Release by the International Centre for Diffraction Data (ICDD) of New Powder Data Mining Tools: the PDF-4/Full File and PDF-4/Organics Databases

Brian O'Connor*, Camden Hubbard#, Tim Fawcett and John Faber
International Centre for Diffraction Data, Newtown Square, PA, USA.
* Curtin University of Technology, Perth
# Oak Ridge National Laboratory, Tennessee

Topics to be Covered in Talk

- ICDD organisation
- The Powder Diffraction File
- The ‘new’ PDF-4 databases
- Materials identification with the PDF
- Database structure, and use in data mining
- Data mining examples
- PDF-4 demonstration

What is the ICDD?

The ICDD comprises a volunteer scientific membership of 350 diffractionists from 41countries, plus a headquarters staff of 38.

Dedicated to -

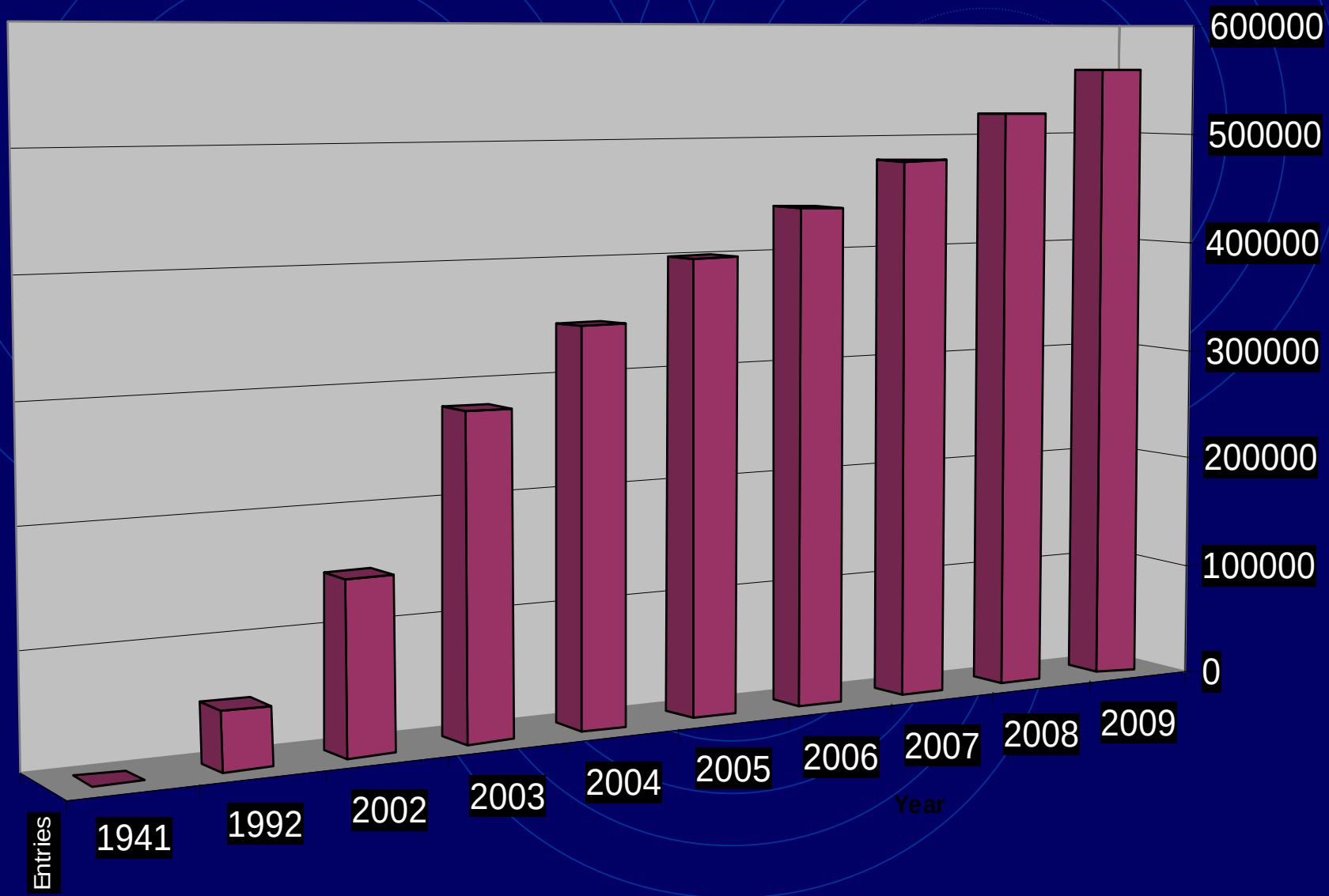
- On-going development of the PDF, including its Grant-in-Aid program
- Advancement of x-ray instrumental science
- X-ray analytical science education – workshops, Denver X-ray Conference, etc
- International networking – conference and workshop support, etc
- Supporting PhD students through its scholarship program

The PDF Data Base in 2003

- PDF is designed to identify crystalline material.
- Currently contains 279,854 unique data sets characterised by x-ray diffraction
- Physical properties, experimental preparation and literature citations included.

For the next several years ~ 30,000 materials will be added to the PDF per year.

Growth of the PDF



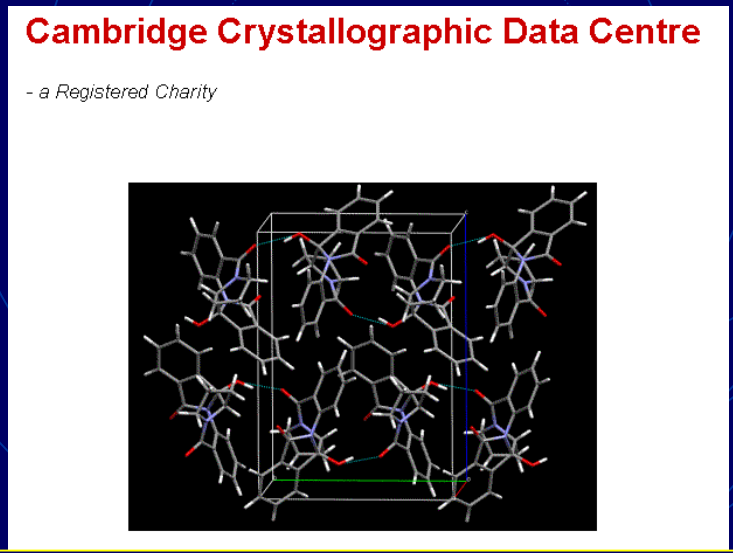
Searchable PDF Descriptors

In addition to 280,000 data sets, you will find:

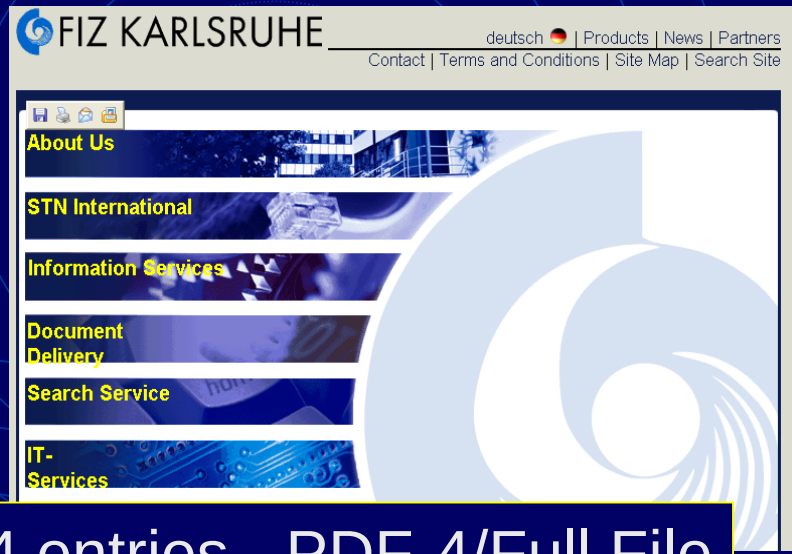
- 200,000 densities
- 100,000 colour classifications
- 50,000 melting points
- 173,000 distinct empirical formulae
- Over 500,000 reference citations
- 1,602 different searchable journal codens

Database Partners with the ICDD

92,011 entries – ICDD Sources



122,816 entries - PDF-4/Organics



56,614 entries - PDF-4/Full File

8,423 entries - PDF-4/Full File

ICDD-MPDS Collaboration on Inorganic Materials – *New !*

- Inorganic structures and properties abstracted by the Materials Phases Data System and MPDS-JST (Japan Science & Technology Corporation)
- Access to future abstracts from Linus Pauling File project by MPDS
- S-entry structural data, including atomic coordinates added to the PDF
- Data sets are added to the database after carrying out the duplicate review to exclude overlap with all other sources.
- PDF entries will be updated with synthesis and reference citations
- ~ 100,000 atomic coordinate sets to be added in 2005

PDF-4 Master Database

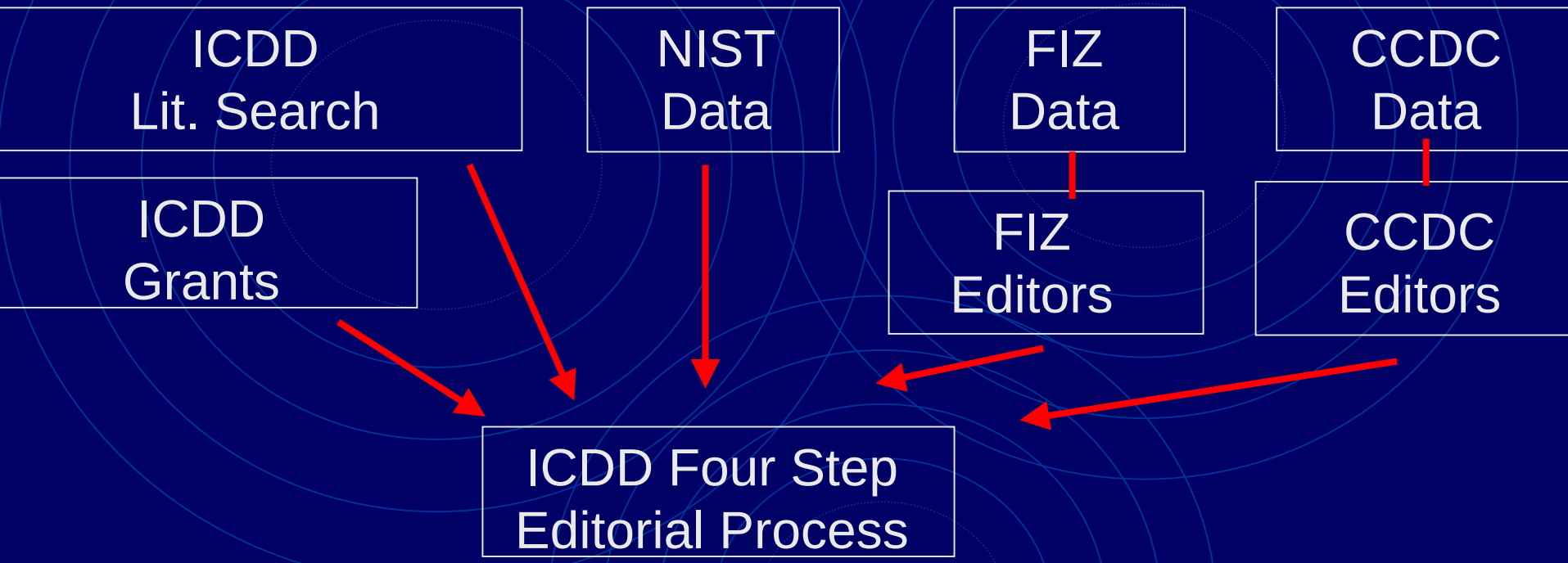
280,000 entries released in three PDF-4 products

PDF-4/Full File
157,048 entries

PDF-4/Minerals
17,535 entries

PDF-4/Organics
147,201 entries

ISO-9001 PDF Review Processes



All data put in a common format, statistically analyzed, and classified in numerous ways

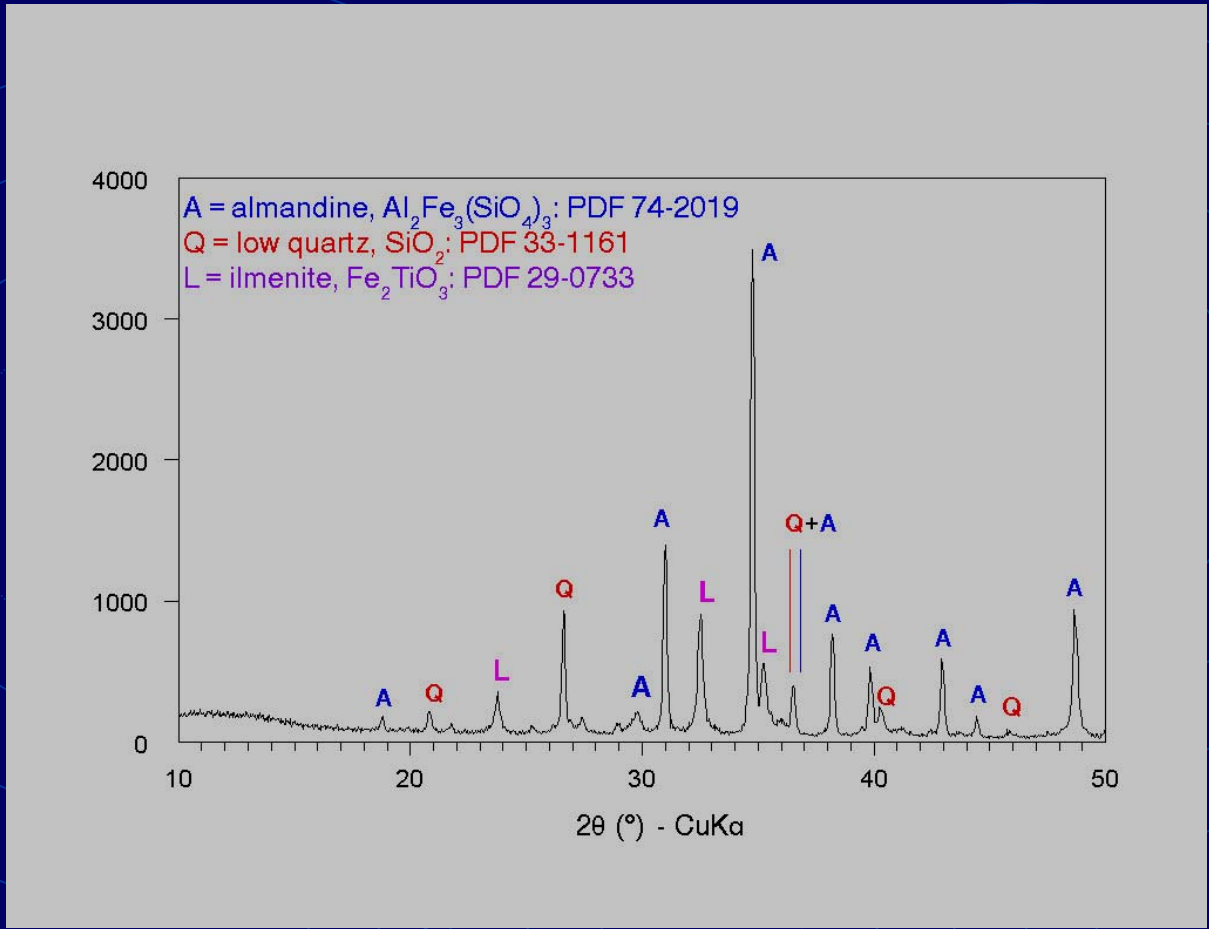
XRD Identification with the PDF

Example

Red coloured material covering beach at the mouth of the Chapman River, Geraldton, Western Australia.

Residents expressed concern that the material might be polluting residue from mineral sands Industry.

Powder data analysed with the MDI search/match identification program and the PDF-4 database



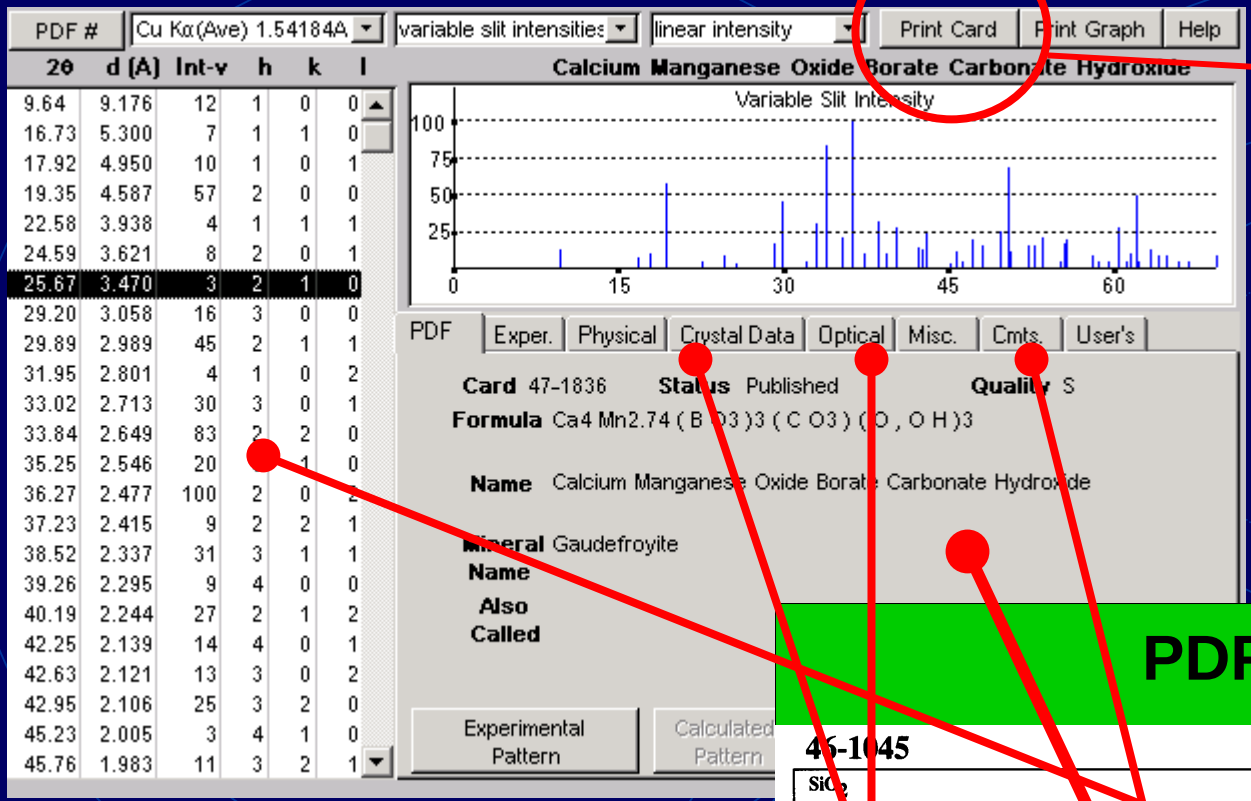
Result:

- The pattern is dominated by the phases almandine, quartz and ilmenite. Almandine and ilmenite are present as minor phases in 'normal' beach sand.
- Erosion of the beach by water outflow from the river, plus wave action, has preferentially removed the less-dense quartz. There is no evidence of mineral processing residues.
- The garnet almandine accounts for the red coloration.

PDF-4 Format

The PDF-4 format is a new relational database container for the PDF

- The data are arranged in a series of database tables.
- The format allows searchable access to all physical property data fields.
- The PDF-4 offers 30+ separate searchable diffraction and physical property fields.
- Integrated display, search and indexing software is standard with the PDF-4 format. (Contact your software distributor for automatic search/match software.)



Still have "card" option

New format – data rearranged in tabular sections with point and click interfaces

PDF Card

46-1045 ★

SiO_2

Silicon Oxide

Quartz, syn

Rad. $CuK\alpha_1$ λ 1.540598 Filter Ge Mono. d-sp Diff.
Cut off Int. Diffractometer I/I_{cor} 3.41
Ref. Ke[illegible] A., Eysel, W., Mineralogisch-Petrograph. Inst., Univ. Heidelberg, Germany, *ICDD Grant-in-Aid*, (1993)

Sys. Hexagonal S.G. $P3_221$ (154)
a 4.91344(4) b c 5.40524(8) A C 1.1001
α γ Z 3 mp
Ref. Ibid.

D_x 2.65 D_m 2.66 SS/FOM F_{30}=539(0.002,31)

εα nωβ 1.544 εγ 1.553 Sign + 2V
Ref. Swanson, Fuyat, *Natl. Bur. Stand. (U.S.), Circ. 539*, **3** 24 (1954)

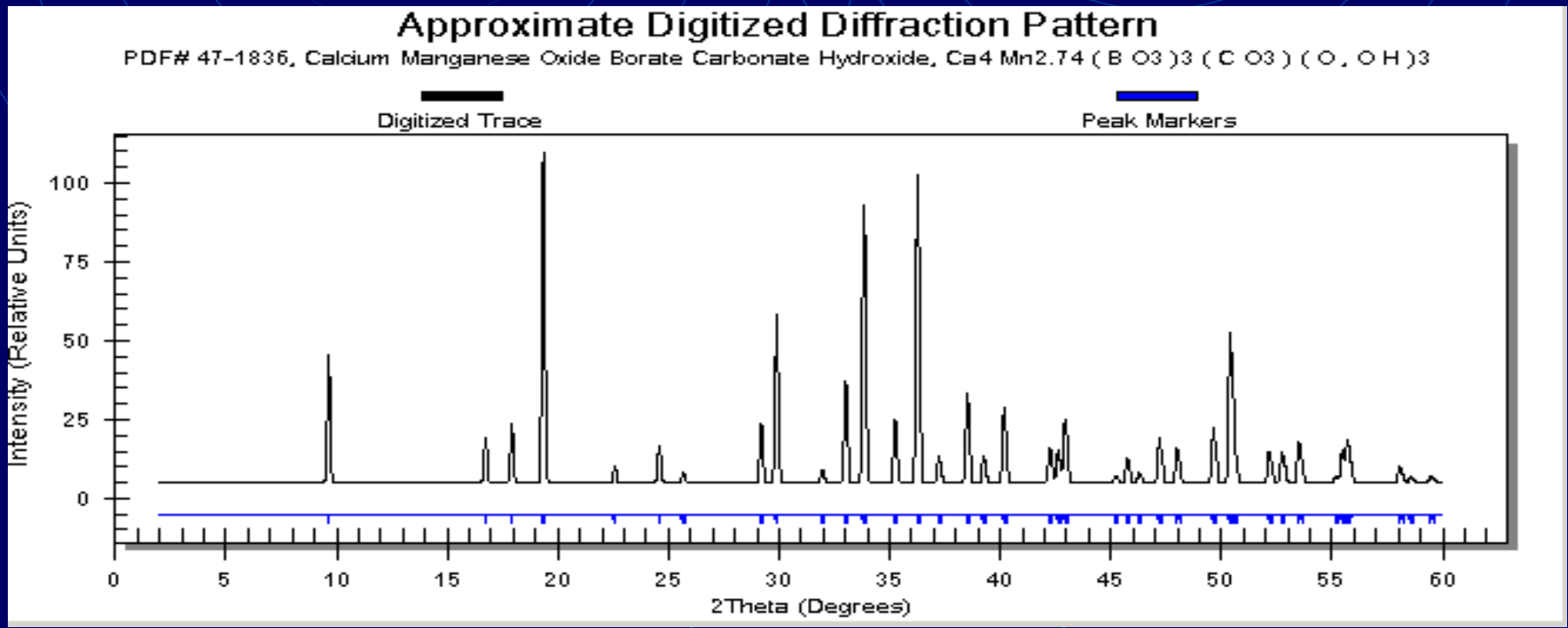
Color White
Integrated intensities. Pattern taken at 23(1) C. Low temperature quartz. 2θ determination based on profile fit method. O_2Si type. Quartz group. Silicon used as internal standard. PSC: hP9. To replace 33-1161. Structure reference: Z. *Kristallogr.*, **198** 177 (1992).

See following card.

dÅ	Int	hkℓ	dÅ	Int	hkℓ
4.2550	16	100	1.1530	1	311
3.3435	100	101	1.1407	<1	204
2.4569	9	110	1.1145	<1	303
2.2815	8	102	1.0816	2	312
2.2361	4	111	1.0638	<1	400
2.1277	6	200	1.0477	1	105
1.9799	4	201	1.0438	<1	401
1.8180	13	112	1.0346	1	214
1.8017	<1	003	1.0149	1	223
1.6717	4	202	0.9896	<1	115
1.6592	2	103	0.9872	<1	313
1.6083	<1	210	0.9783	<1	304
1.5415	9	211	0.9762	<1	320
1.4529	2	113	0.9608	<1	321
1.4184	<1	300	0.9285	<1	410
1.3821	6	212	0.9182	<1	322
1.3750	7	203	0.9161	2	403
1.3719	5	301	0.9152	2	411
1.2879	2	104	0.9089	<1	224
1.2559	3	302	0.9009	<1	006
1.2283	1	220	0.8972	<1	215
1.1998	2	213	0.8889	1	314
1.1978	<1	221	0.8814	<1	106
1.1840	2	114	0.8782	<1	412
1.1802	2	310	0.8598	<1	305

On-the-Fly Pattern Calculation

- Generate both “stick patterns” and digital patterns
- Digital patterns can be user modified to simulate the users instrumental conditions
- Sophisticated data mining ability (upcoming slides)
- Visualization software included and integrated



PDF-4 Master Database

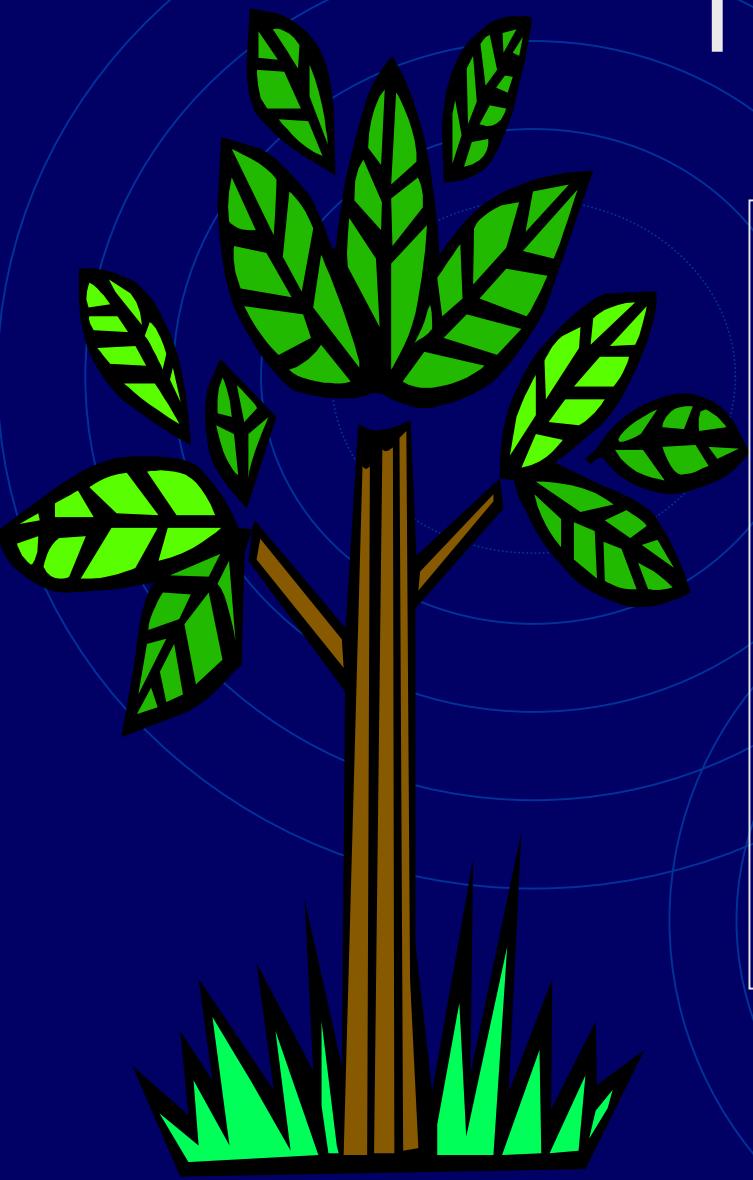
272,000 entries released in three PDF-4 products

PDF-4/Full File
157,048 entries

PDF-4/Organics
147,201 entries

PDF-4/Minerals
17,535 entries

The PDF Tree



Trunk = PDF Entry Number

Branches = 30+ Searchable Property Fields

Twigs = Classification Systems of the Property Fields

Leaves = Individual Material Property Values

Branches & Twigs

Searchable Property Fields

14 Colors
26 Subfiles
458 Mineral Classifications
230 Space Groups
206 Organic Functional Groups
1620 Journals Codes
> 500,000 Reference Citations

Continuous ranges of melting points, densities, I/Ic, unit cell parameters, temperature and pressure

Leaves

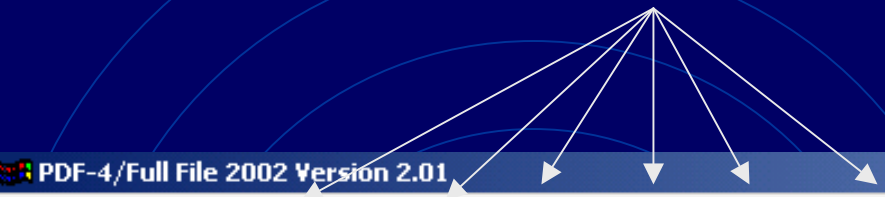
Types of Searches

- Type of Element
- Number of Elements
- Empirical Formula
- Common Name
- Compound Name
- Mineral Name
- Mineral Classification
- d-Spacing
- PDF Quality Mark
- Reference Intensity Ratio, I/Ic
- PDF-4 Sub-files
- Space Group
- Pearson Symbol Code
- Prototype Structure
- Reduced Cell Parameters
- Temperature and Pressure
- Density
- Melting Point
- Color
- Author Name
- Journal

<u>Additional Searches for PDF-4 Organics</u>

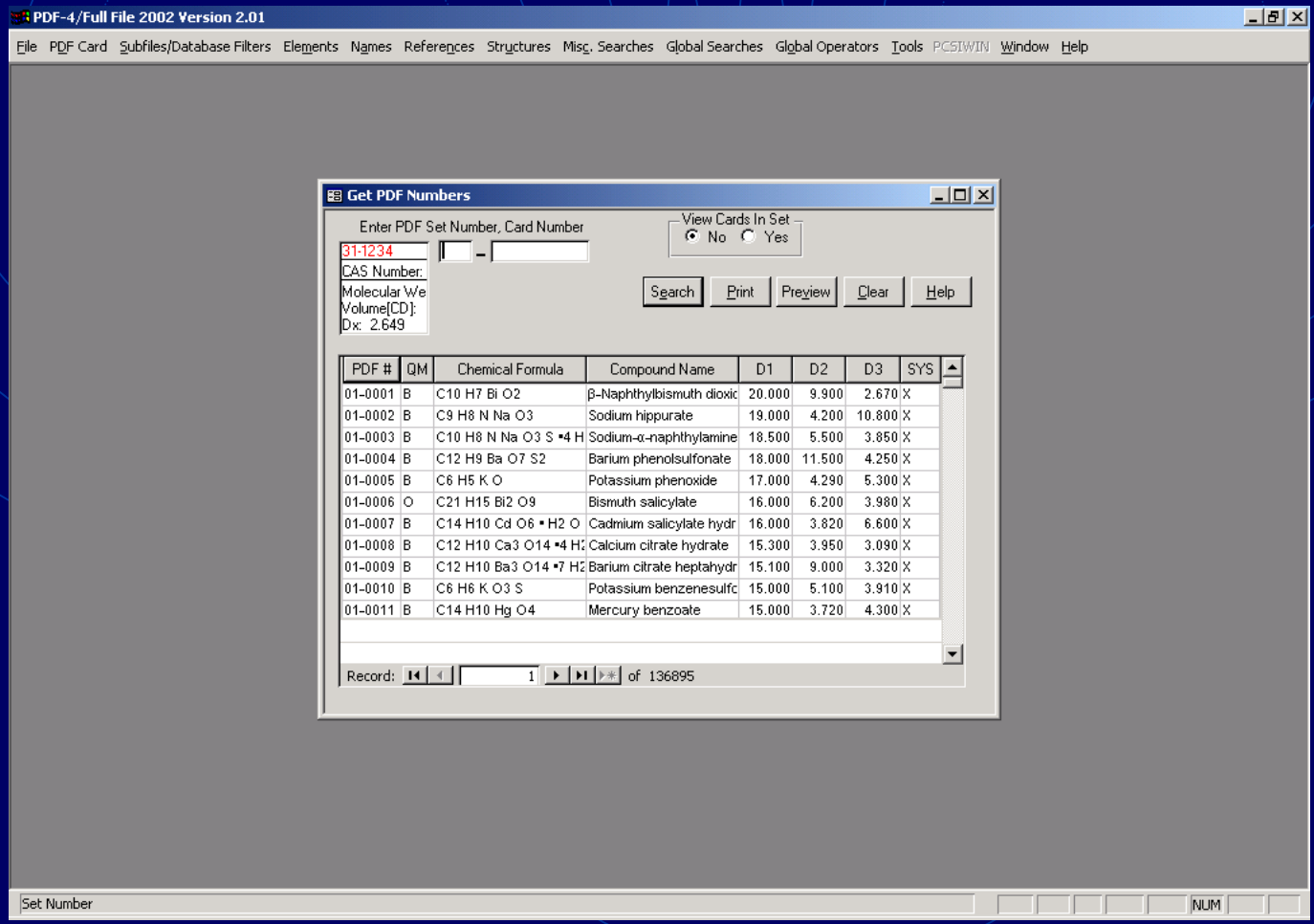
- Organic Functional Group (206)
- Pharmaceutical
- Drug Activity
- Polymers
- Pigments
- Forensics

The Main Branches

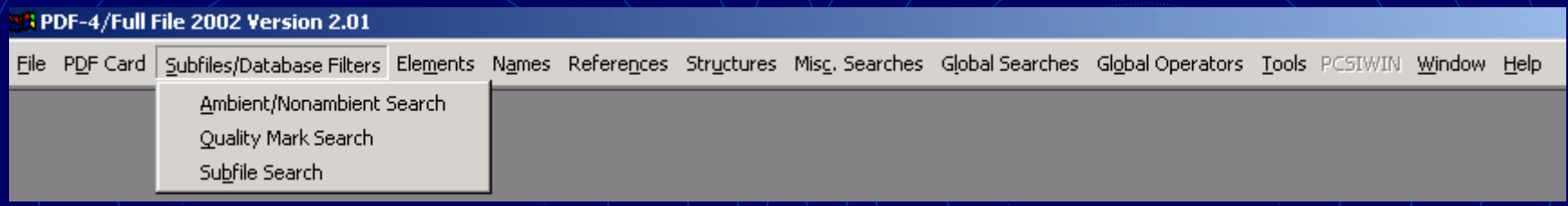


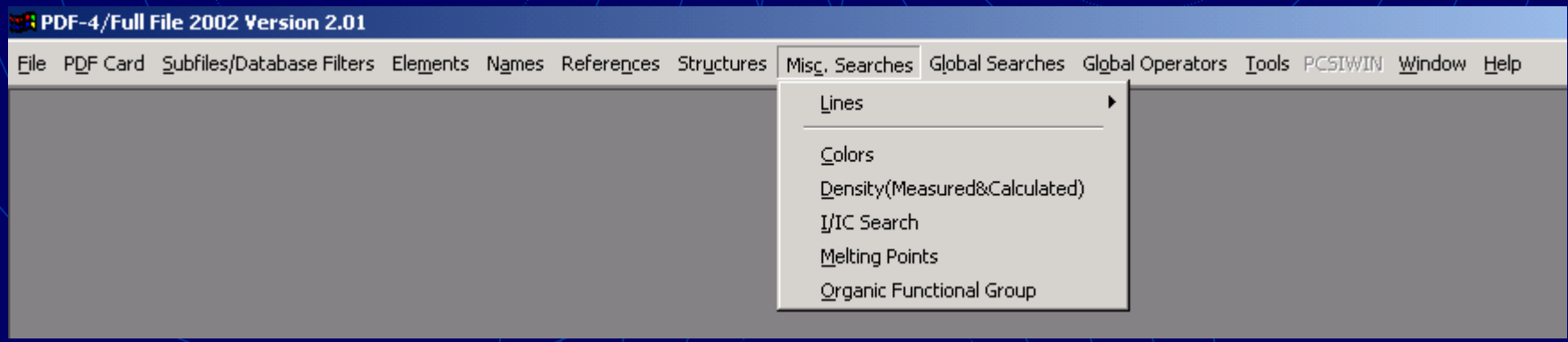
PDF-4/Full File 2002 Version 2.01

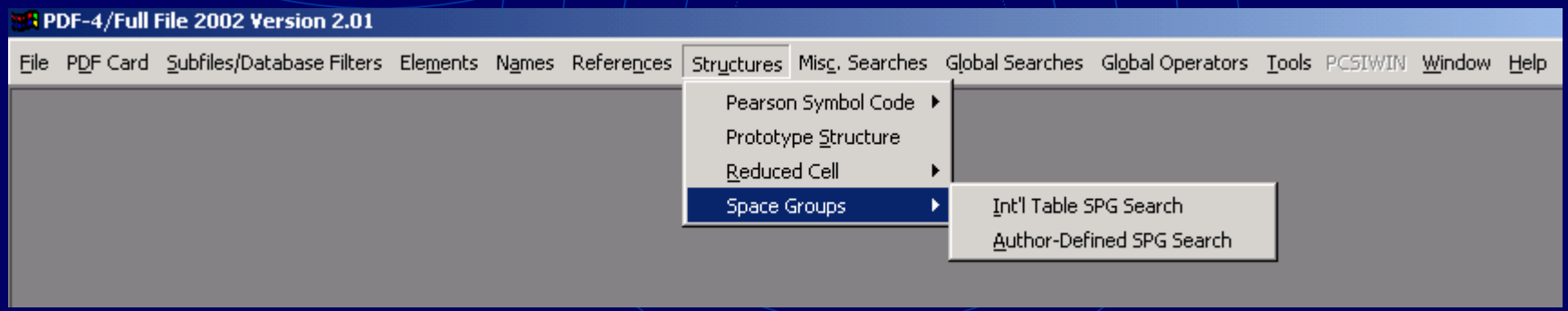
File PDF Card Subfiles/Database Filters Elements Names References Structures Misc. Searches Global Searches Global Operators Tools PCSIWIN Window Help



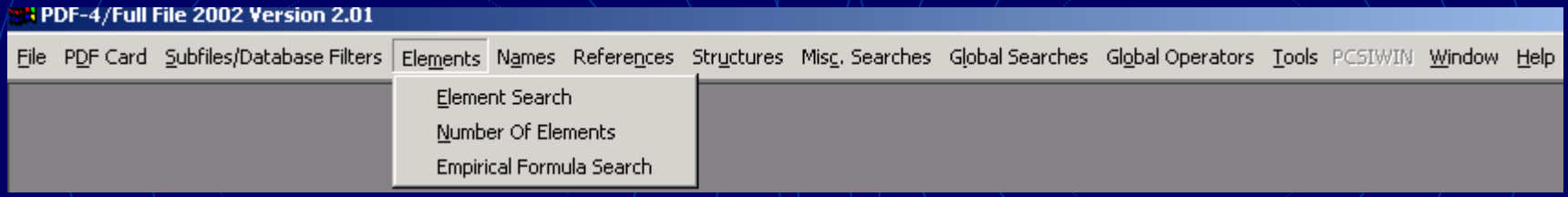
More Branches

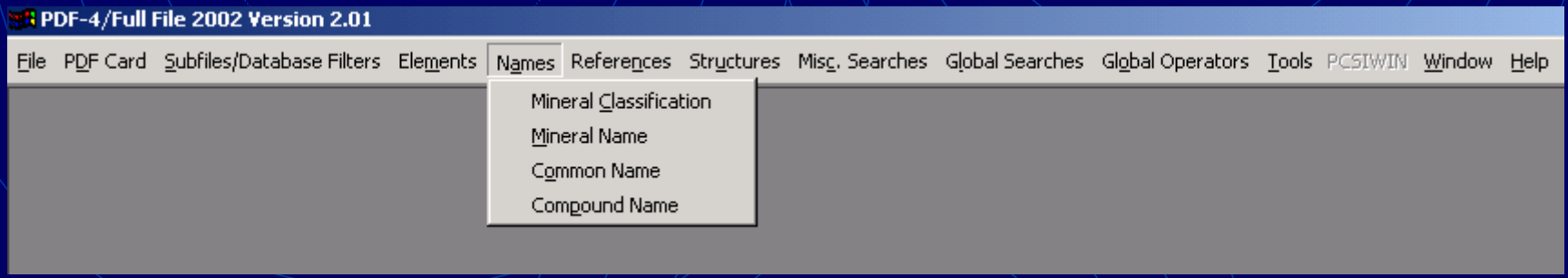


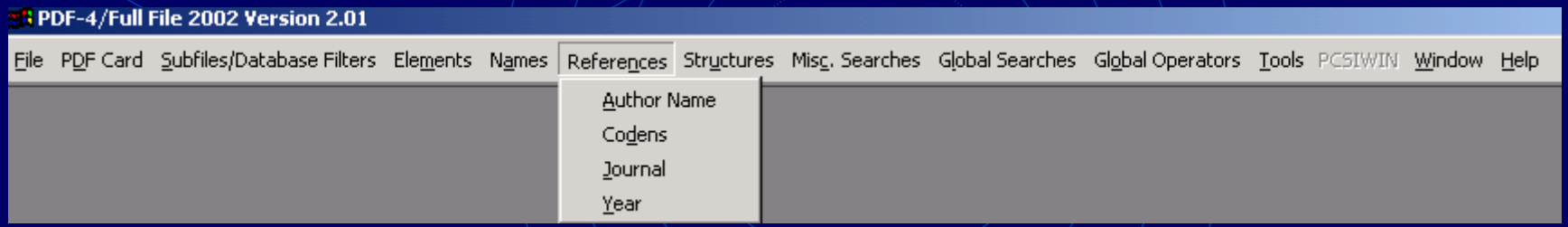




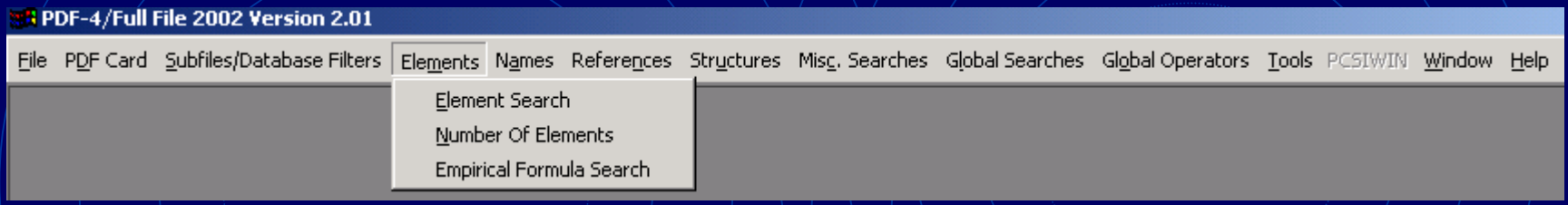
And Even More Branches



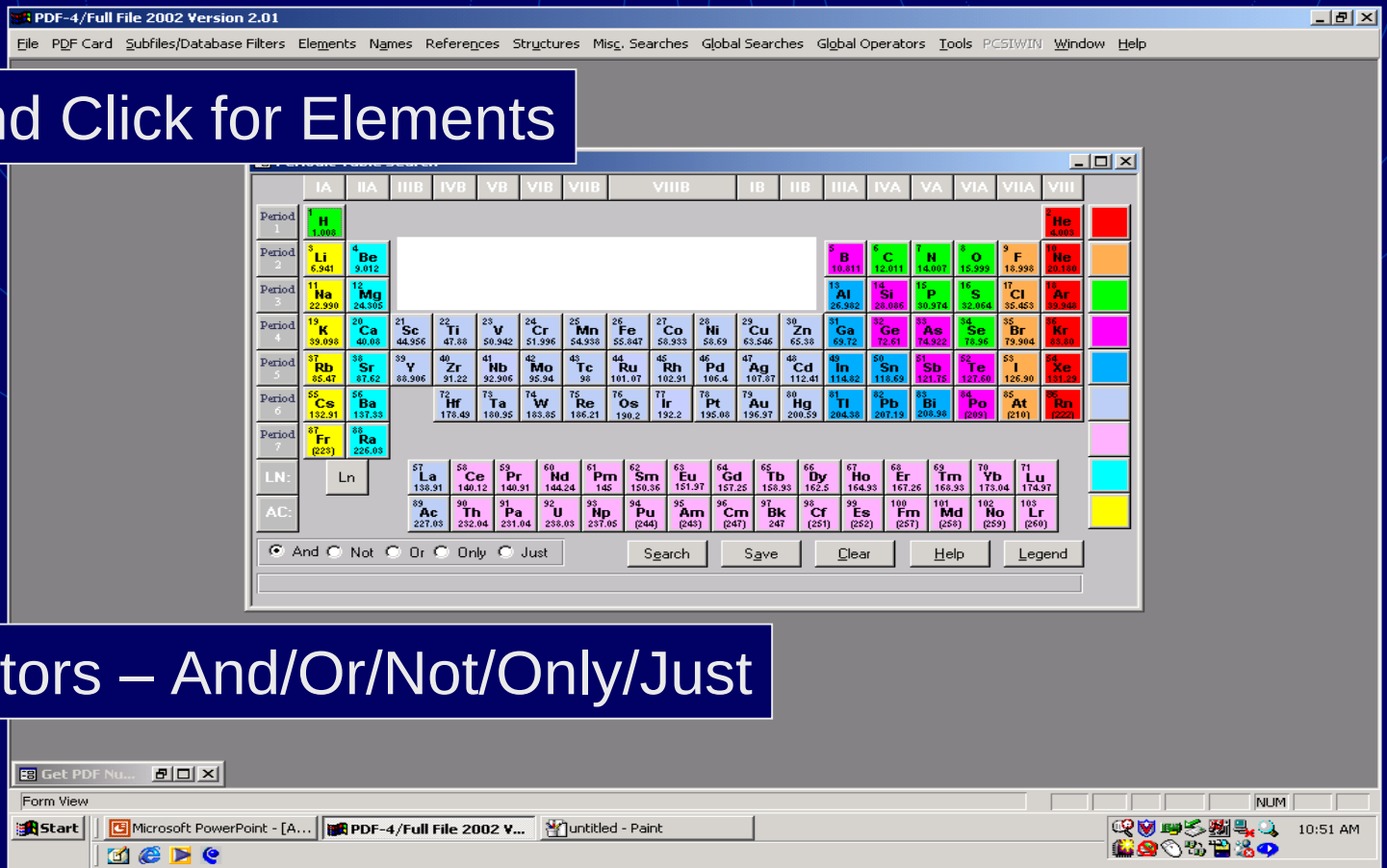




Example: Mining the Database Start with an Element Search

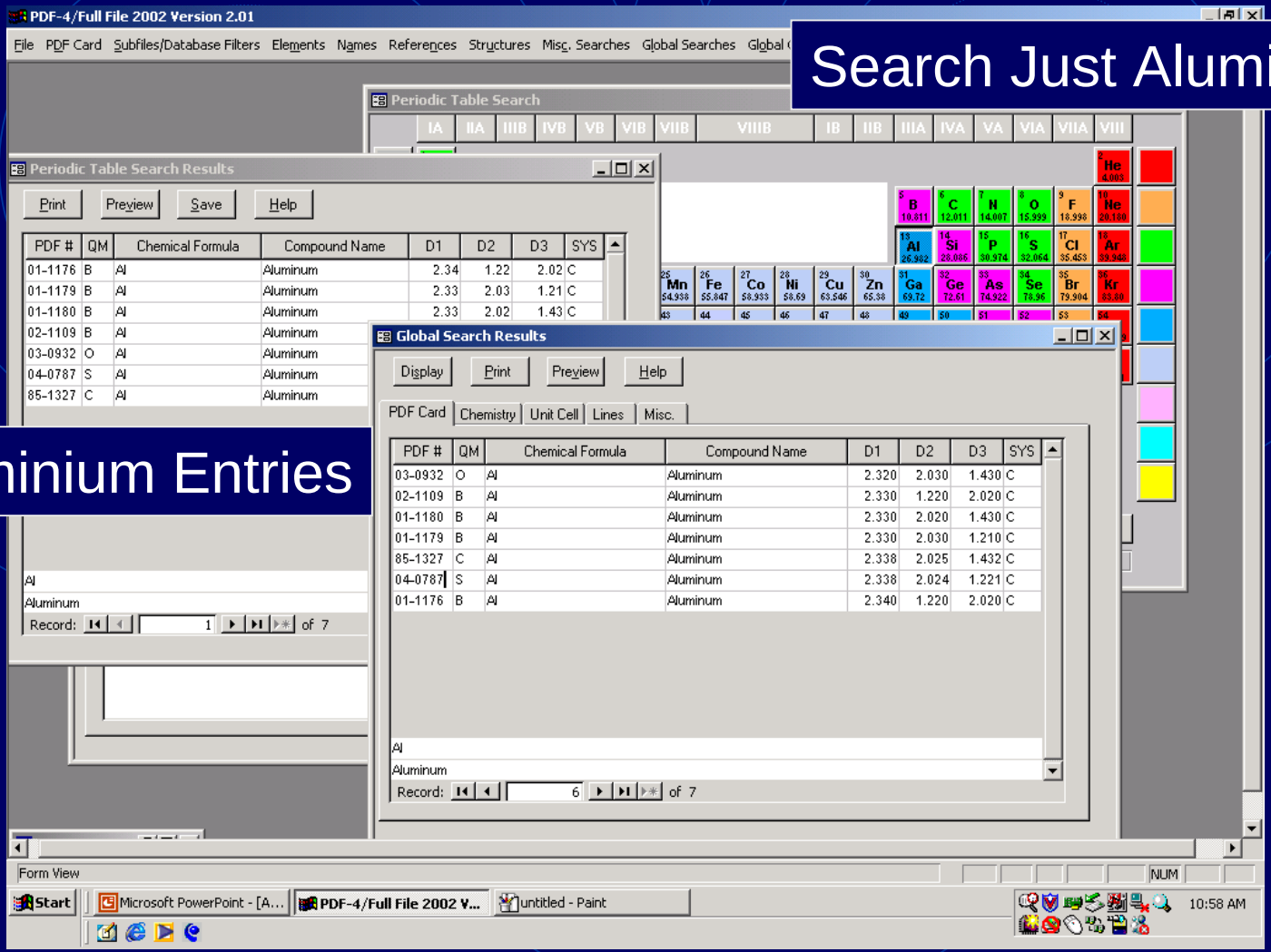


Point and Click for Elements



Operators – And/Or/Not/Only/Just

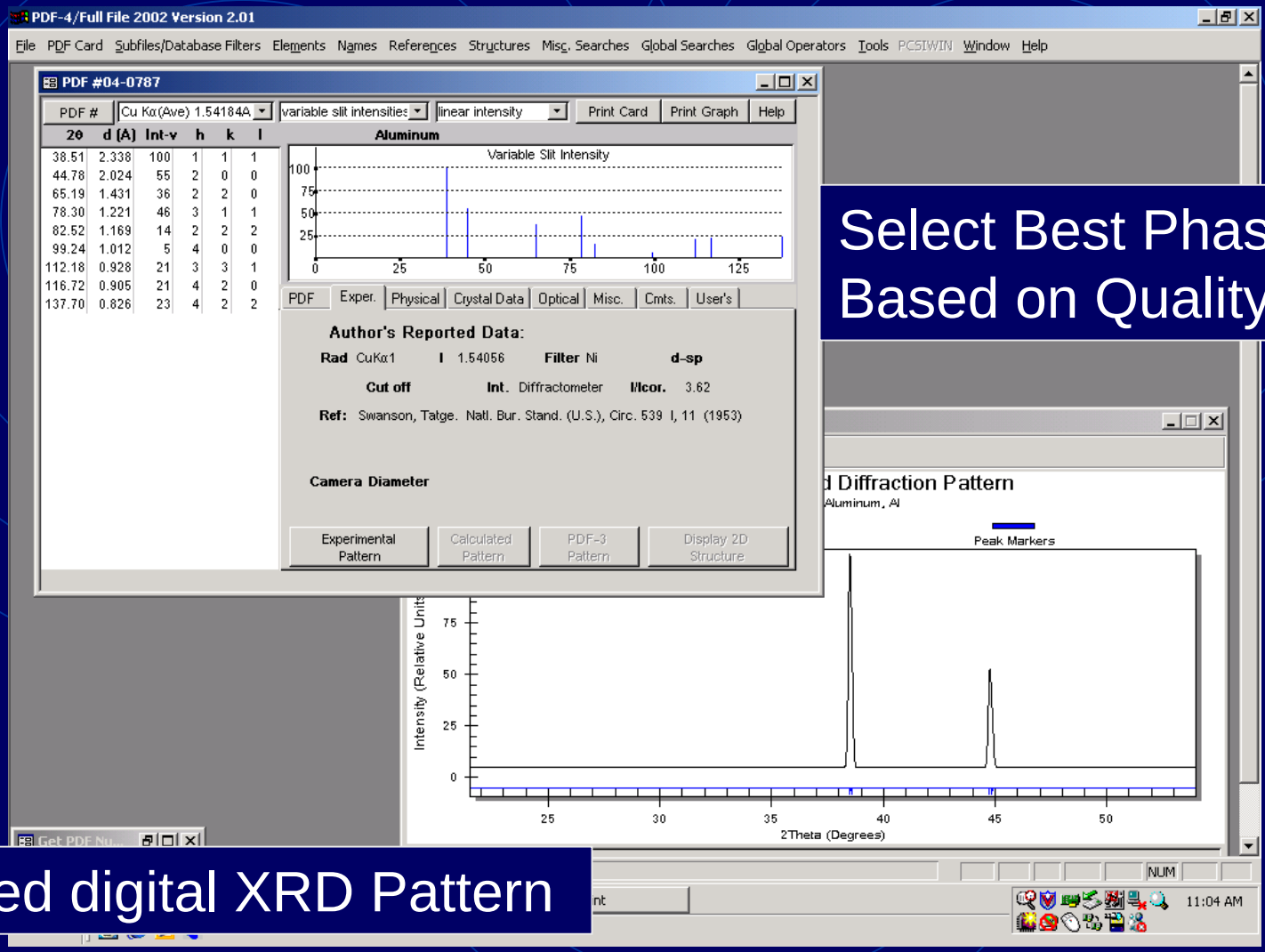
Mining the Database for Aluminium



Search Just Aluminium

7 Aluminium Entries

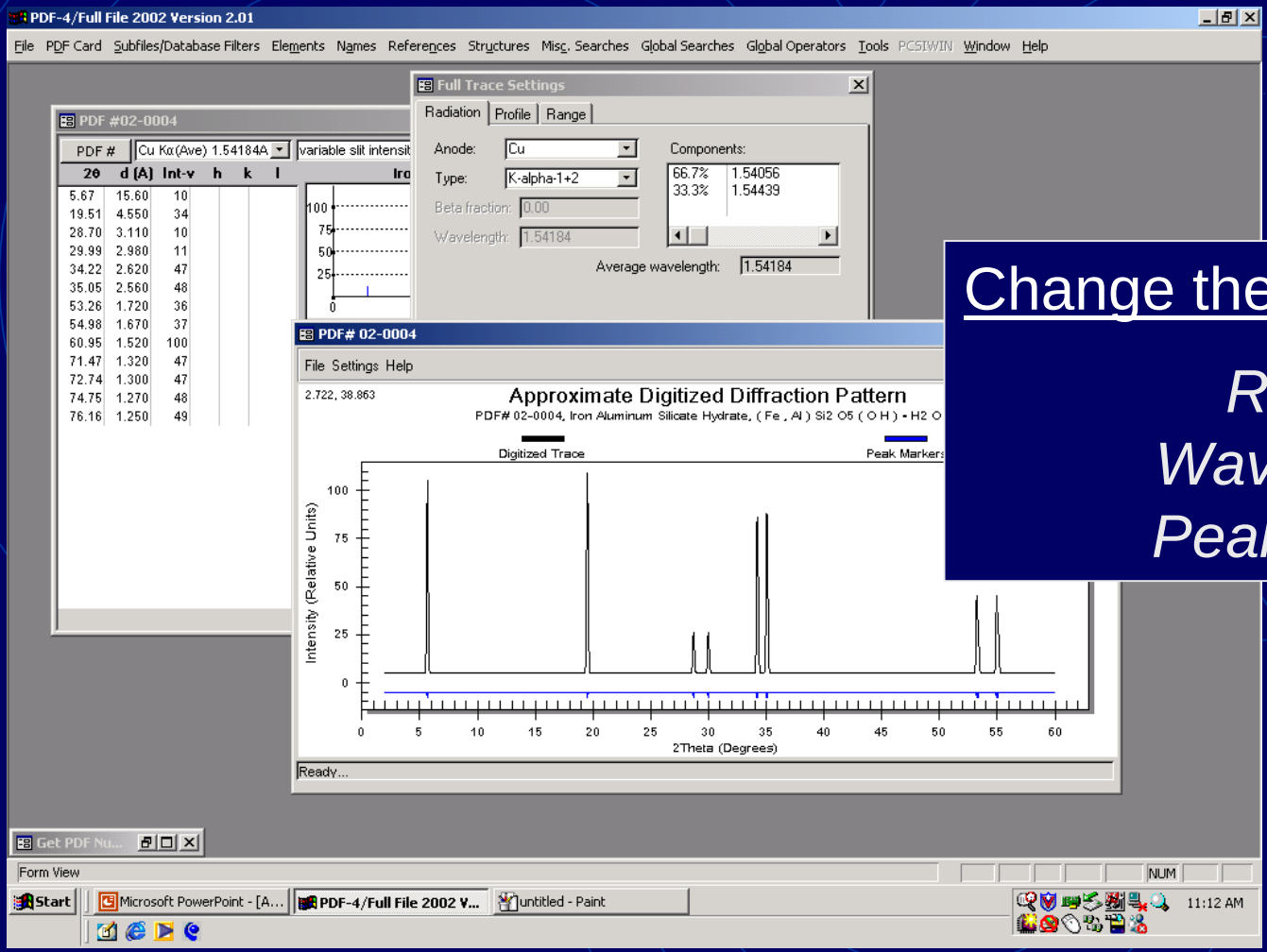
Mining the Database for Aluminium



Select Best Phase – Based on Quality Mark

Generated digital XRD Pattern

Mining the Database for Aluminium



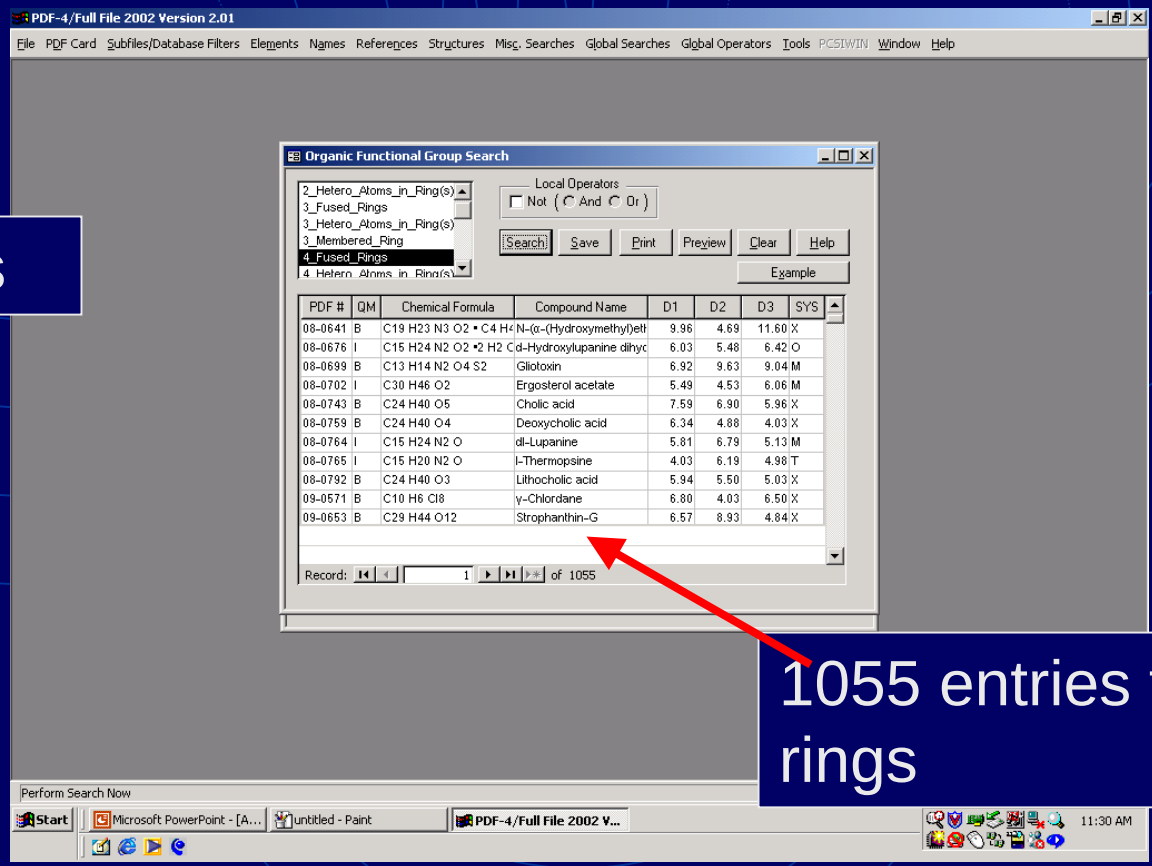
Change the XRD Pattern

Range
Wavelength
Peak Profile

PDF-4/Organic Functional Group Search

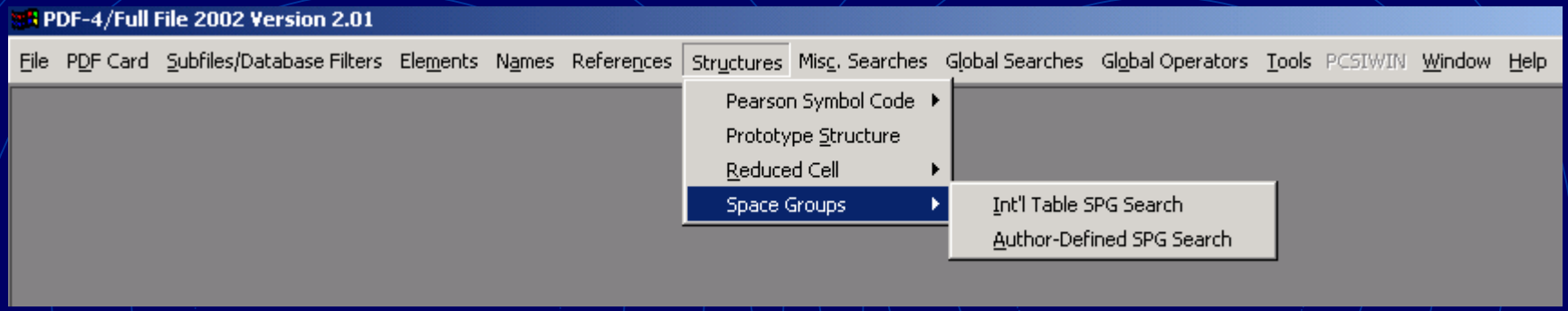
206 Functional Group Classifications

Fused Rings



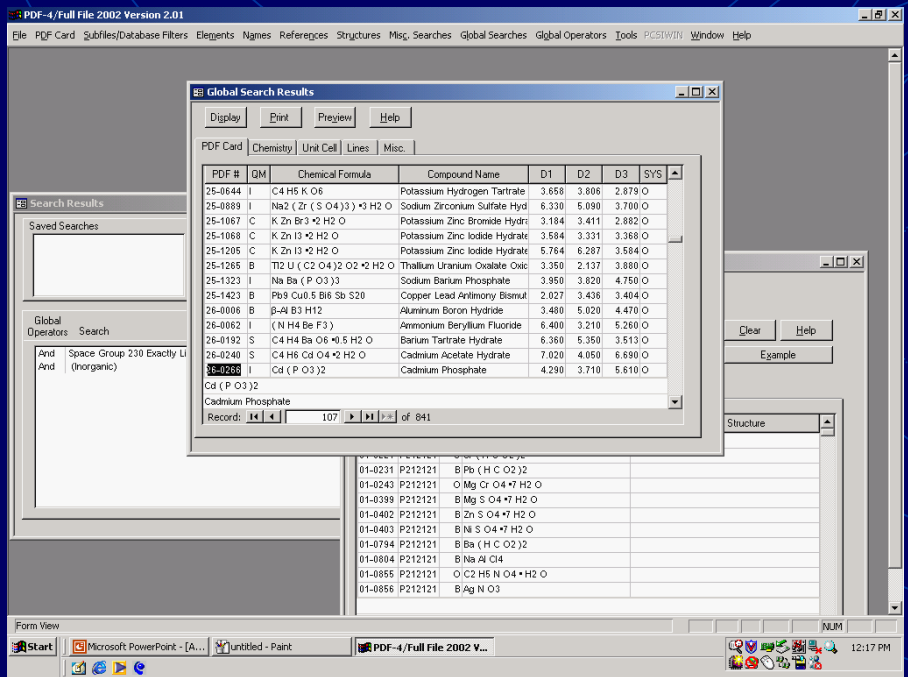
1055 entries for fused rings

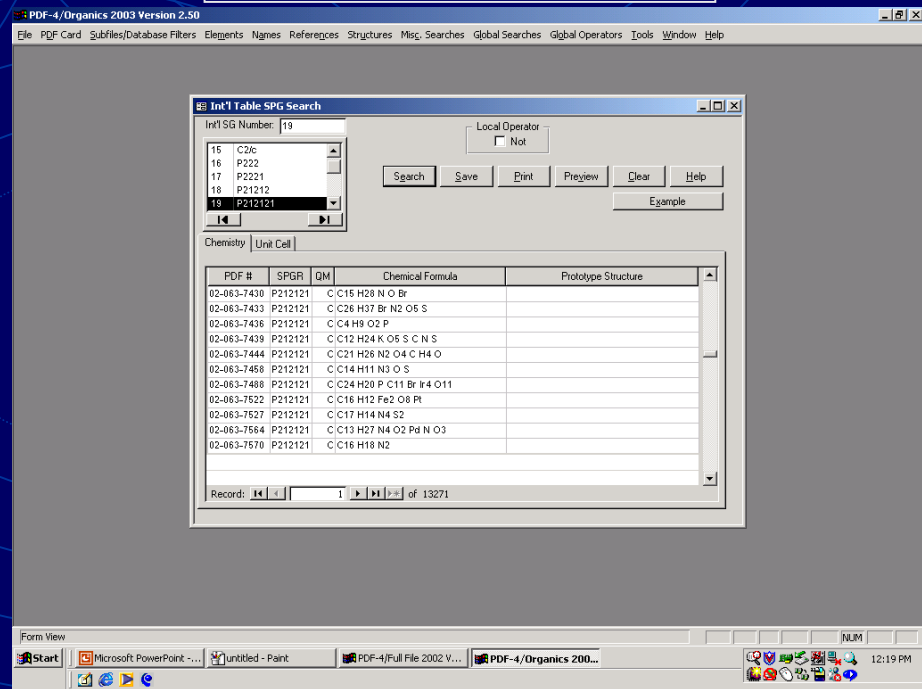
Space Group Search – $P2_12_12_1$ (19)



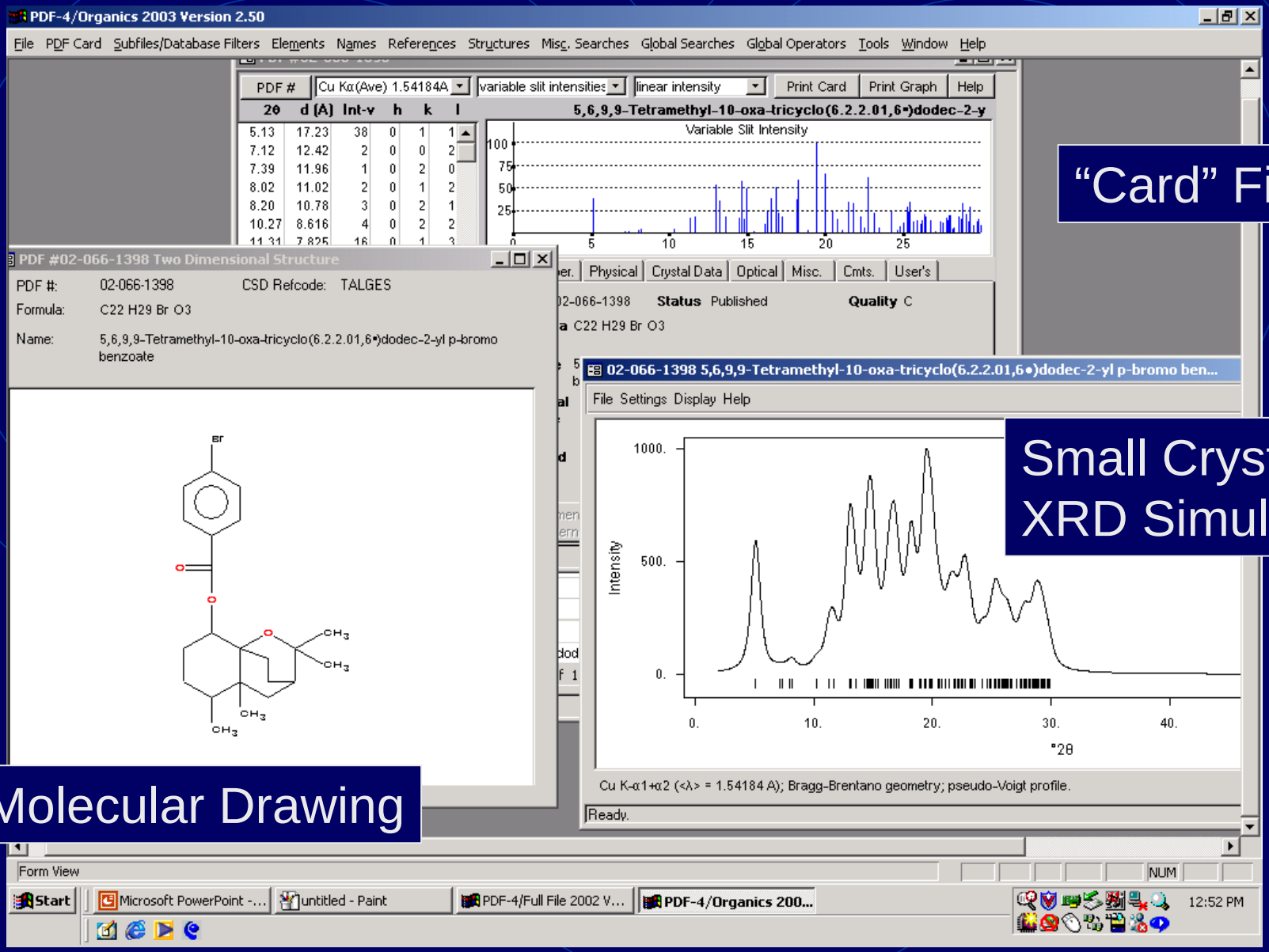
PDF-4/Full File
841 Inorganics

PDF-4/Organics
13,271 Organics





PDF-4 Organics – File Search

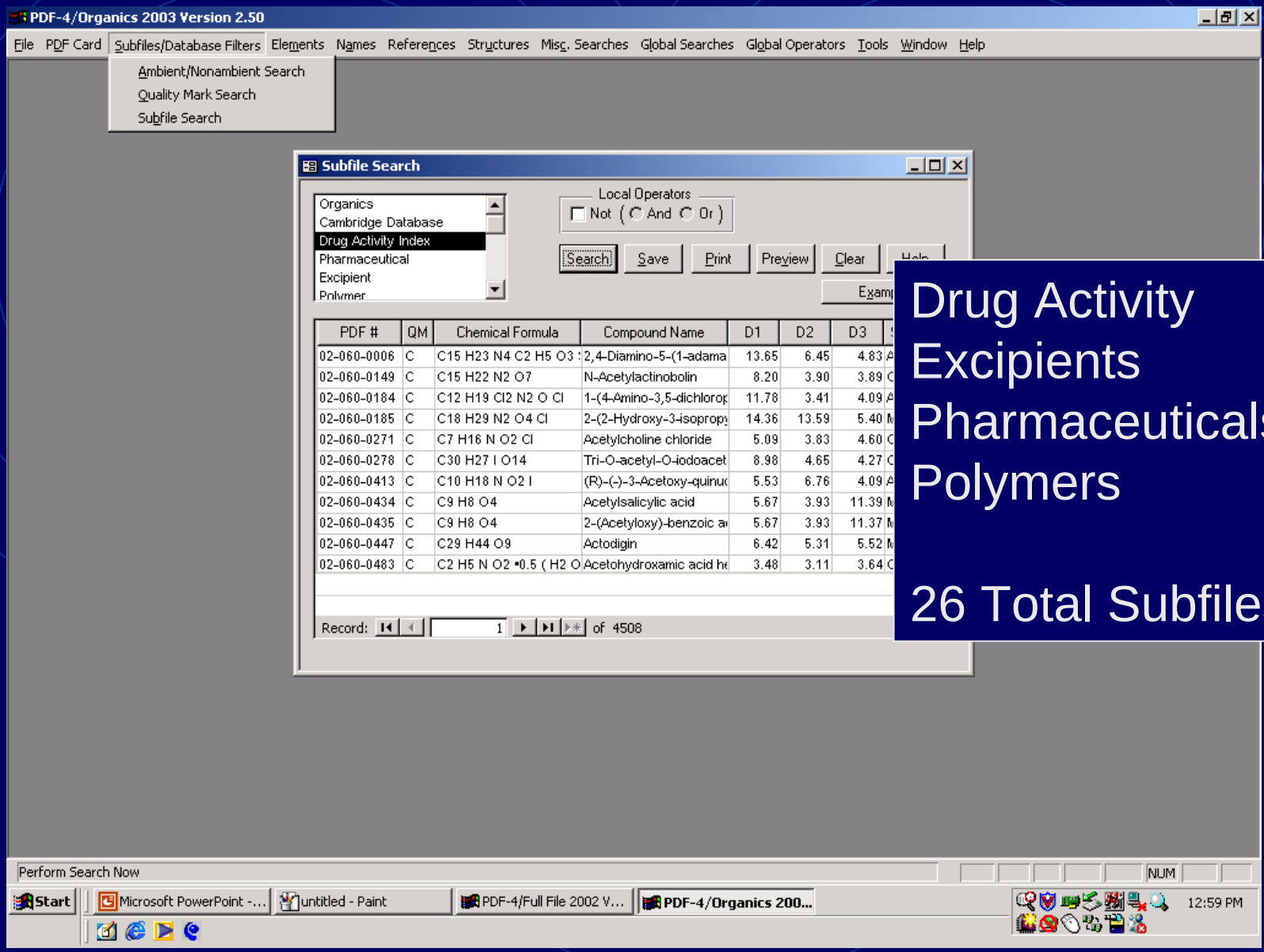


"Card" File

Small Crystallite XRD Simulation

Molecular Drawing

PDF-4 Organics – File Search



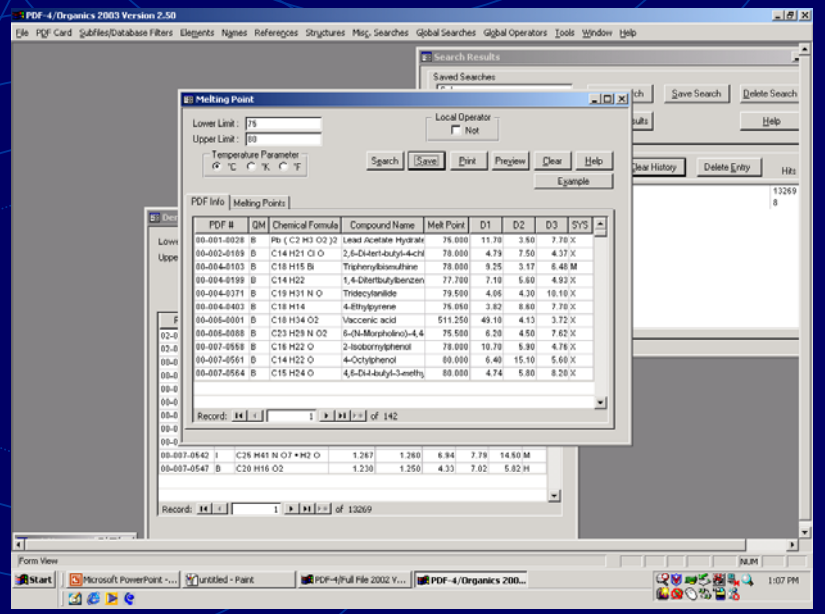
Data Mining – Global Searches

Organic Crystal Identification (1)

Pink crystal
Density = 1.25 – 1.30 g/cc
Melting point = 75 – 80 °C

PDF-4 Organics – 147,201 Entries

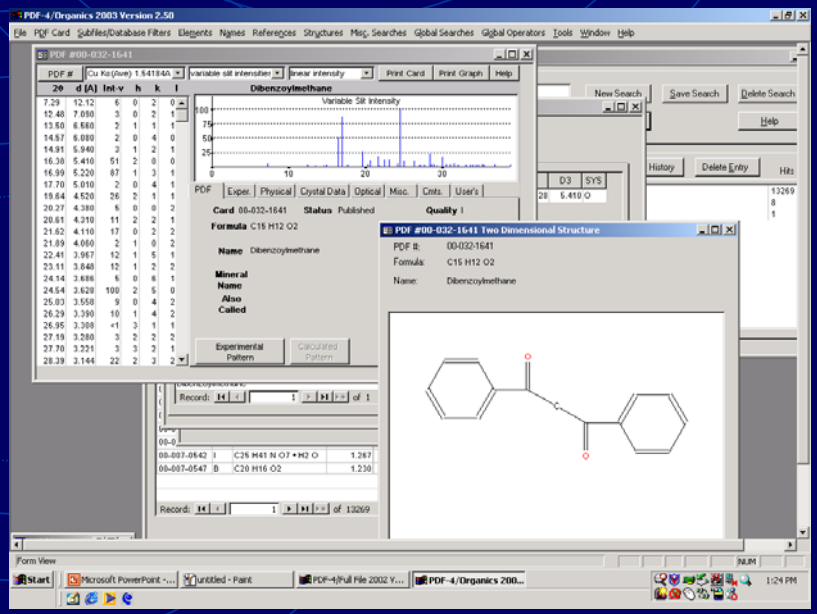
Density Window – 13,269 Materials



Organic Crystal Identification (2)

MP Window – reduces to 8 materials

Colour Window – reduces to 1 material
dibenzoxylmethane, $C_{15}H_{12}O_2$



Search for Passive Solar Home Heating Materials

Step 1 – Melting point search (home comfort)

Step 2 – Eliminate photodegradable hydrocarbons: extended application life

Step 3 – Review search and eliminate potentially harmful materials

1. Element Search

↓

2. Melting Point Search

↓

Produces a list of 17 candidates,
8 contain heavy metals,
1 is an explosive

All 8 candidate materials are referenced in the patent literature for passive solar heating applications

Demonstration