Overview of the role of data reviews and tutorial reviews in improving crystallographic science training

Petra Bombicz

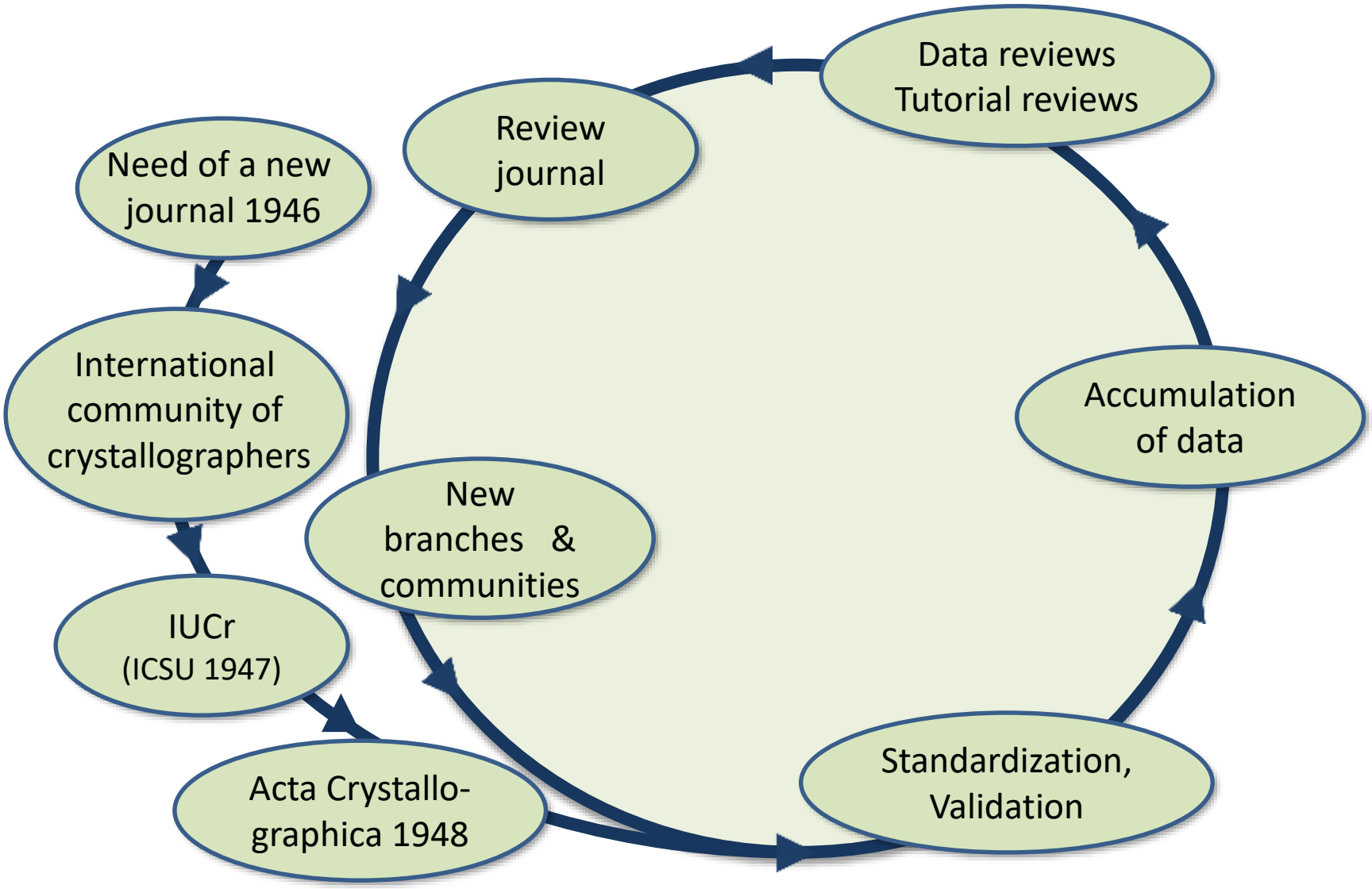
Editor of Crystallography Reviews

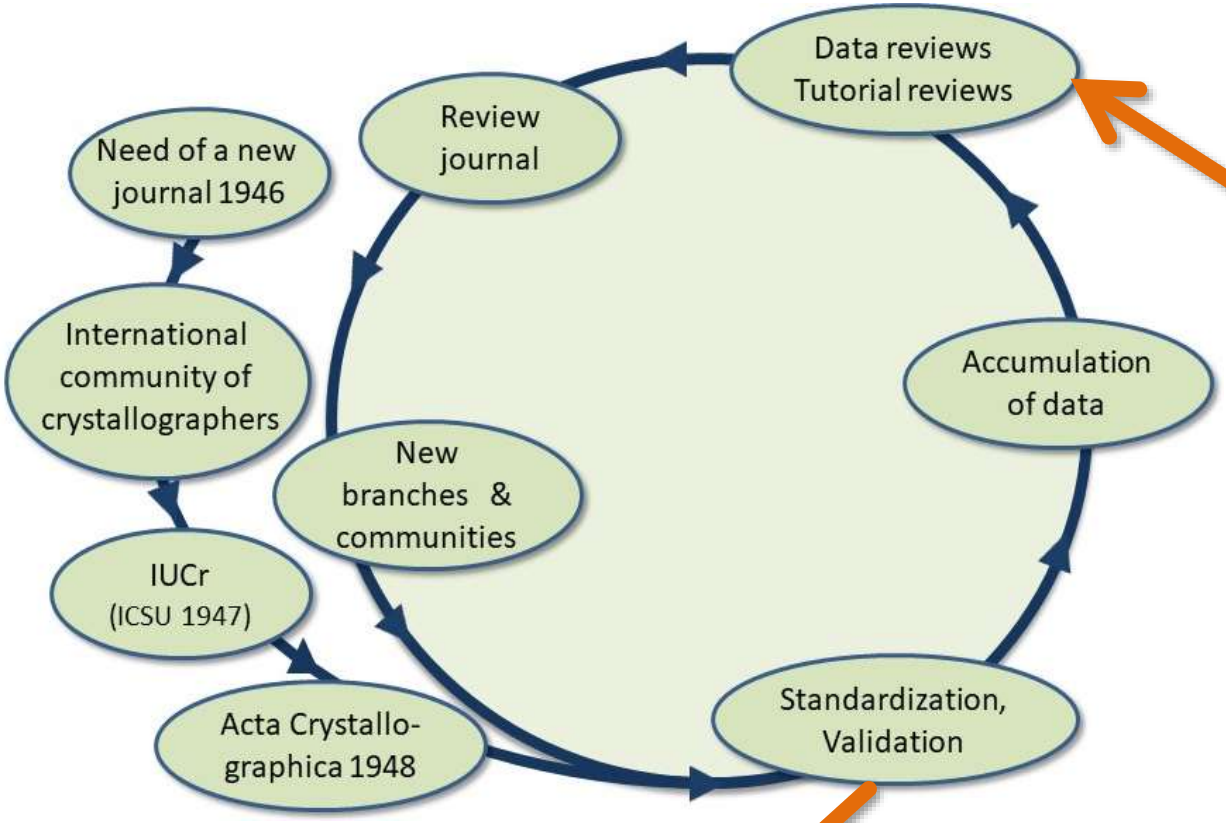
Research Centre for Natural Sciences, Hungarian Academy of Sciences

Workshop on Data Science Skills in Publishing:
for Authors, Editors and Referees

ECM32 Vienna

18th August 2019


Need of a new journal 1946
International community of crystallographers
IUCr (ICSU 1947)
Acta Crystallo-graphica 1948
Review journal
Data reviews Tutorial reviews
Accumulation of data
Standardization, Validation
New branches & communities



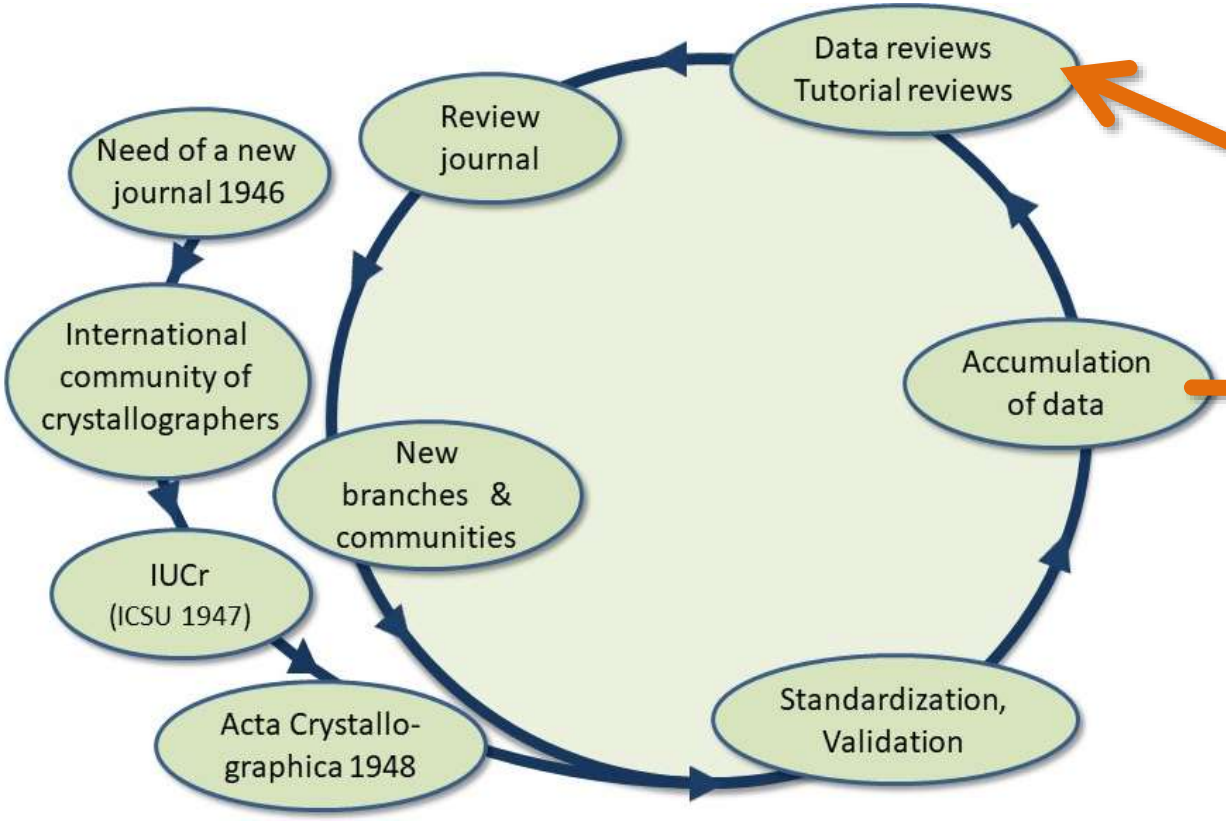
Information retrieval

Searchable data
Improving quality of data

1991 cif
precise definition of data items
standardization of terminology

1992 Introduction of electronic submission of articles

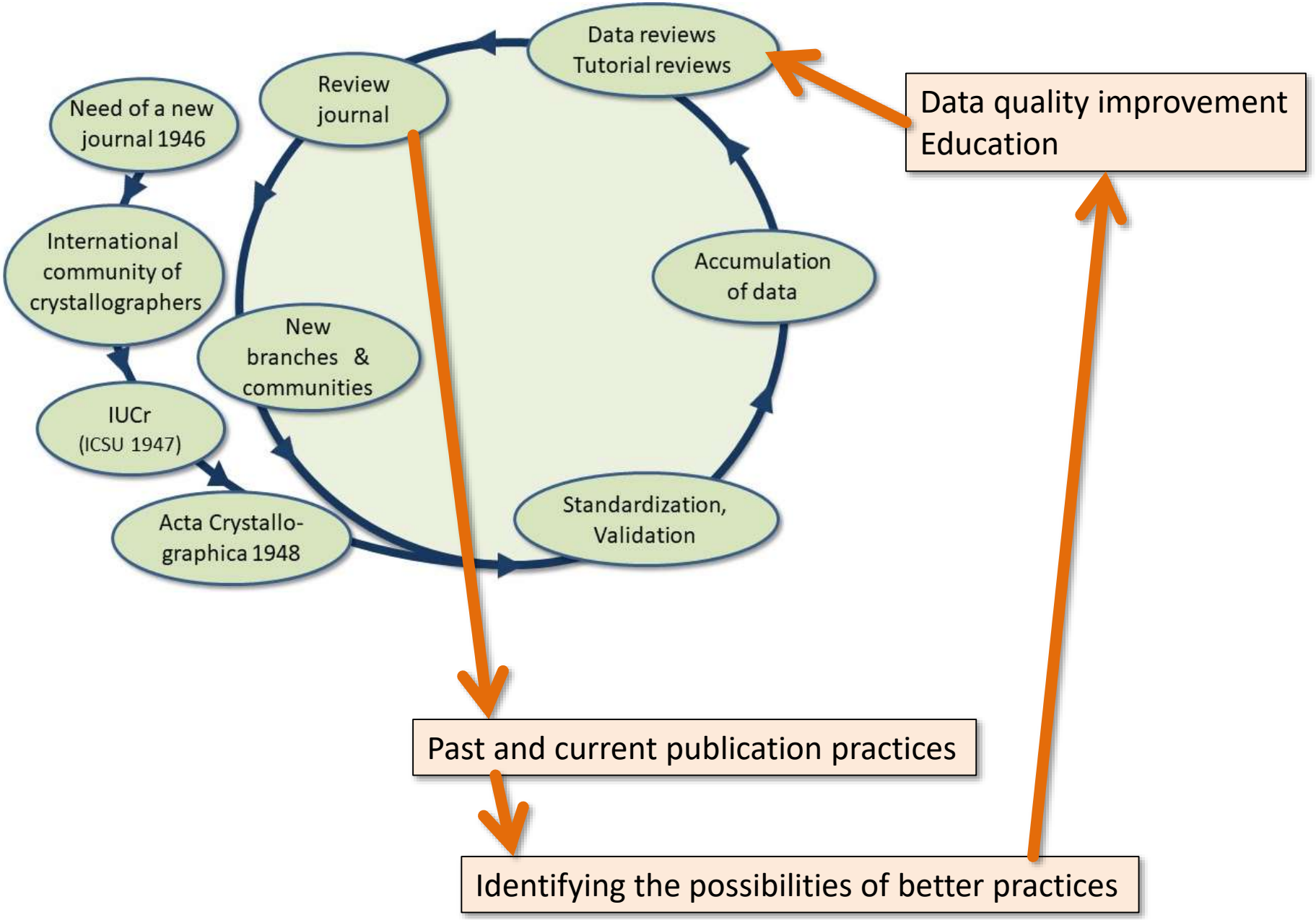
1996 checkcif
automated structure validation in chem cryst
level of experience: authors, reviewers, coeditors



Experimental diffraction data and calculated structural data need to be accurate, complete, consistent and correct.

FACT=Fairness, ACcuracy and Transparency

FAIR=Findable, Accessible, Interoperable and Reusable


Need of a new journal 1946
International community of crystallographers
IUCr (ICSU 1947)
Acta Crystallo-graphica 1948
Review journal
Data reviews Tutorial reviews
Accumulation of data
Standardization, Validation
New branches & communities
Data quality improvement Education
Past and current publication practices
Identifying the possibilities of better practices

Crystallography Reviews

The first issue was published in 1987.

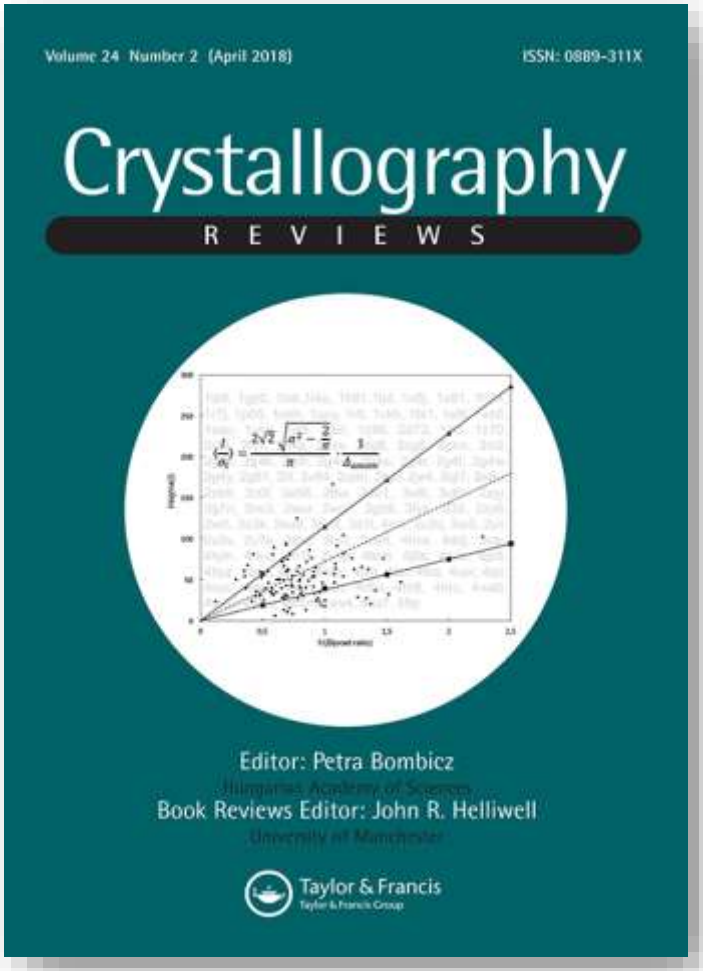
Editors:
Moreton Moore (1987), John R. Helliwell (2007)

Reviews:
established area and ones of emerging importance

Reviews on topics:
crystallography and crystal growth,
biological, chemical, physical, mineralogical and industrial crystallography
covers all theoretical and applied aspects

Online publication, unlimited use of colour,
printed copy edition is published as a single volume.

Regular quarterly issues:
since T&F has taken over the publication

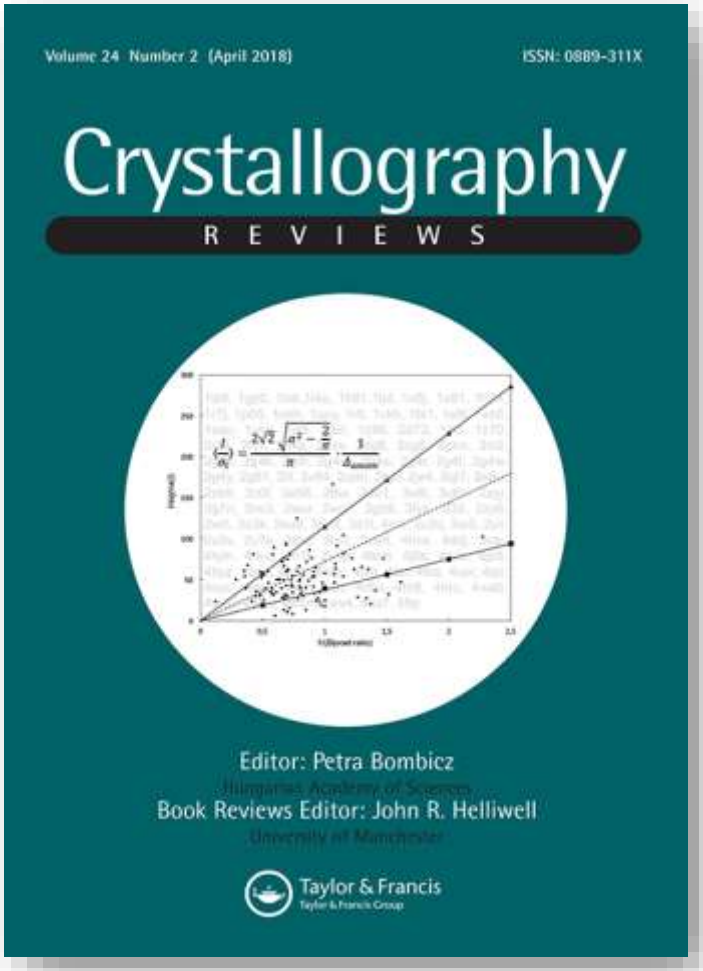


Range of review types:

- full review articles
- shorter reviews:
  topical, historical, tutorial, evaluation,
  data reviews (most recently)
- book reviews

Crystallography Reviews' full reviews:
total length can be around 25 000 words
occasionally even longer, up to 35 000 words

It is a unique feature of Crystallography Reviews not published by other journals
(typical book size)

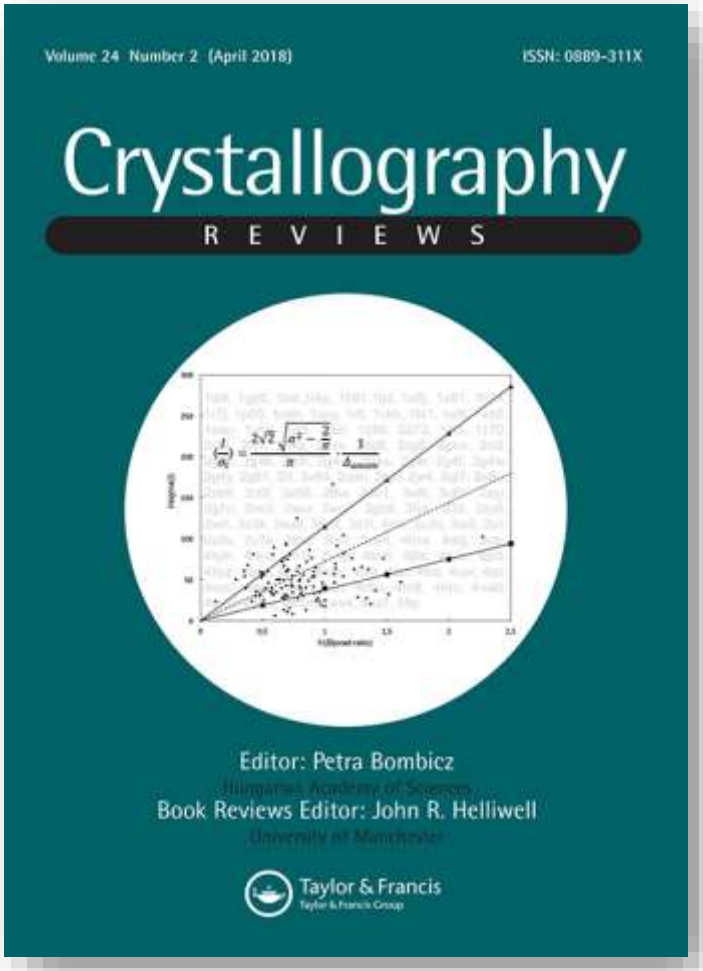


The reviews:

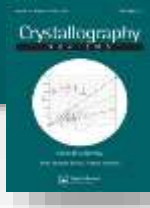
- tutorial reviews:
  direct educational training at all career stages
- data reviews:
  insights to the collections of
  published crystallographic data

Tutorial and data reviews from Crystall Reviews:

- articles mentioning some specific
  authors' conclusions
- special relevance to improving the publishing of data
  - general data weaknesses
  - specific error cases that are found
- identifying possibilities of better practice



Series of ’Data science skills’



Data science skills for referees: I. biological X-ray crystallography

by John R Helliwell CRYSTALLOGRAPHY REVIEWS, *2018, VOL.24, NO.4, 263–272*

Views: 141

a summary of the data evaluation checks, the referee is expected to do
suggestion: a referee would undertake one cycle of model refinement
of the authors’ coordinates against the authors’ processed diffraction data
to referee diffraction data, coordinates and PDB validation reports

certified referees with data science skills

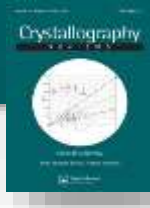
Data science skills for referees: II. powder X-ray diffraction

Data science skills for referees: III. electron crystallography

Data science skills for referees: IV. chemical crystallography

Data science skills for referees: ….. ?

Chemical Crystallography: when are ‘ bad data’ ‘good data’?



Amber Thompson CRYSTALLOGRAPHY REVIEWS, 2019, VOL. 25, NO. 1, 3–53

Difficult data can be associated with:

* sample, * data collection, * image reduction, * model, * chemistry

Discussions:

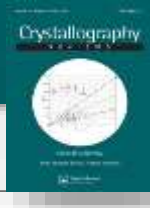
- quality of the achievable diffraction data
- methods for analysing data
- some of the influences on data quality
- attempt to determine whether
  some of the validation tropes widely used are sound advice, or ‘Chinese Whispers’

Overview:

- the mounting of the crystal
- data collection strategy
- effect of temperature
- effect of exposure time
- redundancy
- completeness
- demand what defines ‘good’ data

cost-benefit analysis before starting an experiment (fit for purpose):
not to sacrifice quality on the altar of high-throughput

Metrics for crystallographic diffraction- and fit-data: a review of existing ones and the need for new ones



Julian Henn CRYSTALLOGRAPHY REVIEWS, 2019, VOL.25, NO.2, 83–156

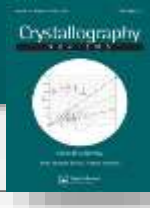
data quality improvement of diffraction data
how to detect and discribe the degree of systematic errors

- standard metrics are revisited
- new developments in metrics of SXRD are suggested
- tentative standard protocol for fit data analysis
- detection and hierarchy of systematic errors are given
- general requirements for good refinements are provided
- two case studies (IUCrJ, IUCrData)

the largest systematic error in most data sets is the incorrectly estimated s.u.s.
it blocks the investigation of other systematic errors

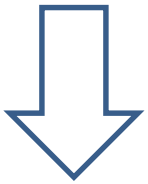
adequate s.u.s for the observed intensities:
much too long neglected
solution in data integration and data processing steps

Metrics for crystallographic diffraction- and fit-data: a review of existing ones and the need for new ones

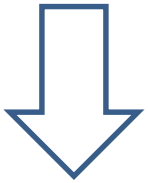


Julian Henn CRYSTALLOGRAPHY REVIEWS, 2019, VOL.25, NO.2, 83–156

the responsibility of the diffractometer manufacturers' integration software
to provide better error estimates for SXRD data
the responsibility of the scientists
to be more realistic about their results



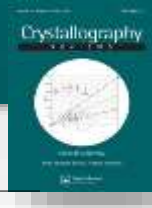
the need for availability of raw diffraction data
differing processing packages can be compared
in their estimation of the σ–s on Bragg reflection intensities



archiving raw diffraction images

the community is aware of the need
of improvement in hardware as well as in software

Refining the macromolecular model – achieving the best agreement with the data from X-ray diffraction experiment

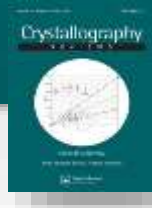


Ivan G. Shabalin, Przemyslaw J. Porebski, and Wladek Minor
CRYSTALLOGRAPHY REVIEWS, 2018, VOL.24, NO.4, 236–262

Tutorial review:

- describes the most practically important concepts of protein structure refinement
- offers guidelines for choosing the best settings for the reciprocal-space refinement of macromolecular models
- provides practical tips for
  - manual model correction
  - modelling of side-chains with poor or missing density
  - ligand identification
- guidelines for choosing the best refinement settings

The signal-to-noise ratio in SAD experiments



Andrzej Olczak and Michele Cianci
CRYSTALLOGRAPHY REVIEWS, 2018, VOL.24, NO.2, 73-101

Extensive survey of 115 sulphur SAD (Single-wavelength Anomalous Dispersion) depositions (PDB)

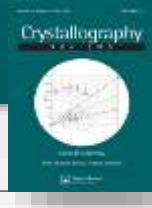
The statistical distributions of the experimental values of $|\Delta I_{anom}| / \sigma(\Delta I)$

The mathematical model allows a-priori prediction of the overall $I/\sigma(I)$,
necessary for successful SAD phasing
this requires the correct description of
the sample composition, number of protein atoms and anomalous scatterers
the experimental conditions (λ especially)

The model described is valid for phasing experiments
with anomalous signal at any X-ray source including XFEL.

Application of the SAD experiment predictor, where
a sample is particularly X-ray radiation sensitive
high $I/\sigma(I)$ cannot be assumed

Should we remediate small molecule structures?
If so, who should do it?



Carle H Schwalbe, CRYSTALLOGRAPHY REVIEWS, 2018, VOL.24, 4, 217-235

1942-2019

Views: 152

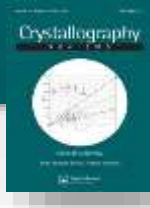
Checkcif: it should highlight errors prior to deposition or publication
errors appear even in some recent structures
Deposition of structure factors along with cifs is increasingly frequent
it enebles retprospective detection and correction

- Missing symmetry
  Richard Marsh: 10% of structures described in *P*1 1999-2004 needed revision
  2% of structures described in *P*1 2010-2013 needed revision
- Location of H
  misplaced H atoms of acids and imidazole derivatives,
  (in carboxylic acids COH *vs* C=O bond length, clashing H atoms, OH…A)
  misplaced H atoms in dihydrogen phosphates, water aggregates
- Mis-identified elements

- ❑ Should we remediate small molecule structures?
- ❑ If so, who should do it?
- ❑ Which structures should be remediated?

carefully selected examples of detecting and correcting errors
suggestions for more comprehensive detection and correction of errors in deposited data

X-ray scattering characterisation of nanoparticles



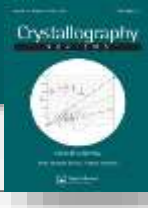
Bridget Ingham CRYSTALLOGRAPHY REVIEWS, VOL.21, No. 4, 229–303

Top cited article for impact factor calculations in the last years. Citations: 43

Tutorial review

- theoretical and practical overview of X-ray scattering methods used to characterize nanoparticles covering XRD and SAXS
- case studies of basic science and industrial applications on a wide variety of nanoparticle materials and synthesis methods
- recent advances in related techniques such as anomalous scattering and pair distribution function analysis

Crystal chemistry of organic minerals – salts of organic acids: the synthetic approach



Oscar Enrique Piro and Enrique José Baran
CRYSTALLOGRAPHYREVIEWS, 2018 VOL. 24, NO. 3, 149–175

modern data collection, advanced SG and crystal structure determinations and routine untwining procedures

Full crystallographic characterization of minerals is frequently hampered by:

- unavailability of natural samples (single crystals),
- lack of purity and other disorders of these materials

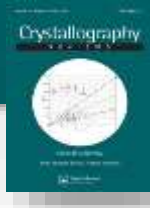
Detailed crystal structure determination of synthetic analogues of organic minerals before the discovery of their natural counterparts:

$Mg(HCOO)_2 \cdot 2H_2O$; $Cu(CH_3COO)_2.H_2O$; $CaCu(CH_3COO)_4.6H_2O$;
$Na_2C_2O_4$; $Ca(C_2O_4) \cdot 3H_2O$; $MnC_2O_4 \cdot 2H_2O$, NaMg[Fe(C2O4)3] · 9H2O
NaMg[Al*x*Fe1−*x*(C2O4)3] · 9H2O

Synthesis of mineral analogues, followed by SXRD provides a powerful approach

- to uncover the full crystal and molecular structure of natural minerals
- providing structural information on related natural analogues hitherto undiscovered

Book reviews



The book reviews:
help readers select which books they should spend their time on and learn from

Watson Fuller CRYSTALLOGRAPHY REVIEWS 2009, VOL. 15, NO. 3, 219–222
on
Molecular and cellular biophysics, CRC Series in Pure and Applied Physics, Chapman & Hall, CRC Press, 2008, by Jack A. Tuszynski

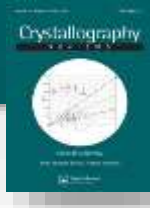
Views: 39

„… the reader can reasonably wish to be assured that the material presented is accurate and the conclusions drawn are the product of informed and mature analysis.”

Apects:

- Understanding - deep
- Facts - accurate
- Organisation - logical
- References - precise
- Figures and Tables - informative

Book reviews



The book reviews: Data opens new opportunities

David W.H. Rankin CRYSTALLOGRAPHY REVIEWS 2009, VOL.15, NO.3, 223–224
on
CRC handbook of chemistry and physics, (The Rubber Book) 89th edition,
edited by David R. Lide, Boca Raton, CRC Press, Taylor and Francis Group, 2008 2736 pp.

Views: 2362

”It contains a mass of information. If you want to know about the absolute flux of the hard component of cosmic rays, … or the natural logarithm of π to 50 decimal places, they are all there. Well, nearly all.
There is a section on solid-phase structures, but the data on bond lengths in crystalline organic compounds have not been updated. This section still consists of … an analysis of interatomic distances in the CSD, 1985 version. That is getting on for a quarter of a century ago, so surely something better could be done now?”

100th edition, June 7, 2019, Editor-in-Chief John R. Rumble (1532 pages, -2000ps)
comprehensive one-volume reference resource for science research
”*Handbook* contained myriad information for every branch of science and engineering. Later editions focus almost exclusively on chemistry and physics topics and eliminated much of the more ’common’ information.”

Summary

the role of data reviews and tutorial reviews

- Initialise check of data quality may result in data improvement
- Improvement of the refinement of the model
- Provide statistical distributions
- Contribute to the ability of prediction
- Suggest to remediate structures
- Help to the work of referees and editors
- Book improvement (new edition)
- etc …

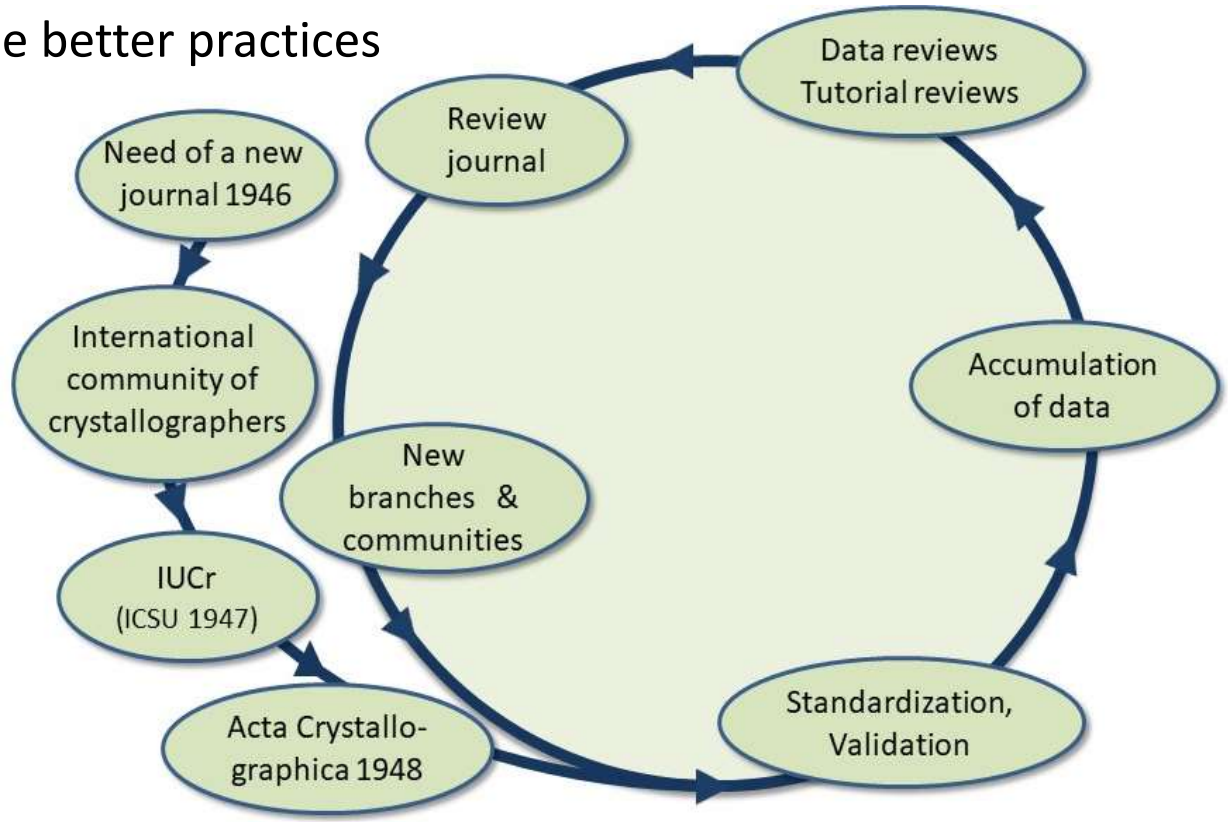
Summary

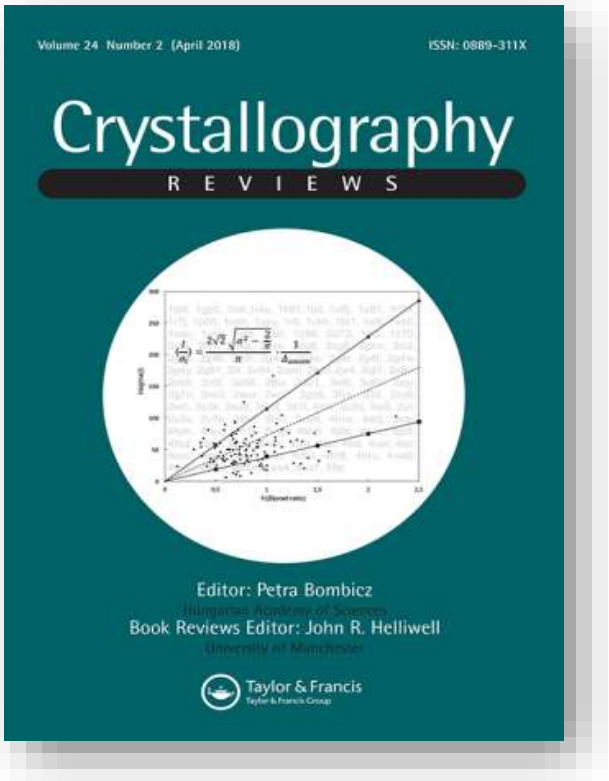
the role of data reviews and tutorial reviews

The data reviews in particular illustrate:
the current and past publication practices of our crystallographic fields and subfields.

The reviews contribute to

- data quality improvement
  - general data weaknesses
  - specific error cases that are found
- the identification of possible better practices
- education





Acknowledgements:

Prof John R Helliwell
Prof Moreton Moore

Taylor & Francis
Managing Editors and Publication Editors

Data reviews and tutorial reviews' role in improving crystallographic science training

Book reviews

John R Helliwell Acta Cryst. (2014). A70, 92–94
on
Early Days of X-ray Crystallography. IUCr/Oxford University Press, 2013.
By André Authier.

„It should be studied by all with an interest in where we came from in our field of crystallographic science, and is a guide to where we are going.”

missing:
The nature of the uncertainty of the determination of atomic positions is an odd omission, not of the author but of accomplished physicists of the time.

Since crystal structure analysis allows one to not only see atoms but also to know the precision of the method, and thereby deliver the various standard uncertainties of atomic positions and *B* factors.