



MMS

Merged Markush Service

MMS on UNIX

Workshop Manual

Merged Markush Service



MMS

Combining the patent office expertise of

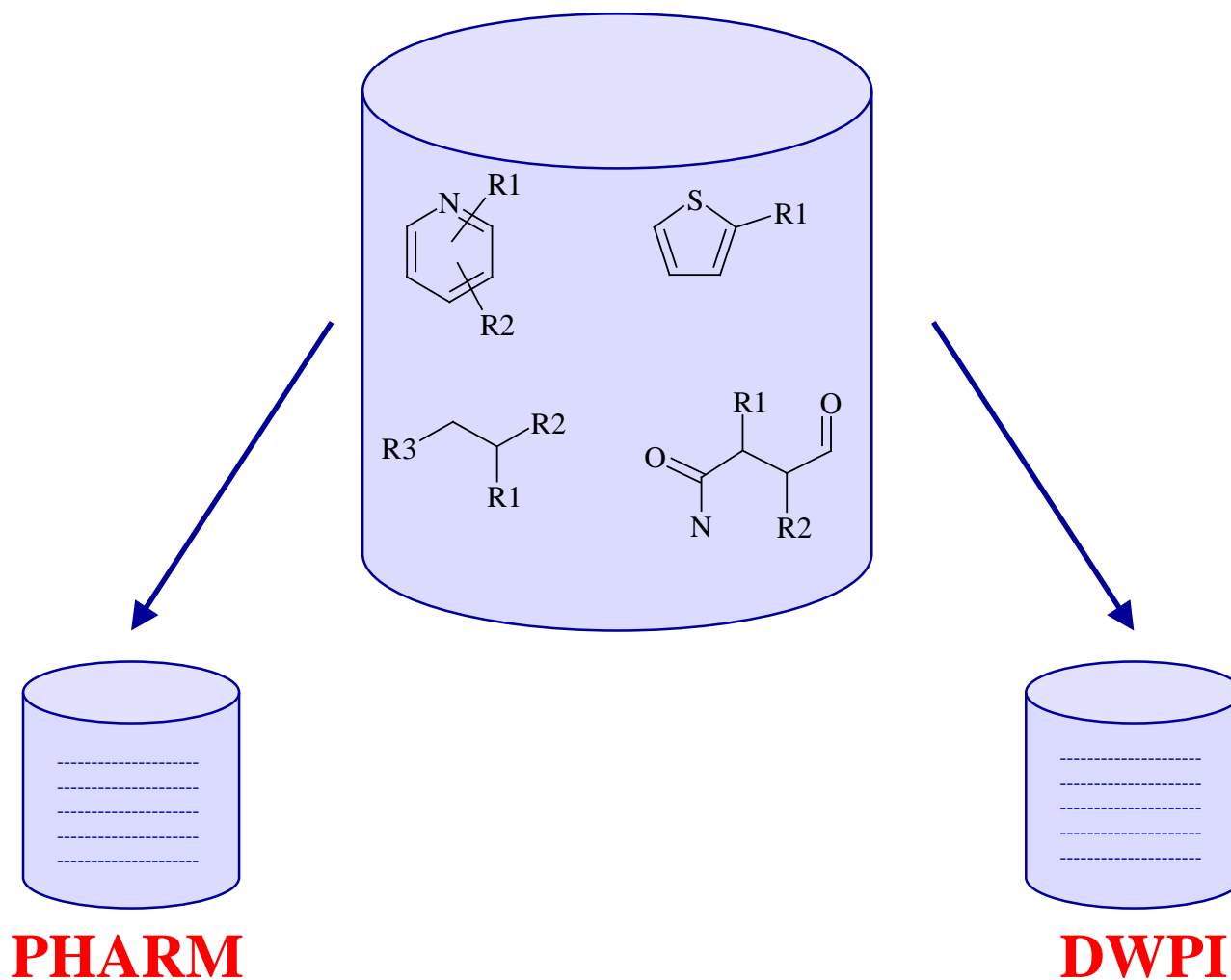


with the indexing expertise of

THOMSON

— TM

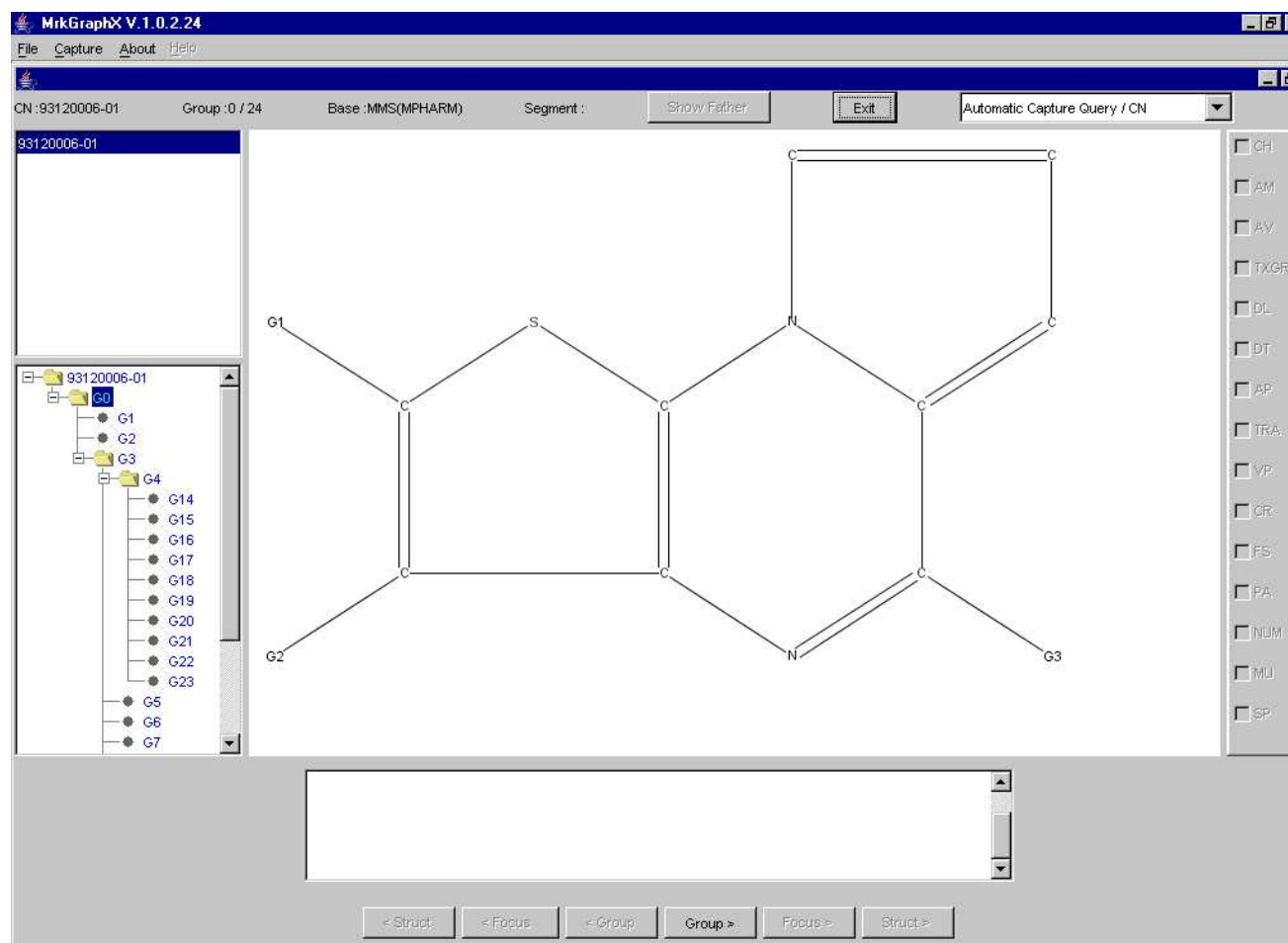
MERGED MARKUSH SERVICE



Merged Markush Service



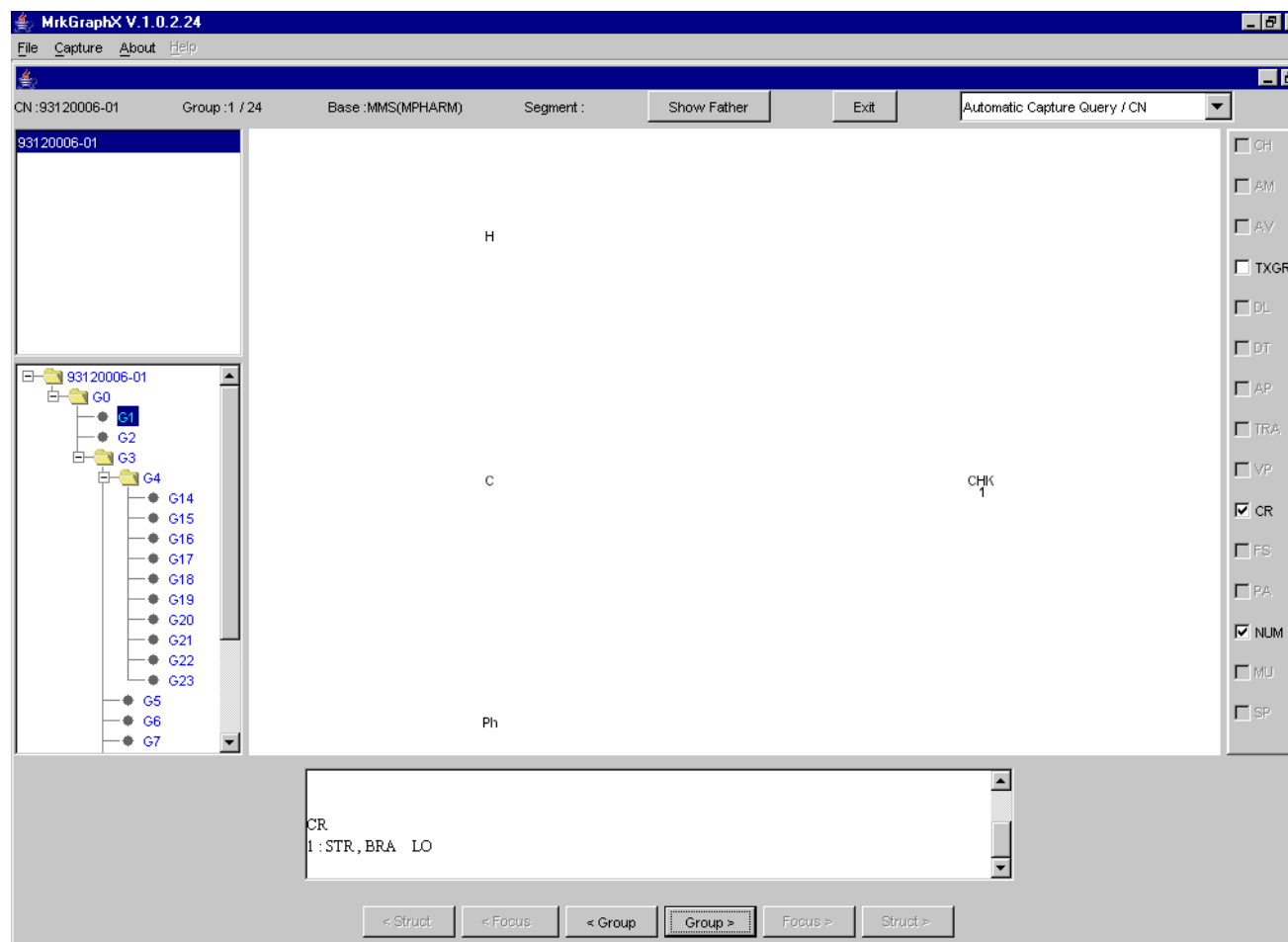
Approx. 2,600,000 structure records



Merged Markush Service



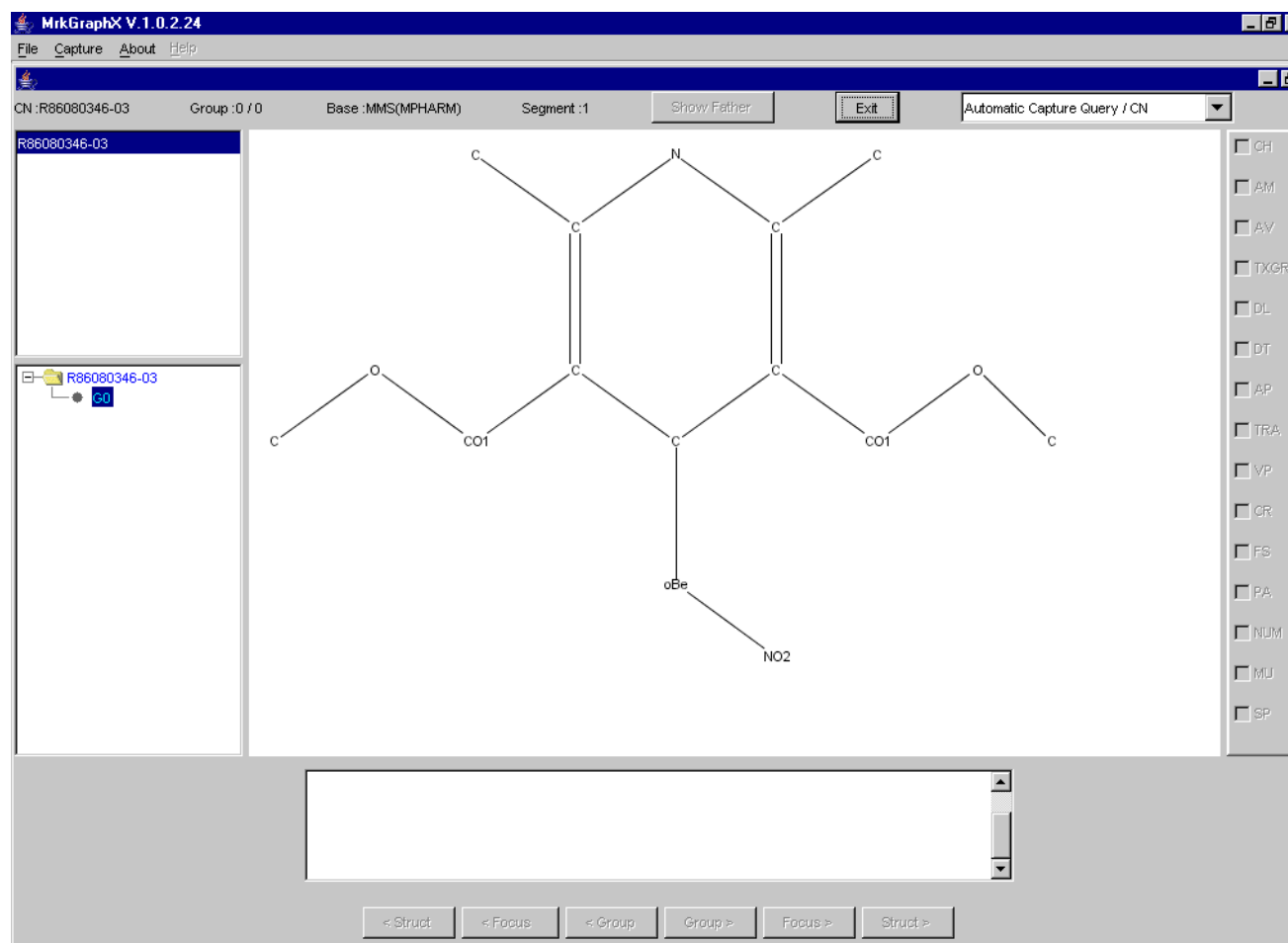
Approx. 2,600,000 structure records



Merged Markush Service



Approx. 2,600,000 structure records



Merged Markush Service



MMS contains:

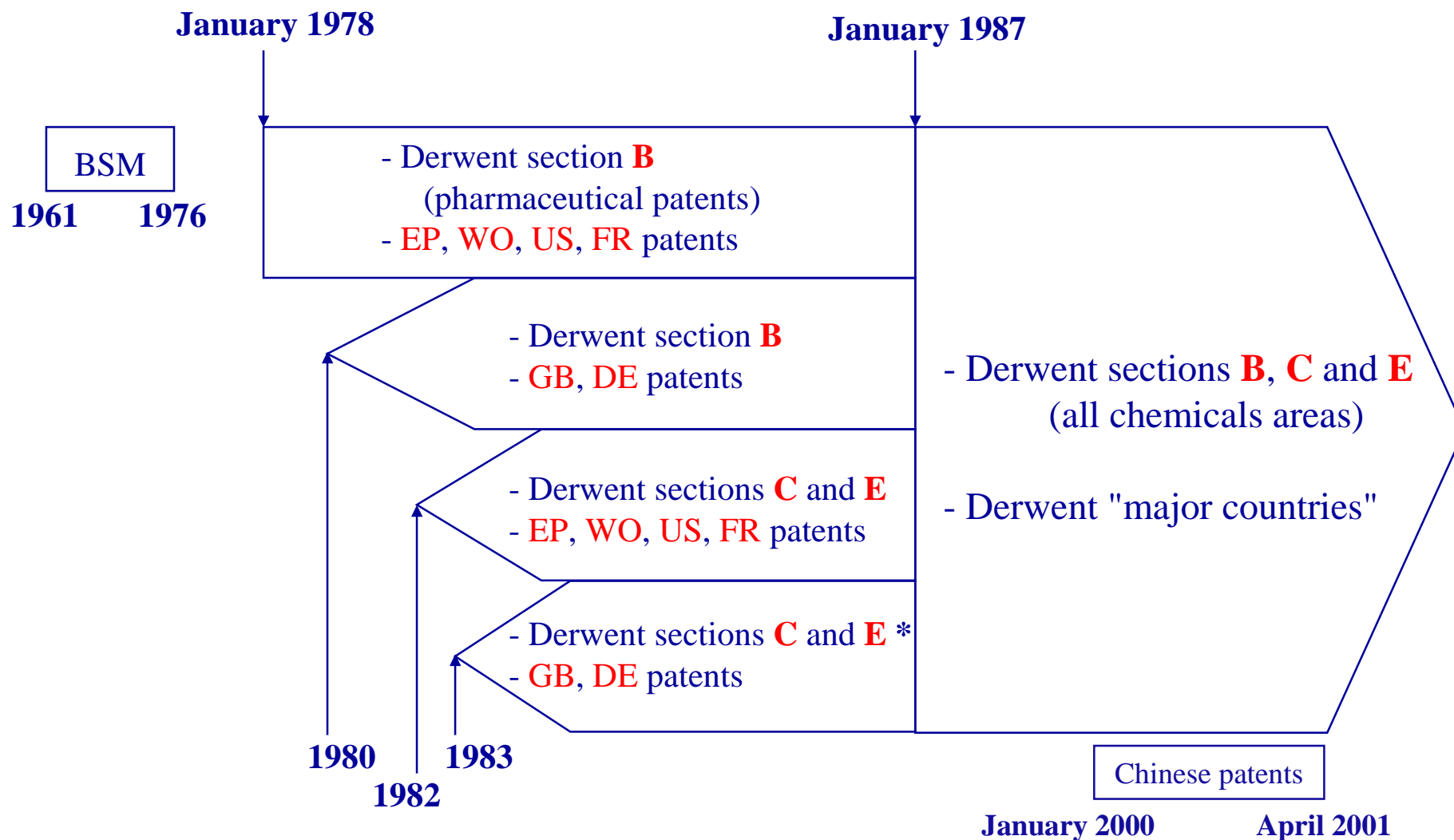
- \approx 60% Markush structures
- \approx 40% single compounds

MMS overview



- Content of MMS
 - Subject, period and country coverage
 - Structure coverage
 - Depth of indexing
- Indexing policy
 - Markush structures
 - Indexing tools
 - Structure representation conventions
- Structure of the MMS file

Content of MMS



* For sections C and E, DE and GB patents from February 1985 to June 1986 are not included.

Structure coverage

- Compounds described as new
- Products of new processes, including materials purified in new ways
- Compounds which are removed*
- Compounds used to effect removal*
- Compounds that are analyzed or detected*
- Compounds used in analysis or detection*
- Catalysts that are new
- Important ingredients of new compositions
- Known compounds with new activities

** when this is important to the novelty of the invention*

Content of MMS



Depth of indexing:

- Claims (generic and specific terms)
- Examples
- Disclosure (listed compounds)

Indexing policy

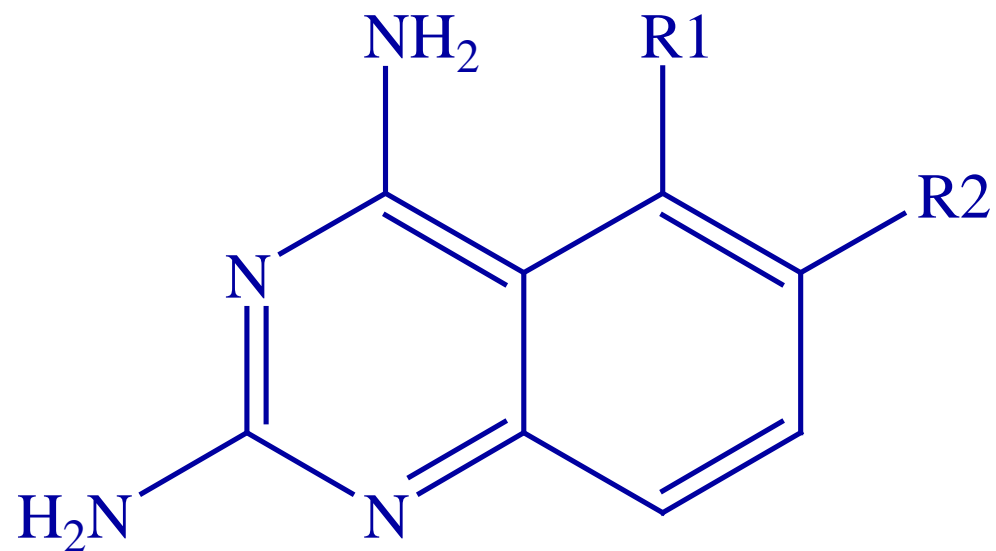


- **Markush structures**
 - Markush structure description
 - Markush structures in MMS
- **Indexing tools**
- **Structure representation conventions**

Indexing policy

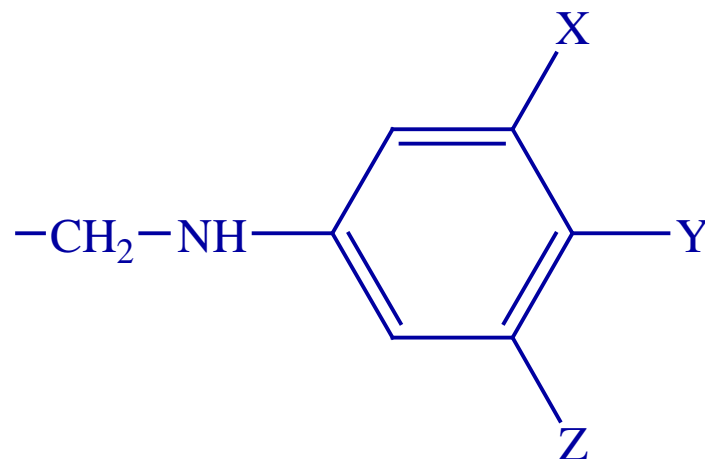


Markush structure description



Markush structure description

- **R1** is methyl, ethyl or a lower alkyl group
- **R2** is cyano, Br, Cl, halogen or a group of formula:

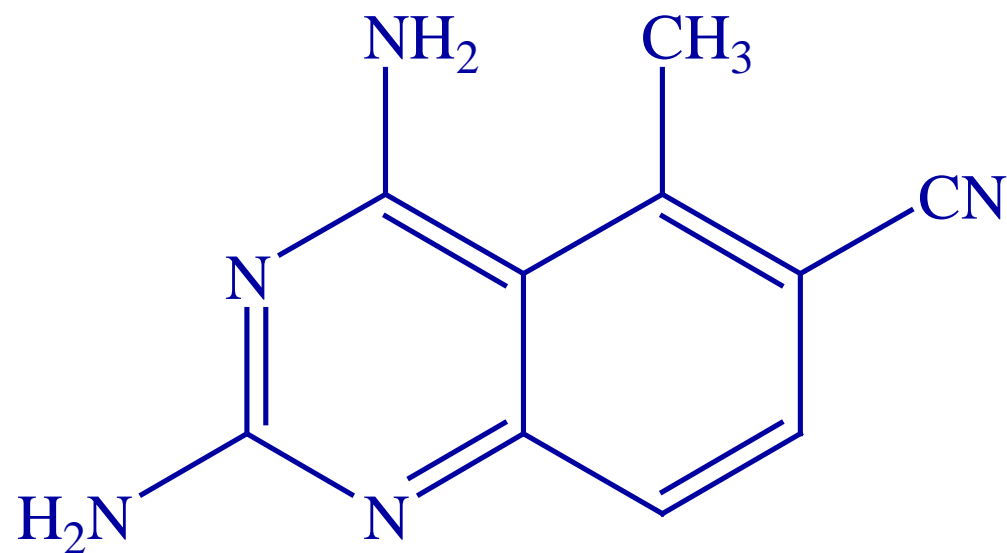


X, Y and Z: H, Br, Cl, halogen, $-\text{O}-\text{CH}_3$, alkoxy or $-\text{CF}_3$

Indexing policy



Markush structure description



R1: methyl / **R2:** cyano

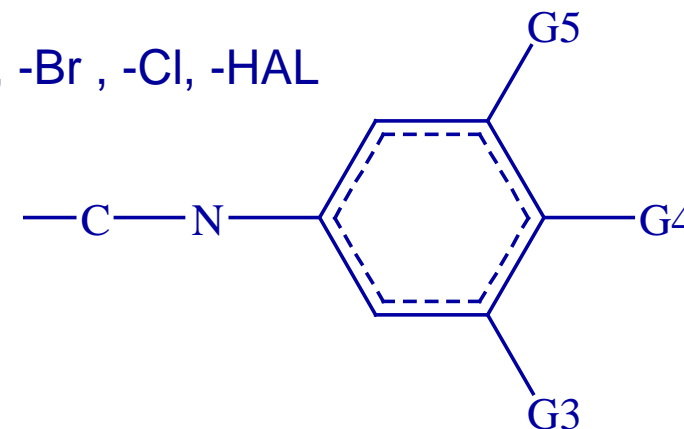
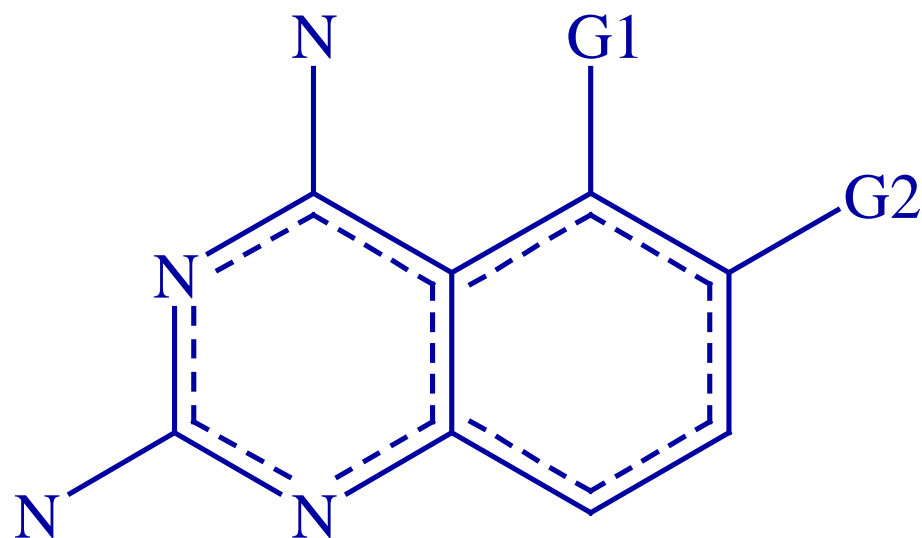
Indexing policy



Markush structures in MMS

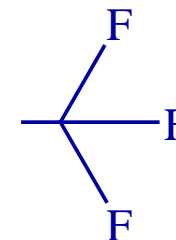
$G_1 = -C, -ET, -CHK^{LO}$

$G_2 = -CN, -Br, -Cl, -HAL$



$G_3 = G_4 = G_5 = -H, -Br, -Cl, HAL,$

$-O-C, -O-CHK^{LO}$



Indexing tools

- Atoms
- Shortcuts
- Superatoms
 - Superatom attributes
 - Text notes
- G groups

- **Atoms:**

All atoms from the Mendeleiev periodic table

- **Shortcuts:**

- Carbon chains: ET, NPR, IPR, NBU, SBU, TBU, IBU, Cn (C1, C2 ...)
- Benzene ring: PH, OBE, MBE, PBE
- Functions: CO1, CO2, SO2, SO3, PO3, PO4, CN, NO2, ACE

Indexing tools: Superatoms

- **Acyclic hydrocarbons:**

CHK Alkyl or alkylene

CHE Alkenyl or alkenylene

CHY Alkynyl or alkynylene

- **Carbocyclic hydrocarbons:**

ARY Carbocyclic system, optionally fused, containing at least one benzene ring (aryl)

CYC Cycloaliphatic, optionally fused, non-aromatic carbocycle

Indexing tools: Superatoms

- **Heterocyclic systems:**

HEA Monocyclic aromatic heterocycle (heteroaryl)

HET Non-aromatic monocyclic heterocycle

HEF Fused heterocycle

Indexing tools: Superatoms

- **Metals:**

MX Any metal

AMX Alkali or alkaline earth metals

A35 Group IIIA - VA metals

TRM Transition metals (excluding lanthanum)

LAN Lanthanides (including lanthanum)

ACT Actinides

Indexing tools: Superatoms

- **Other:**

HAL Halogens

ACY Acyls

PRT Protecting group

POL Polymer or polypeptide residue

DYE Dye group residue

XX Any atom or group excluding hydrogen (display only)

Indexing tools: Superatom attributes

- **Acyclic hydrocarbon attributes:**

- Chain length

LO	Low (1 to 6 carbon atoms)
MID	Middle (7 to 10 carbon atoms)
HI	High (11 or more carbon atoms)

- Chain type

STR	Straight
BRA	Branched

Indexing policy



Indexing tools: Superatom attributes

- **Cyclic system attributes:**

- Ring type

MON Monocyclic

FU Fused

- Degree of saturation

SAT Fully saturated

UNS Unsaturated

Indexing tools: Text notes

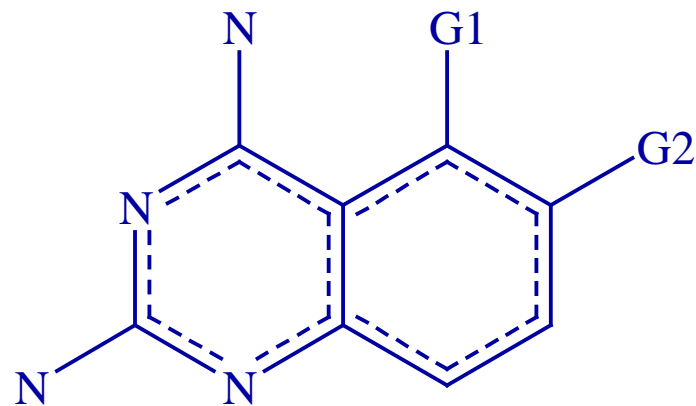
- **Number of carbon atoms:** CHK, CHE, CHY, CYC
- **Number of double bonds:** CHE, CHY
- **Number of triple bonds:** CHY
- **Number of rings:** ARY, CYC, HEF
- **Number of ring members:** ARY, CYC, HEA, HET, HEF
- **Number of specific heteroatoms:** HEA, HET, HEF
- **Atom of attachment:** HEA, HET, HEF, CYC
- **Number of repetitions for repeating units**

Indexing policy



G groups

Fixed part



Variable part

G1 = -C ; -ET ; -CHK^{LO}

50 G groups, 4 levels of nesting, 1023 atoms

Indexing policy



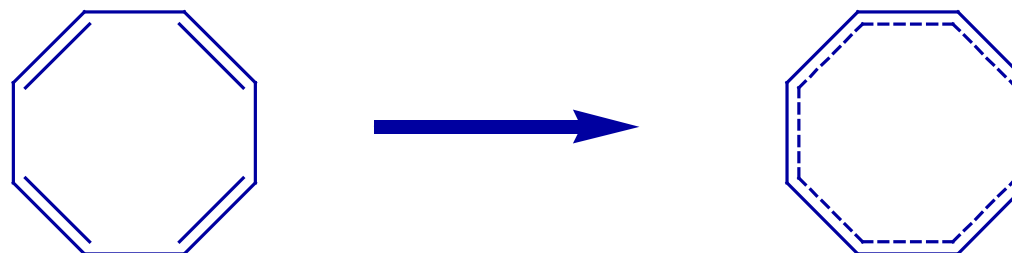
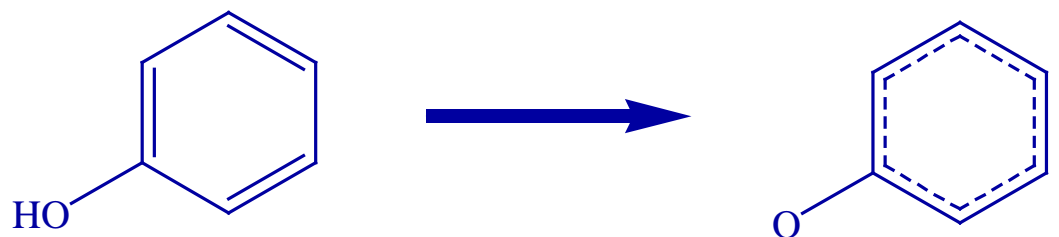
Structure representation conventions: bond normalization

- Aromaticity
- Tautomerism
- Enol-Keto tautomerism

Structure representation conventions

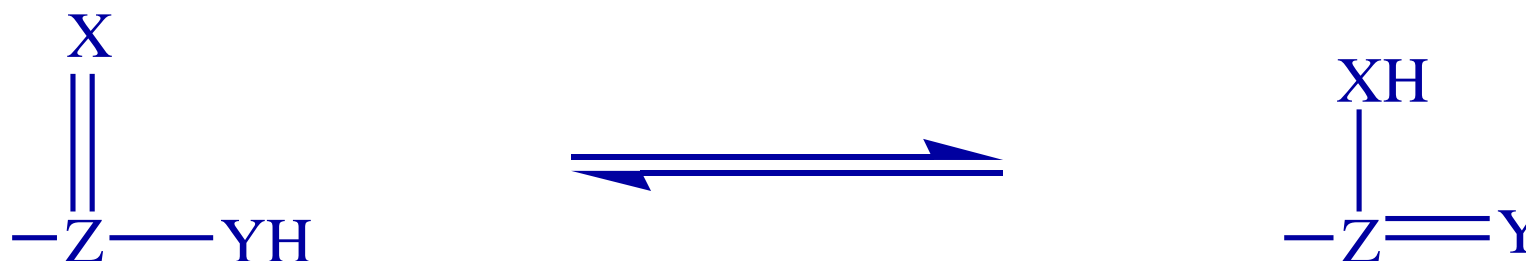
- **Aromaticity**

The bonds in rings containing $2n$ atoms with n alternating single and double bonds are normalized. This normalization takes priority over other rules such as tautomerisation.



Structure representation conventions

- **Tautomerism**



Where:

X, Y are O, S, Se, Te, N

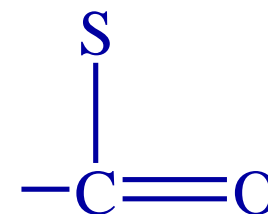
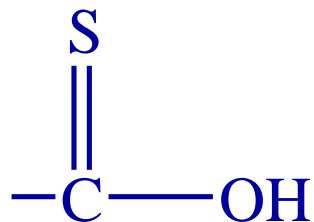
Z is B, C, Si, N, P, As, S, Se, Te, F, Cl, Br or I

Structure representation conventions

- **Tautomerism**

If X and Y are different, the double bond is placed preferentially on the bond to the atom which is first in the sequence:

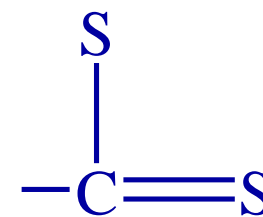
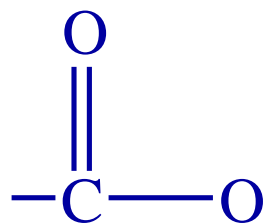
O, S, Se, Te, N



Structure representation conventions

- **Tautomerism**

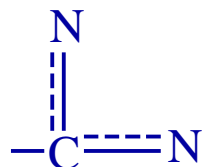
If X and Y are the same but not N, one double bond and one single bond are used.



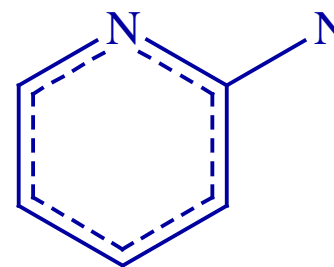
Structure representation conventions

- **Tautomerism**

If both X and Y are N, normalized bonds are used.



but



Note: This rule IS NOT applied to amino groups on normalized ring systems in which case the amino group is attached by a single bond

Structure representation conventions

- **Enol-Keto tautomerism**

If the heteroatom is O, S, Se, or Te, the keto form is preferentially indexed.



X = O, S, Se, Te

MMS overview



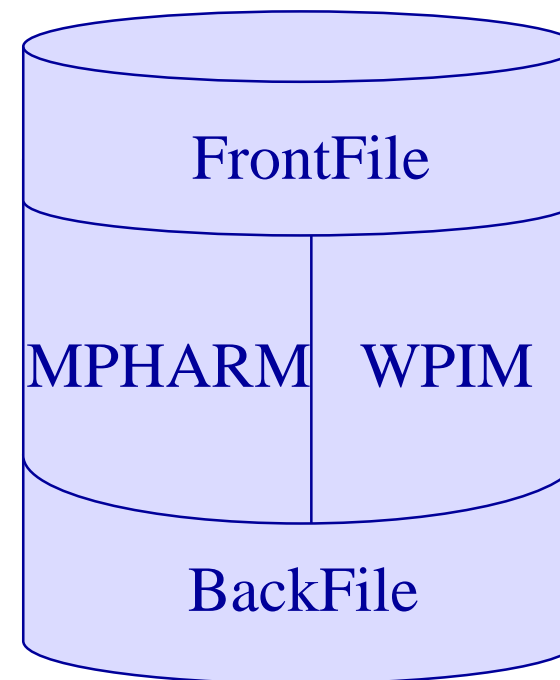
- Introduction
- Content of MMS
 - Subject, period and country coverage
 - Structure coverage
 - Depth of indexing
- Indexing policy
 - Markush structures
 - Indexing tools
 - Structure representation conventions
- Structure of the MMS file

Structure of the MMS file

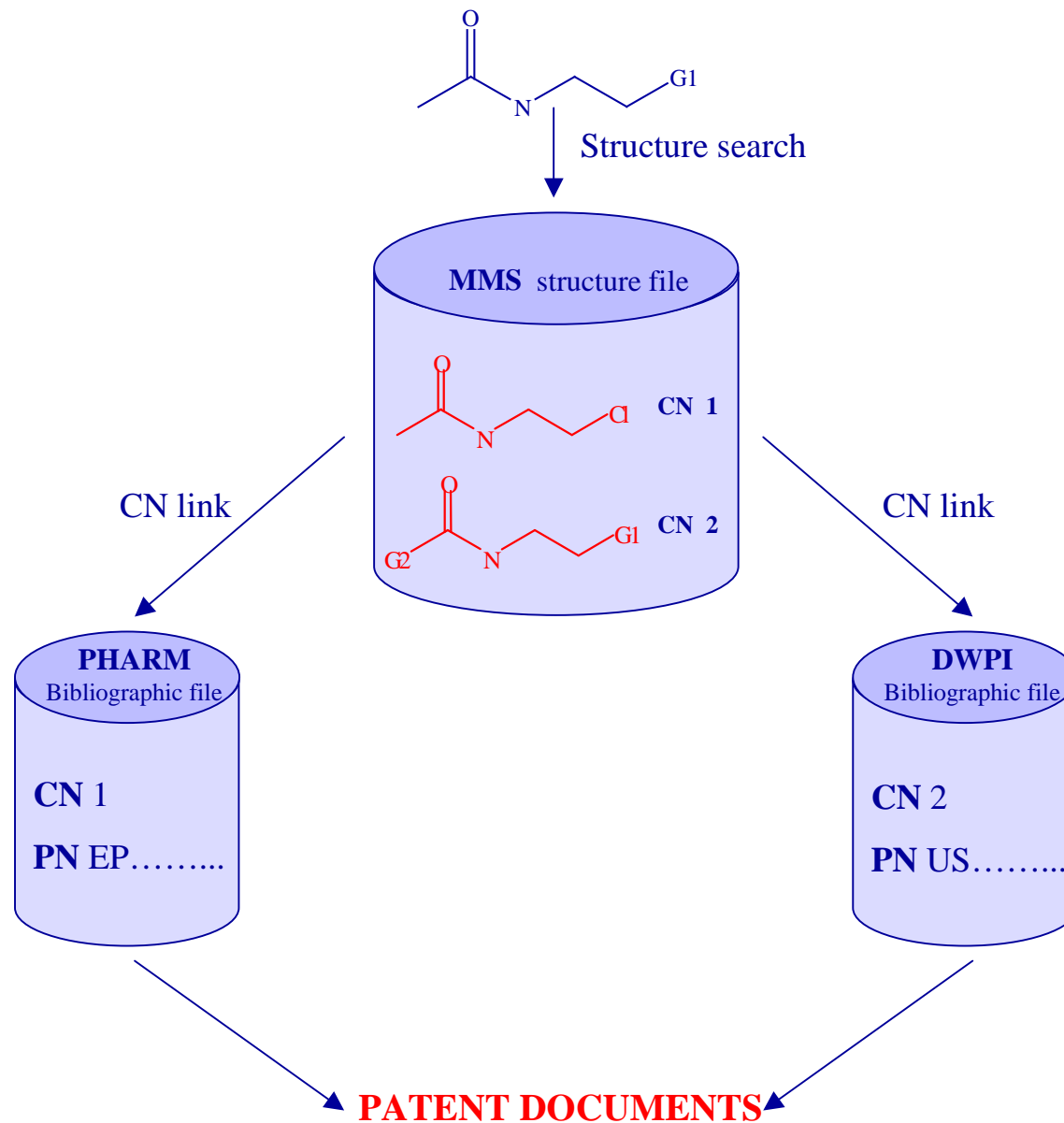


Four segments:

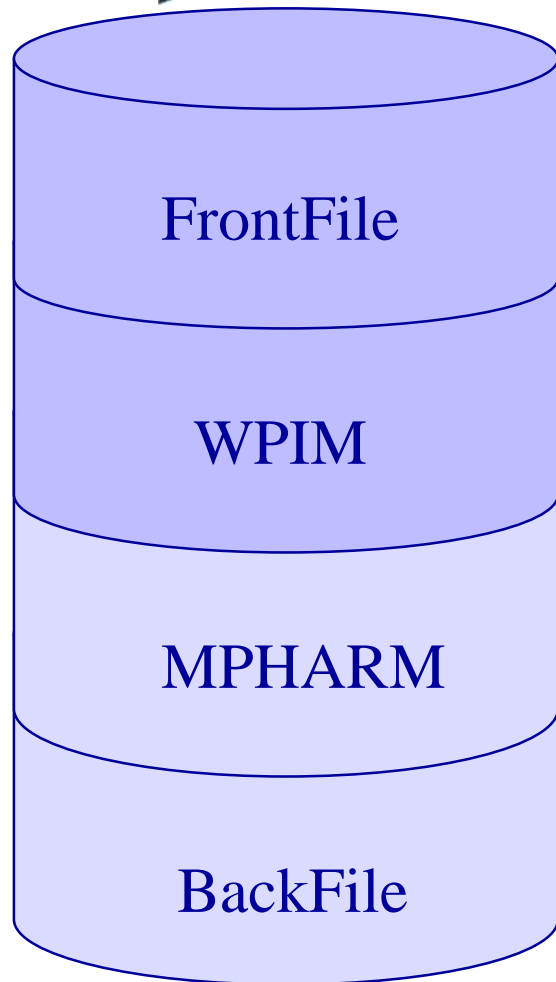
- FrontFile - from Derwent week 9816
- MPHARM – 1987-1999
- WPIM – 1987-1998 (week 15)
- BackFile – Before 1987



Link to bibliographic files



Link to bibliographic files

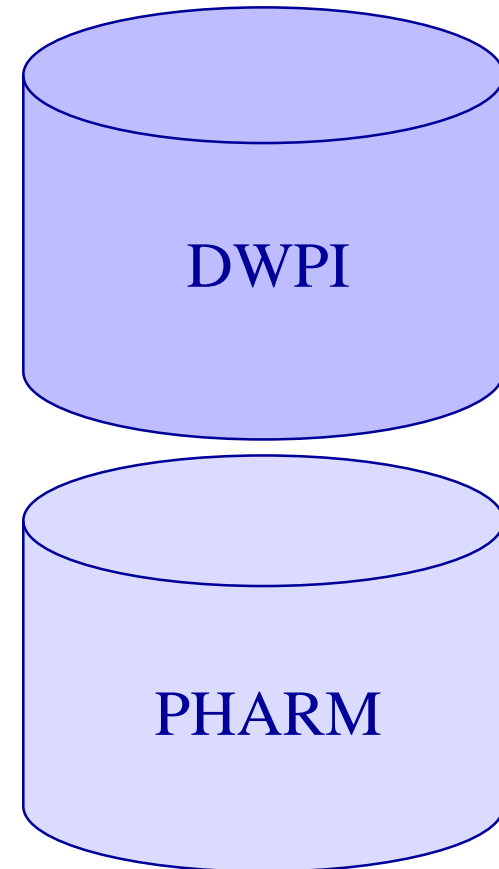


MMS STRUCTURE FILE

CN XXXX-XXXXX (generic)
or **RXXXXX (specific)**

4 CN formats

CN XXXXXXXXX-XX (generic)
or **RXXXXXXXXX-XX (specific)**



BIBLIOGRAPHIC FILES