-identical internal structure), which preserves the semantic labelling of the original CIF. These macros are maintained in a separate file referenced and executed by every invocation of the typesetting program.

In contrast, Fig. 5.7.3.3 shows part of the SGML now used to typeset *Acta Cryst. C* and to generate HTML versions of the articles online. It is immediately seen that the markup emphasizes typographic style and positioning, and there is no explicit labelling by semantic element. Additional labelling is now found in the document structure; the individual items are marked up as 'list items' (`<li>`), but the arrangement of this list into a tabular form is a feature of the typesetting engine, not the SGML.

It is clear that the TEX macros provide a representation of the contents of the CIF that could easily be converted back to the initial input CIF. At present, such bidirectional translation is not possible from the SGML file.

Clearly, therefore, a mapping to SGML that preserved semantic markup would be preferable. It is most likely that suitable bidirectional translations would be based on XML.

```
data_I
_symmetry_cell_setting            orthorhombic
_symmetry_space_group_name_H-M    'P 21 21 21'
_symmetry_space_group_name_Hall   'P 2ac 2ab'
_symmetry_Int_Tables_number       19

loop_
 _symmetry_equiv_pos_as_xyz
 'x, y, z'
 '-x+1/2, -y, z+1/2'
 '-x, y+1/2, -z+1/2'
 'x+1/2, -y+1/2, -z'

_chemical_formula_moiety           'C10 H8 Br N S'
_diffrn_radiation_type              MoK\a
_exptl_absorpt_coefficient_mu       4.280

_cell_length_a                    5.7339(7)
_cell_length_b                    14.8229(15)
_cell_length_c                    23.469(2)
_cell_angle_alpha                 90.0
_cell_angle_beta                  90.0
_cell_angle_gamma                 90.0
_cell_volume                      1994.7(4)
_cell_formula_units_Z             8
_cell_measurement_temperature     296(2)
_cell_measurement_reflns_used     58
_cell_measurement_theta_min       4.806
_cell_measurement_theta_max       11.635

_exptl_crystal_description        plate
_exptl_crystal_colour             colourless
_exptl_crystal_size_max           0.40
_exptl_crystal_size_mid           0.32
_exptl_crystal_size_min           0.04
_exptl_crystal_density_meas       ?
_exptl_crystal_density_diffrn     1.693
_exptl_crystal_density_method     'not measured'
_exptl_crystal_F_000              1008
```

(*a*)

Crystal data

$C_{10}H_8BrNS$	Mo $K\alpha$ radiation
$M_r = 254.14$	Cell parameters from 58 reflections
Orthorhombic, $P2_12_12_1$	$\theta = 4.8$–11.6°
$a = 5.7339\ (7)$ Å	$\mu = 4.28$ mm^{-1}
$b = 14.8229\ (15)$ Å	$T = 296\ (2)$ K
$c = 23.469\ (2)$ Å	Plate, colourless
$V = 1994.7\ (4)$ Å^3	$0.40 \times 0.32 \times 0.04$ mm
$Z = 8$	
$D_x = 1.693$ Mg m^{-3}	

(*b*)

Fig. 5.7.3.1. Typesetting of structural data. The contents of the CIF (*a*) are transformed into a typeset representation (*b*) that omits, annotates or reorders the incoming data according to context and the style rules of the journal.

5.7.3.2. CIF and XML

XML is a specific concrete implementation of SGML suitable for generation of online browsable content. Mature style transformation mechanisms for XML exist and others are under active development.

```
\structno
\noindent{\nineit Crystal data}\nobreak\par
\vskip2pt\begindoublecolumns\twocoltrue\defaultfont
\raggedright
\everypar={\global\parindent=0pt\hangindent=1em
          \hangafter=1 }\noindent

\chemformiupac{C$_{10}$H$_{8}$BrNS}
\chemform{C$_{10}$H$_{8}$BrNS}
\chemformsum{C$_{10}$H$_{8}$BrNS}
\molwt{254.14}
\system{Orthorhombic}\def\sgsetno{2}
\defaultfont
\sgHM{$P2{}_{1}2{}_{1}2{}_{1}$}
\cella{5.7339 (7)}
\cellb{14.8229 (15)}
\cellc{23.469 (2)}
\cellalpha{90.00}
\cellbeta{90.00}
\cellgamma{90.00}
\cellvol{1994.7 (4)}
\cellz{8}
\dx{1.693}
\dm{missing}
\densmetha{not measured}
\radiationtype{Mo {\it K}$\alpha$}
\wavelength{0.71073}
\cellrefl{58}
\cellthetamin{4.8}
\cellthetamax{11.6}
\absorpmu{4.28}
\celltemp{296 (2)}
\shape{Plate}
\colour{Colourless}
\sizemax{0.40}
\sizemid{0.32}
\sizemin{0.04}
\sizerad{missing}
\origin{see text}
```

Fig. 5.7.3.2. Part of a TEX file used to print the article shown in Fig. 5.7.3.1(*b*).

```
<sec id="sec2.1">
 <sec id="sec2.1.1">
  <st><?print tpct=0pt>Crystal data</st>
  <p>
   <l id="l1" type="unord">
    <li><p>C<inf arrange="stagger">10</inf>H<inf
        arrange="stagger">8</inf>BrNS</p></li>
    <li><p><it>M</it><inf
        arrange="stagger"><it>r</it></inf> =
        254.14</p></li>
    <li><p>Orthorhombic, 
     <fi type="tex" print-info="tth" img.type="tth"
        img.data="teximages/ga1014fi3.tth">
        P2_{1}2_{1}2_{1}</fi></p></li>
    <li><p><it>a</it> =
        5.7339 (7) &Aring;</p></li>
    <li><p><it>b</it> =
        14.8229 (15) &Aring;</p></li>
    <li><p><it>c</it> =
        23.469 (2) &Aring;</p></li>
    <li><p><it>V</it> =
        1994.7 (4) &Aring;<sup
        arrange="stagger">3</sup></p></li>
    <li><p><it>Z</it> =
        8</p></li>
    <li><p><it>D</it><inf
        arrange="stagger"><it>x</it></inf> =
        1.693 Mg m<sup
        arrange="stagger">&minus;3</sup></p></li>
    <li><p>Mo <it>K</it>&alpha; radiation</p></li>
    <li><p>Cell parameters from 58 <?print show
        &softreturn;>reflections</p></li>
    <li><p>&thetas; = 4.8&ndash;11.6&deg;</p></li>
    <li><p>&mu; = 4.28 mm<sup
        arrange="stagger">&minus;1</sup></p></li>
    <li><p><it>T</it> =
         296 (2) K</p></li>
    <li><p>Plate, colourless</p></li>
    <li><p>0.40 &times; 0.32 &times;
         0.04 mm</p></li>
   </l>
  </p>
</sec>
```

Fig. 5.7.3.3. Part of the SGML file used to print the article shown in Fig. 5.7.3.1(*b*).

Section 5.3.8.2.1 describes one transformation to XML in the biological structures field, designed primarily for database interchange rather than publication. This transformation preserves the underlying data model of an mmCIF very closely, and one might anticipate similar XML transformations for small-molecule CIF applications and for publications. It is even possible that the XML transformations referred to in Chapter 5.3 could be used for publishing articles if suitable style transformations are developed, but this has not been tested yet.

One difficulty with a simple CIF-to-XML transformation is that it could be easily adapted to the publication of structure reports in dedicated journals, but would not necessarily be compatible with other XML implementations developed by an unspecialized publishing house. This could be avoided by the registration of an XML name space covering transformed CIF data and the production of portable stylesheet transformations that could be adopted and modified to meet the requirements of different publishing houses. As yet, we know of no initiatives in this direction.

XML name spaces have been registered to safeguard the development of subject-specific methods of representation as part of a project by the International Union of Pure and Applied Chemistry (Becker, 2001). One markup language that falls within the scope of this project is Chemical Markup Language (CML) (Murray-Rust & Rzepa, 1999, 2001).

Further discussions of the relationship between CIF and XML representations and a proposal for extensions to certain CIF data values to accommodate the wider range of data structures permitted in XML are given by Bernstein (2000).

We acknowledge the guidance, enthusiasm and dedication of past and present members of the editorial boards of *Acta Crystallographica Sections C* and *E* in developing the journals along the path described in this chapter. Particular tribute must be paid to Syd Hall, George Ferguson, Bill Clegg, David Watson and Tony Linden. We are very grateful to Ton Spek for his close involvement with the development of checking software, and also wish to acknowledge George Sheldrick, Mario Nardelli, Eric Gabe, Peter White, Yvon Le Page, Alan Mighell, Vicky Karen, Doug du Boulay, Mike Dacombe and Charlie Bugg for their help in the early days of automated structure checking. We wish also to pay tribute to the dedication and effort of our colleagues in the IUCr editorial office: Gillian Holmes, Sean Conway, Amanda Berry, Sarah Froggatt and Lisa Stephenson; and we thank the many authors who have been willing to test new approaches through the years.

Appendix 5.7.1
Request list for *Acta Crystallographica Section C*

Table A5.7.1.1 contains the request list for *Acta Crystallographica Section C* as given in the 2005 *Notes for Authors*. This list is appropriate for a single-crystal X-ray diffraction study and gives all the data items that are displayed in an article *if they are present in the CIF*. In principle, a smaller set of mandatory data items could be supplied as a separate request list. However, certain items may be