other information analysis centers, we are confident that such benefits are attainable. Anyone who is genuinely concerned with, or involved in, the search for solutions in this area would be hard put to deny that such benefits are desirable.

It would seem that all that remains is for those of us who believe so firmly in the value of this research tool to convince the potential users that an effective world food information analysis center would be an investment well worth making. Thus far, our efforts in this regard have not borne fruit. However, we have not given up and we do not intend to. At some point in time—hopefully soon—sufficient support for such a center will develop on the part of industry government or both to make the dream a reality. When this happens, a lot of us who are trying to do a better job of agricultural and food research will have a powerful new resource to draw upon.

CRYSTAL DATA Editor—Automatic Proofreader for the 1968 Edition

LEONARD B. TAYLOR and GERALD G. JOHNSON, JR.
Materials Research Laboratory & Department of Computer Science,
The Pennsylvania State University, University Park, Pa. 16802

Received March 18, 1968

A computer program in FORTRAN IV on the IBM 360/67 has been developed to proofread the third edition of the crystallographic reference *CRYSTAL DATA*. Proofreading is accomplished in four phases: data preparation, record searching routine, numerical verification preparation, and numeric validation. The merging of the character-processing ability of computers and crystallographic theory yields increased reliability and accuracy as well as a decreased time delay for future editions.

The third edition of *CRYSTAL DATA*, a crystallographic reference publication, is scheduled to be published this year, under the joint sponsorship of the American Crystallographic Association and the National Standard Reference Data System of the National Bureau of Standards.[1] This edition is of particular interest to crystallographers, physicists, and chemists since it will contain over 30,000 entries with information current to January 1968. Previous editions were not as large nor as current as the third edition, owing to the time required for proofreading and constructing indices. It is now possible to reduce the time previously needed to publish large current and accurate editions by exploiting the character-processing ability of electronic computers.

Prior to this edition, proofreading of the text was done manually, consuming many man-hours; the level of accuracy using this method was often questionable. A computer program to perform this proofreading of the more than 30,000 entries of *CRYSTAL DATA* has been developed by the authors at The Pennsylvania State University. This program will enable proofreading of all input records to the third and subsequent editions, utilizing a standard level of accuracy which is compatible with crystallographic theory.

The program currently processes one record per second on an IBM 360/67 at The Pennsylvania State University. The source program was written in FORTRAN IV, LEVEL G, RELEASE 12.

DATA PREPARATION

The input data is subjected to four phases of verification prior to final acceptance of the entry. Each of these phases—translation, record searching routine, numeric verification preparation, and numeric validation—concentrates on specific sections of each record to verify the existence and accuracy of items within each entry. Figure 1 shows a sample page of *CRYSTAL DATA*, unjustified for publication. Each entry begins with the characters *** and may be of any reasonable length. The input to the program, *CRYSTAL DATA* Editor, is in the form of octal tapes prepared at the Government Printing Office by the National Bureau of Standards. To produce a compatible code for the IBM 360, these tapes are subjected to a translation routine, phase one of the program. This phase is simply a character-for-character translator from octal code to hexadecimal code, the internal notation utilized for the IBM 360/system. Figures 2 and 3 show the output of the translation section. Figure 2 was printed with the 64-character upper-case print chain, and Figure 3 utilized the 120-character upper and lower case print chain. Either form of input is acceptable, and

***A,ED,NM,OR,000109
Ammonium nickel cyanide (2·1·4) 3-hydrate
&&&.5828 .4801 8.80 15.10 7.25 ... 4 qual. 1.706, 21° 1.741
92°37′ 95°23′ 88°59′
Diammonium nickel cyanide trihydrate, $(NH_4)_2Ni(CN)_4 \cdot 3H_2O$ (Brasseur & de Rassenfosse, Mem. Soc. Roy. Sci. Liege, 4, 397 1941:15.10, 8.80, 7.25; 95°23′, 92°37′, 88°58.5′; $(a:b:c)_{morph}$ =1.686 : 1 : 0.824)0$\bar{1}$0/$\bar{1}$00/00$\bar{1}$. Yellow orange. Twinning. Opt. neg. (546 mμ) =1.473, 1.597; 2V 25°. Pseudo-hexagonal.
***A,ED,NM,OR,000110
Rubidium nickel cyanide (2·1·4) 3-hydrate
&&&.5838 .4824 8.99 15.40 7.43 ... 4 qual. 2.455, 20° 2.480
92°22′ 95°29 89°37′
Dirubidium nickel cyanide trihydrate, $Rb_2Ni(CN)_4 \cdot 3H_2O$ (Brasseur & de Rassenfosse, Mem. Soc. Roy. Sci. Liege, 4, 397, 1941: 15.40, 8.99, 7.43; 95°29′, 92°22′, 89°36.5′, $(a:b:c)_{morph}$ =1.705 : 1 : 0.8345)0$\bar{1}$0/$\bar{1}$00/00$\bar{1}$. Twinning. Opt. neg. Pseudo-hexagonal.
***A,ED,NM,OR,000111

Figure 1. Unjustified page of *CRYSTAL DATA*

```
***C,ED,NM,IN,020860
                  1
SAMARIUM OXIDE (2@3)

&&&10.932 9 @IA3. @16. QUAL. MN2O3 . .
          51  5  1 51
   SAMARIUM SESQUIOXIDE, SM2O3 (TEMPLETON & DAUBEN, J. AM. CHEM. SOC. 76, 5237, 1954). POWDER
7                    1
DATA ONLY. PREVIOUS WORKERS QUOTE A 10.922 AND A 10.915, BUT THE AUTHORS SUGGEST THAT BECAUSE OF THE

INCREASED PURITY OF THE SAMPLE THE PRESENT VALUE IS MORE RELIABLE. ALSO REPORTED: 10.928 1 (BRAUER &

GRADINGER, Z. ANORG. ALLGEM. CHEM. 276, 209, 1954). CP. M63.9023.
                                                        21
```

Figure 2. Upper case printing of output from translation of National Bureau of Standards tape

```
***C,ED,NM,IN,020862

Americium oxide (2•3)

&&&11.03±1 Ia3 16 quant. Mn2O3

  Americium sesquioxide (cubic form), Am2O3 (Templeton & Dauben, J. Am. Chem. Soc. 75, 4560, 1953).
7                                      1
Reddish brown. Powder data only, indexed. Structure is based on that of bixbyite with which it is

isostructural. Confirmed by Eyring et al (J. Am. Chem. Soc. 74, 1186, 1952).
```

Figure 3. Upper and lower case printing of output from translation of National Bureau of Standards tape

the option is exercised by altering the translation table to include or exclude lower case letters and subscripts.

RECORD SEARCHING ROUTINE

This translated data is then processed by the record searching routine whose primary function is to break the free-formated input record into a maximum of 34 items, each addressable in core storage. Editing of the record is initiated in this phase. The following tests are performed by this section:

1. Verify the existence of a crystal system designation letter, editing instructions, substance type, chemical type, and reference number.
2. Verify the existence of up to four chemical names.
3. Extract the numerical data of the entry according to each crystal's format.
4. Verify changes to bold-face type.
5. Extract chemical formula to be used to compute the crystal's density.
6. Verify existence of a reference section, and extract the year of publication within the reference section. If the entry is being published for the first time, then extract numerical data from the reference section also.
7. Extract a transformation matrix.
8. Verify existence of any miscellaneous information.

After each record is separated into its components, the record searching routine passes half of the information to the next phase. The remaining sections are accumulated on external storage until completion of the program.

NUMERICAL VERIFICATION PREPARATION

The information from each entry passed to the numerical verification preparation phases will be subjected to the following test:

1. Search for and eliminate multiple decimal points and invalid characters in numeric fields.
2. Determine the number of significant digits of four numeric entries.
3. Verify crystal's space group with input table.
4. Verify correct format of transformation matrix.
5. Prepare an array with numeric data for punching output cards, for 3 crystal systems. These will be entered into a cell-reduction program at the University of Maryland.

This phase is primarily concerned with checking the isolated entry items extracted by the record searching routine. It checks each item for correct content and deletes invalid characters which may have been entered.

NUMERIC VALIDATION

The final phase of the program is concerned with testing the numerical information of the entry based on crystallographic theory. If all necessary numeric information is available, the following tests are made:

1. Test that all given axial ratios are within specified limitations.[2]
2. Compute independent axial ratios and compare these to the input ratios to the proper significance.
3. Test the difference between theoretically calculated density and experimentally measured density input figures.
4. Compute an independent density figure from the input formula and compare this with the input density figure.
5. Test that given angles are within specified limitations.

Item 4 of the Numerical Validation section, computing an independent density figure from the input formula, is of particular interest. This routine will calculate an atomic weight from the input formula exactly, if upper and lower case input is used; or approximately, if all upper case data is used. The routine will correctly evaluate formulas which contain: hydrate radicals, three nested

levels of parentheses, decimal subscripts, and entries of the form (E_1, E_2, E_3), where E_i represents any element. The formulas must be no longer than 100 characters.

Approximate calculations result from upper case input data, since there is no way to determine which letters are not capitals. Thus, a number of atomic weight approximations are generated, depending on the number of non-unique letter pairs in the formula.

The four phases of editing and validation are summarized in Figure 4.

OUTPUT

At the completion of processing all input entries, an accumulation of errors from each phase is printed. Sample output pages are shown in Figures 5 through 9. Figure 5 shows a partial table of space groups used to check the input space group, listed by crystal systems. Figure 6 shows the cross-reference listing of detectable errors, by control sections of the program, with a brief explanation of the probable type of error. The numbers in the KEY column are printed in the error lists in the following figures. Figure 7 is a listing of the errors from the record searching routine. The system's letter, reference number, and axial ratios are listed for reference purposes. The number under the KEY column is used to enter the table in Figure 6 to determine the probable error. Figure 8 is a listing of errors from the numerical verification and preparation routine. The system's letter, reference number, and axial ratios are printed, for reference purposes, along with the crystal's space group, which was not found in the table in Figure 5. Figure 9 is a listing of errors from the numerical validation section of *CRYSTAL DATA* Editor. Again for reference purposes, the crystal system's letter, reference number, and axial ratios are printed. The volume column is the program's volume calculation for the record, used in the density calculation. The DX column is used to print the input experimental density figure if the program's calculation did not fall within the specified limits.[2] The CALDX column lists the program's density calculation. A comparison between the two columns, DX and CALDX, can be made to determine which is correct. The EST.DX

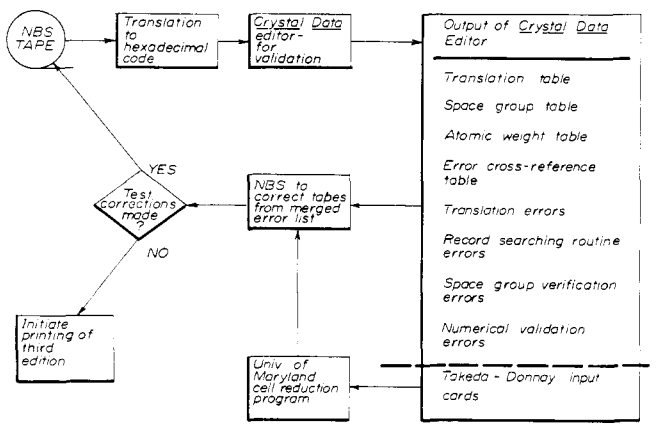

Figure 4. Summarization of the four phases of the *CRYSTAL DATA* editor

```
KEY        ERROR
MAIN PROGRAM:
  1        SYSTEMS LETTER NOT FOUND - GO TO NEXT RECORD
  2        GRID CHANGE TO 7 BEFORE NAMALT - NOT FOUND
  4        REFERENCE EXCEEDS 200 CHARACTERS - GO TO NEXT RECORD
  6        THREE AMPERSANDS NOT FOUND - GO TO NEXT RECORD
  8        MISSING MATRIX (A, M, O)
  9        MATRIX DOESN'T END WITH A PERIOD
 10        YEAR NOT FOUND
 11        MORE THAN 75 BLANKS IN REFERENCE FIELD
 12        DOUBLE BLANK IN TABULAR FIELD
MONOCLINIC ERRORS:
 21        GRID 5 NOT FOLLOWED BY BLANK, @, OR PERIOD
 23        DASH OR CENT SIGN NOT FOLLOWED BY BLANK
 25        ERROR IN TABULAR FIELD
ANORTHIC ERRORS:
 31        GRID 5 NOT FOLLOWED BY BLANK, @, OR PERIOD
 33        DASH OR CENT SIGN NOT FOLLOWED BY BLANK
 35        ERROR IN TABULAR FIELD
ORTHORHOMBIC ERRORS:
 41        GRID 5 NOT FOLLOWED BY BLANK, @, OR PERIOD
 43        DASH OR CENT SIGN NOT FOLLOWED BY BLANK
 45        ERROR IN TABULAR FIELD
HEXAGONAL - TETRAGONAL ERRORS:
 51        GRID 5 NOT FOLLOWED BY BLANK, @, OR PERIOD
 53        DASH OR CENT SIGN NOT FOLLOWED BY BLANK
 55        ERROR IN TABULAR FIELD
CUBIC ERRORS:
 60        GRID 5 NOT FOLLOWED BY BLANK, @, OR PERIOD
 62        DASH OR CENT SIGN NOT FOLLOWED BY BLANK
 69        ERROR IN TABULAR FIELD
ACSUA ERRORS:
 91        GRID 5 NOT FOLLOWED BY @, BLANK, OR PERIOD
 92        DASH OR CENT SIGN NOT FOLLOWED BY BLANK, ( , OR )
VALION ERRORS:
200        A/B GT 1.2 OR C/B GT A/B   (ANORTHIC)
201        C/B GT A/B     (MONOCLINIC)
202        BETA GT 120 DEGREES OR BETA LT 90 DEGREES  (MONOCLINIC)
203        ALPHA LT 90 DEGREES OR BETA LT 90 DEGREES (ANORTHI
204        ABS(DM-DX) GT 0.1*DX
205        ZFO=0 OR VOLUME=0.0
207        CALA/B DIFFERS FROM GIVEN (A,M,O)
208        CALC/B DIFFERS FROM GIVEN  (A,M,O)
210        ATOMIC WEIGHT=0.0,CANNOT TRANSLATE FORMULA
211        CALDX NOT WITHIN LIMITS
212        DX NOT GIVEN, CALDX OR PROGRAM LISTED
213        MULTIPLE VALUES OF DX--EST. OR CALCULATED
T-O ERRORS:
300        DECIMAL POINT INSERTED FOR DM,DX,AB,CB,OR CA
301        DECIMAL POINT INSERTED FOR A, B, OR C
302        NON-STANDARD SPACE GROUP - ANORTHIC
303        DECIMAL POINT IN ZFO DELETED
304        CELL REDUCTION NEEDED - ORTHORHOMBIC,MONOCLINIC
305        NON-STANDARD SPACE GROUP-CUBIC,HEXAGONAL, OR TETRAGONAL
```

Figure 6. *CRYSTAL DATA* editor error--cross reference table used to determine probable errors in record

TABLE OF STANDARD SPACE GROUPS FOR EACH CRYSTAL SYSTEM

MONOCLINIC SYSTEM									
B2	BM	B2/M	BB	B2/B					
PB	P2/B	P21	P21/M	P21/B	P2	PM	P2/M		
ORTHORHOMBIC SYSTEM									
AMM2	AMA2	ABM2	ABA2						
C222	CMM2	CMMM	C2221	CMC21	CMCM	CMMA	CMCA	CCC2	CCCA	CCCM
CUBIC SYSTEM									
F23	FM3	FD3	F432	F4132	F43M	F43C	FM3M	FM3C	FD3M	FD3C
I23	I213	IM3	IA3	I432	I4132	I43M	I43D	IM3M	IA3D	
P23	P213	PM3	PN3	PA3	P432	P4232	P4332	P4132	P43M	P43N
PM3M	PN3N	PM3N	PN3M							

Figure 5. Partial table of 230 standard space groups—minus the triclinic set

ERRORS FROM TRANSLATION AND REFIELD

SYSTEM	REFNO	A/B	C/B	C/A	A	B	C	NAME	KEY
M	002934	1.3927	0.694		11.42	8.20	5.69	C2H12N6NIS2	10
M	002146	1.6043	0.884		13.30	8.29	7.33	C2H6N4PTS	8

Figure 7. Error listing from the record searching routine

THE FOLLOWING ERRORS ARE FROM SPACE-GROUP TRANSFORMATIONS

SYSTEM	REFNO	SPACE GROUP	A/B	C/B	C/A	FORMULA	KEY
M	003030	P21/N	1.7423	0.424	.	CUC2OH1804	304
M	003031	C2/C	1.7513	1.409	.	K2NI(COS)4	304
M	003032	C2/C	1.7763	1.431	.	K2PD(COS)4	304
M	003033	P21/A	1.8125	1.750	.	C6H3(NO2)2OH@C10H6@@NH2	304
M	003034	C2/C	1.8178	1.462	.	K2PT(COS)4	304
M	003035	P21/A	1.9181	1.364	.	KAUCL4	304
M	003036	I2/M	2.0303	2.021	.	CUCL2	304
M	003037	P21/A	2.0540	1.784	.	C6H5NH2@C6H2(NO2)3OH	304
M	003039	P21/A	2.1439	1.083	.	C6H10@32	304
M	003040	P21/A	2.1853	1.084	.	C6H10I2	304
M	003040		2.1853	1.084	.	C6H10I2	303
M	003041	P21/N	2.4017	2.188	.	CUC10H1004	304
M	003042	P21/C	2.5000	1.443	.	C6H4(C6H5)2	304
M	003047	C2/C	2.6530	2.066	.	C5H804	304

Figure 8. Error listing from numerical verification preparation routine

ERRORS FROM VALIDN

SYSTEM	REFNO	A/B	C/B	C/A	VOLUME	DX	CALDX	EST. DX	FORMULA	KEY
M	3028	1.723	1.717	0.0	5116.3437			3.444	CBR4	212
M	3029	1.732	1.732	0.0	6110.8008				CI4	205
M	3038	2.095	1.311	0.0	450.1479			1.447	C6H2(CH3)2(NO2)2	212
M	3040	2.185	1.084	0.0	438.9856				C6H10I2	205
M	3042	2.500	1.443	0.0	625.2683	1.211	1.223		C6H4(C6H5)2	211
M	3043	2.523	0.802	0.0	1832.0564	0.998	1.002		C40H56O	211
M	3047	2.653	2.066	0.0	627.3699	1.388	1.398		C5H8O4	211
M	3048	2.820	2.077	0.0	2658.4507			1.520	C6H2(NO2)3NHC6H5	212
M	3049	2.958	2.040	0.0	2129.1047			1.878	(C3H5)2C(CONH)2CO	213
M	3049	2.958	2.040	0.0	2129.1047			1.492	(C3H5)2C(CONH)2CO	213
M	3049	2.958	2.040	0.0	2129.1047			1.685	(C3H5)2C(CONH)2CO	213
M	2992	0.240	0.212	0.0	1233.9939			2.979	BA(N3)2	212

Figure 9. Error listing from numerical validation routine

column is used to print the program's density calculations when no input figure is given in the record. The accumulated errors are then sent to the National Bureau of Standards to correct the original tapes. Multiple passes over the data may be made to ensure minimal errors prior to initiating publication of *CRYSTAL DATA*.

SUMMARY

The utilization of computer output to drive a photocomposition machine has not solved the problem of verifying the content of the publication, especially in the area of technical reference publications. In the case of *CRYSTAL DATA*, the proofreading and hand checking of calculations is a tedious job, and total reliability is not guaranteed. To eliminate this task, a program has been developed to eliminate most proofreading and hand calculations for the next publication of *CRYSTAL DATA*. The development of this program is being supported jointly by the National Standard Reference Data System of the National Bureau of Standards and the American Crystallographic Association, the originators of the publication.

Direct benefits of the program include:

1. Decreased time between publication and utilization of the most recent data through updating procedures.
2. An over-all cost savings per publication to the National Bureau of Standards.
3. A higher degree of input accuracy will be achieved for this and future editions

Secondary benefits include practically error-free single crystal data master tapes at The Pennsylvania State University which will contain the most recent data available. These tapes will be built from the corrected information at the end of the program. It is anticipated that a future comparison between the Penn State tapes and the American Society for Testing and Materials powder diffraction data files will result in the best possible data source for crystallographic research.

LITERATURE CITED

(1) Donnay, J. D. H., Donnay, Gabrielle, Cox, E. G., Kennard, Olga, King, M. V., "Crystal Data Determinative Tables," Williams and Heintz Map Corp., Washington, D. C., 1963.
(2) "Final Specifications for the NBS Crystal Data Program," Service Bureau Corp., 1966.
(3) Taylor, L. B., "Crystal Data Editor—A Study in Non-Arithmetic Programming," Master's thesis, The Pennsylvania State University, University Park, Pa., 1968.