



## What's Cooking for KNIME at Vernalis?

*Steve Roughley*

*s.roughley@vernalis.com*

*KNIME Cheminformatics SIG*

*Leeds, UK*

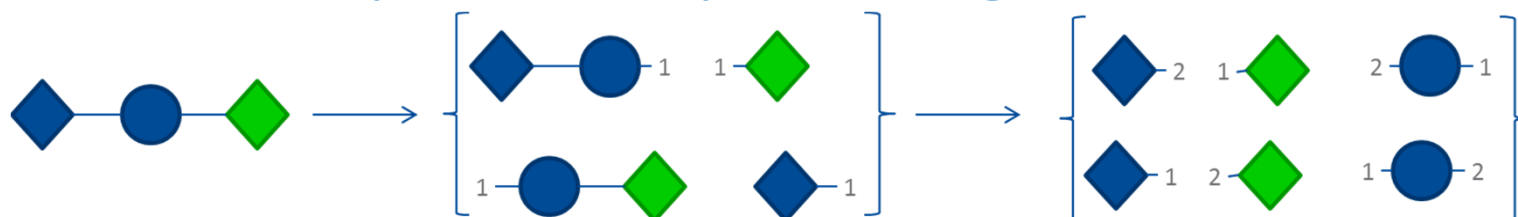
*20<sup>th</sup> May 2015*



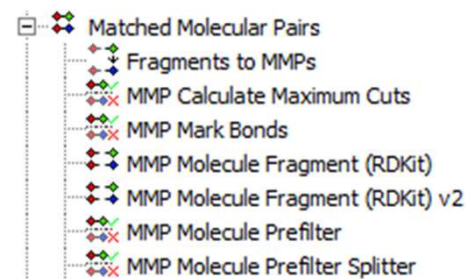
THE QUEEN'S AWARDS  
FOR ENTERPRISE:  
INTERNATIONAL TRADE  
2014

- Version 1.4.0
- New releases since last meeting:
  - Flow control nodes (delays, timed loops, more switches)
  - Read/Write flow variables
  - Matched Pairs
  - Performance Monitoring
- Now stands at 52 nodes in 3 plugins
- Further 79 nodes released internally (and more ongoing)
  - Permission to release ~25 fingerprint related nodes...
  - ...when I find the time to finish tidying and testing

- Working on improvements to the MMP nodes
  - Speed and memory use issues with current implementation
  - Duplication of output for multiple cuts, e.g:



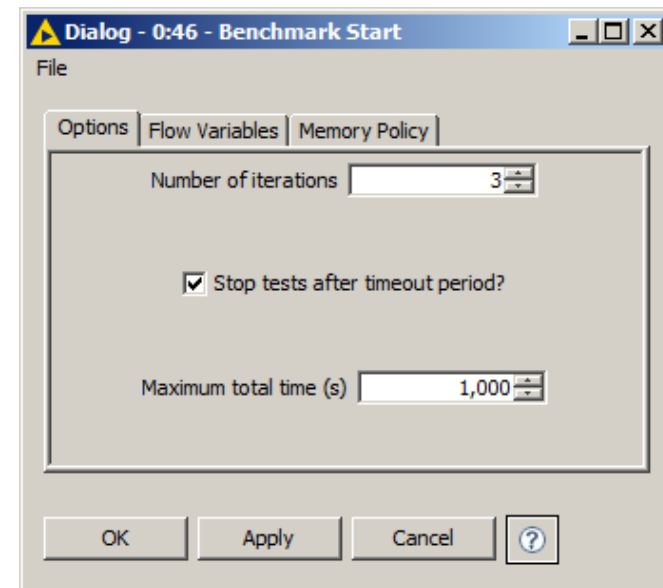
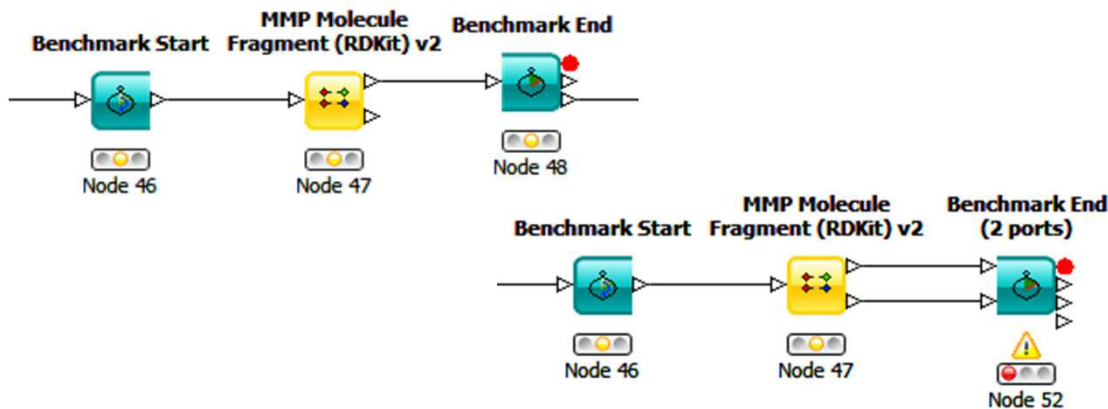
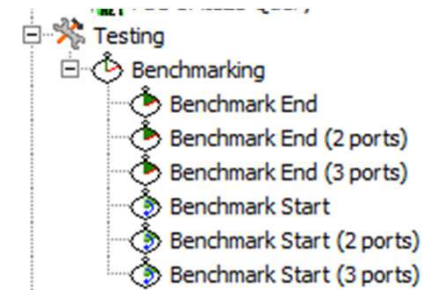
- Currently at ~35-fold speed increase for fragmentation
  - Duplication no longer an issue
  - Fewer SMILES  $\leftrightarrow$  RDKit conversions
- New nodes for pre-filtering, bond highlighting
- We will release in due course



# Benchmarking Nodes

New in v 1.4.0

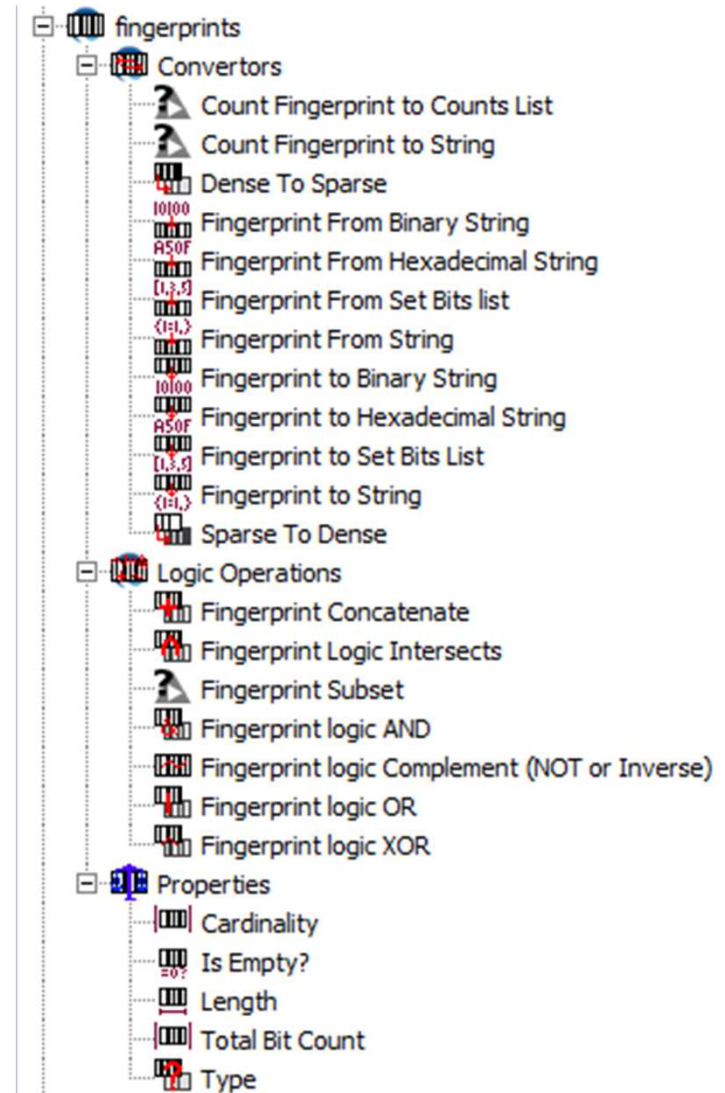
- Time execution of intervening node(s)
  - Multiple iterations
  - Optional time limit (next iteration doesn't start if already elapsed)
  - Min/Max/Mean iteration time
  - Individual iteration times (1<sup>st</sup> output table)
  - Last iteration table(s) passed through



# Fingerprint nodes

PENDING RELEASE

- BitVector and ByteVector
- Type /format conversions
- Logic operations
- Properties
  
- Also aggregators (GroupBy, Column combiner nodes)
  
- Some functionality released by KNIME at around time permission for release obtained
  - “De-duplicate”



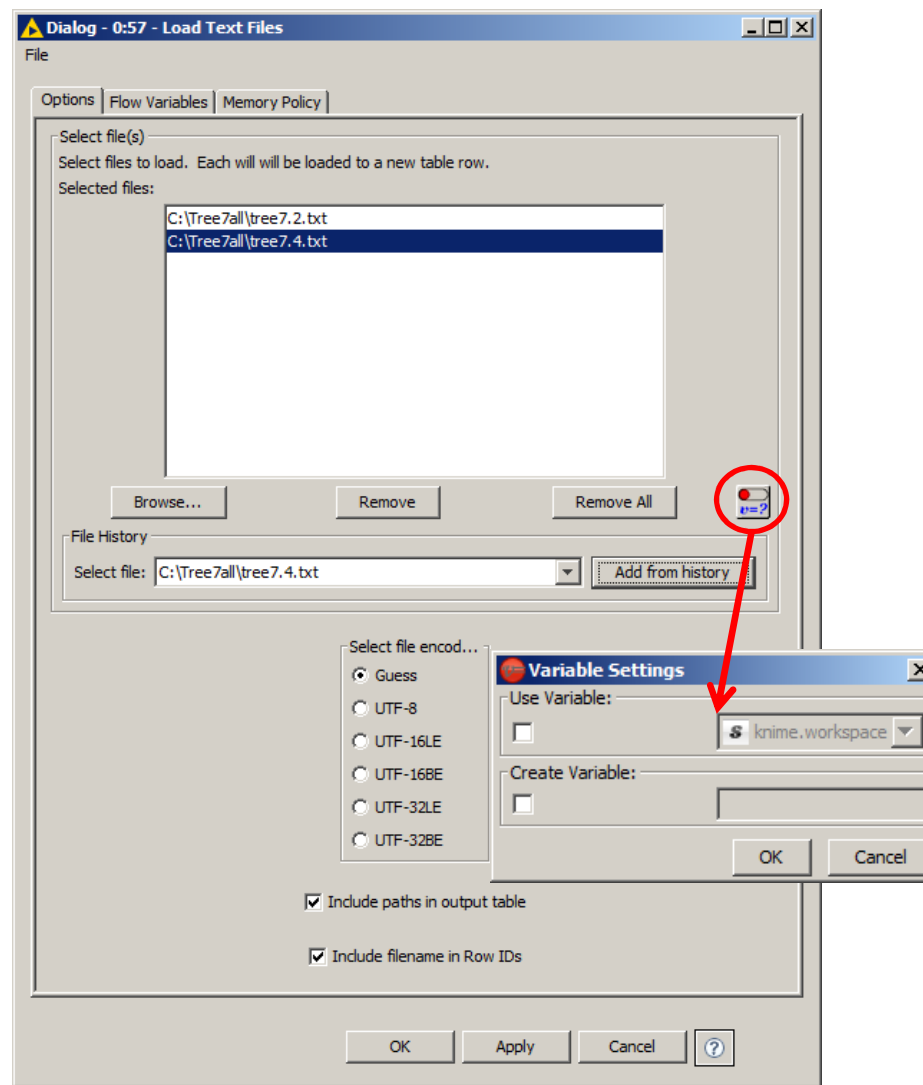
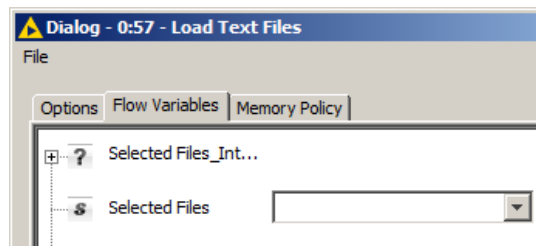
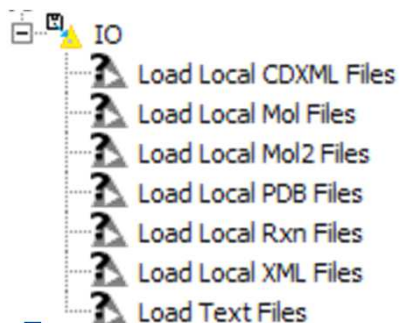
- “Speedy SMILES”
  - ~18 nodes for fast processing of SMILES
  - Filters/Splitters, Manipulators (e.g. invert stereochem, desalt), Property calculators (e.g. charges, HAC, chiral centres etc)
  - **FAST**: 1.4m molecules (ChEMBL 19)
    - Desalted - ~14 s
    - Split chiral/non-chiral molecules - ~11 s
    - Heavy Atom Count ~6 seconds
- RDKit co-ordinate manipulations
  - Rotate, Translate
- Fingerprints
  - Ertl Scaffold Keys, Pharmacophore triangles/tetrads (with ph4 cell type added)

# Text file loading source nodes

Load Text Files

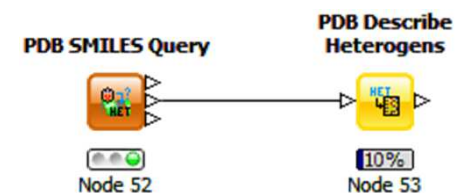
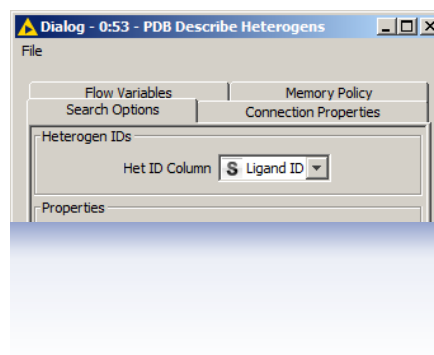
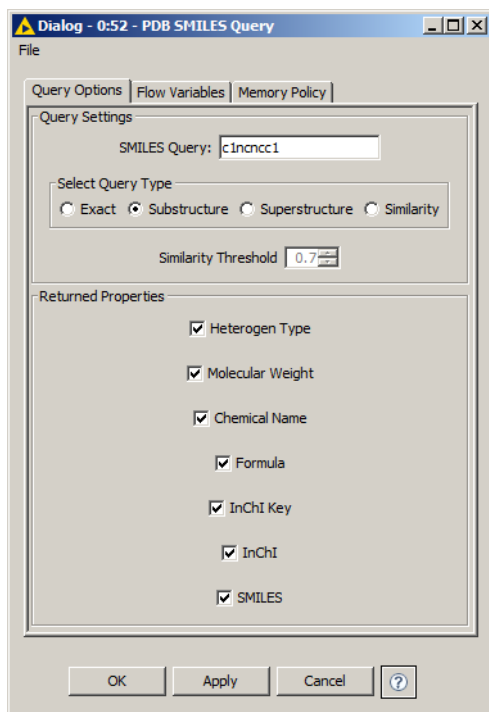


- Complements released Load text-based files manipulator node
- New DefaultDialog component and SettingsModel
  - StringArray, but with flow variable replacement
  - List of files can be passed via flow variable without the flow variables tab 'Array problem'
- Also available in other 'flavours':



# What else? (1)

- More PDB webservice queries
  - SMILES HET query (see <http://www.rcsb.org/pdb/software/rest.do#smiles>)
  - ‘Describe heterogens’ query (see <http://www.rcsb.org/pdb/software/rest.do#descComp>)



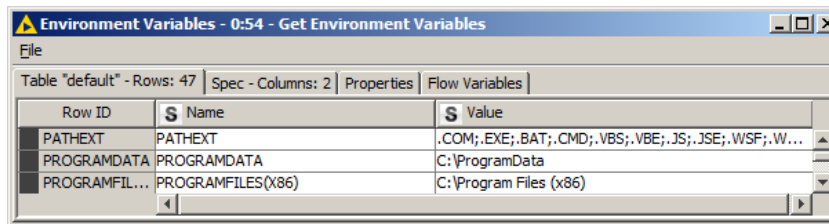


- Utility nodes
  - Garbage collector calls System.gc()

### Get Environment Variables



Node 54



Row ID	Name	Value
PATHEXT	PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.W...
PROGRAMDATA	PROGRAMDATA	C:\ProgramData
PROGRAMFIL...	PROGRAMFILES(X86)	C:\Program Files (x86)

### Get Host Name



Node 55

### Run Garbage Collector



Node 56

```
INFO RunGarbageCollectorNodeModel Memory Status before GC:
INFO RunGarbageCollectorNodeModel Free memory: 566 MB
INFO RunGarbageCollectorNodeModel Used memory: 1414 MB
INFO RunGarbageCollectorNodeModel Allocated memory: 1981 MB
INFO RunGarbageCollectorNodeModel Running Garbage Collection...
INFO RunGarbageCollectorNodeModel Free memory: 1909 MB
INFO RunGarbageCollectorNodeModel Used memory: 86 MB
INFO RunGarbageCollectorNodeModel Allocated memory: 1996 MB
INFO RunGarbageCollectorNodeModel Max. available memory: 1996 MB
```

## Implementation details

```
public class SettingsModelStringArrayFlowVarReplacable extends SettingsModelString
    implements SettingsModelFlowVariableCompatible {
    ...
    /* CONSTRUCTORS */
    SettingsModelStringArrayFlowVarReplacable(String, String[]) //Standard delimiter (;)
    SettingsModelStringArrayFlowVarReplacable(String, String[], String) //User-defined delimiter

    public String[] getStringArrayValue() {
        String stringValue = super.getStringValue();
        if (stringValue == null) {
            return null;
        }
        if ("".equals(stringValue)) {
            return new String[0];
        }
        ... PROCESS TO AN ARRAY HERE ...
    }

    public void setStringArrayValue(String[] values) {
        String stringVal;
        ... ...
        super.setStringValue(stringVal);
    }

    /**
     * @return The delimiter used for serialization of Arrays to Strings and
     *         vice versa
     */
    public String getDelimiter()
```