FoX, CML and tools from eMinerals

Andrew Walker
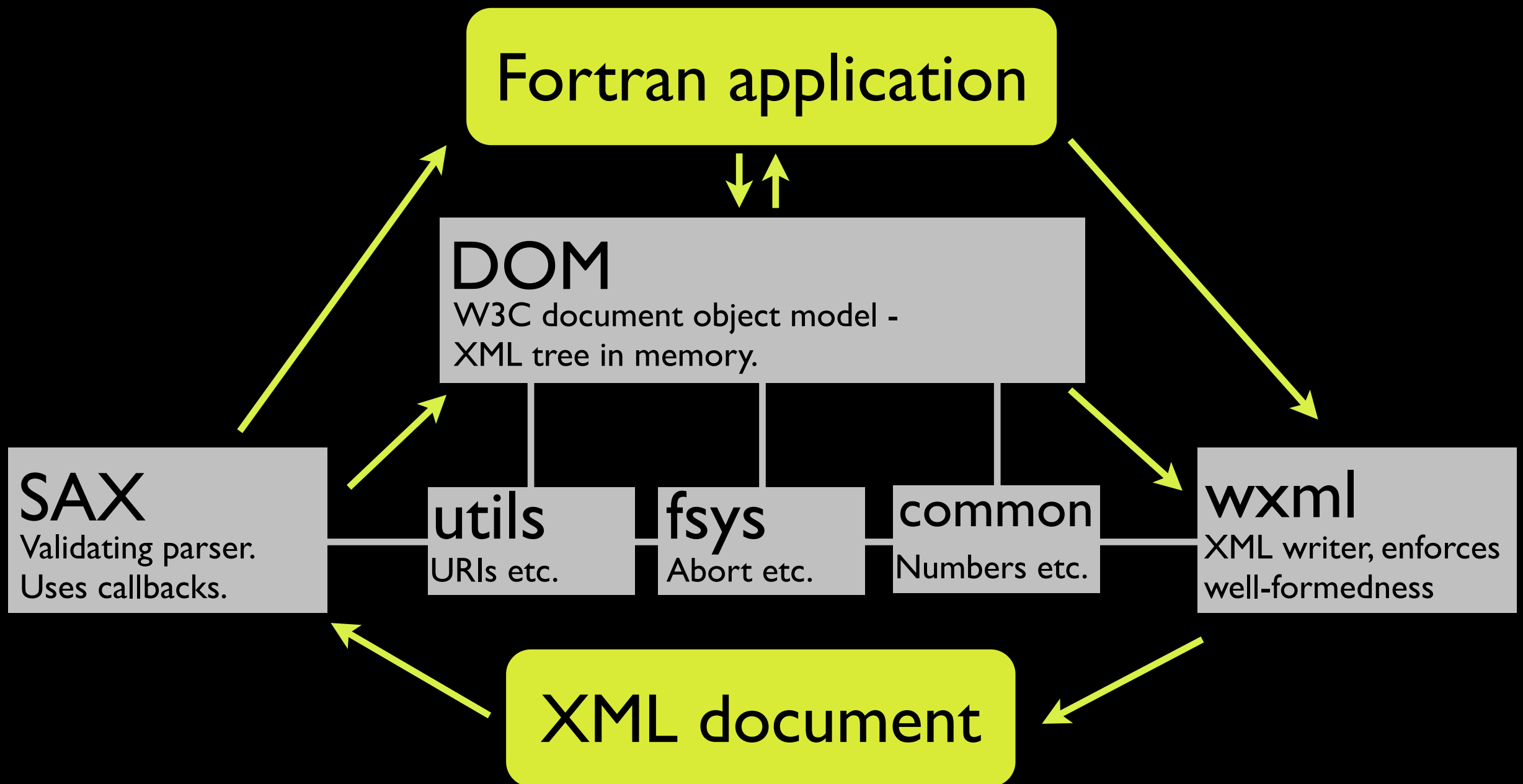
andrew.walker@bristol.ac.uk

Big picture

Computational
chemistry
application

output
data

Data
management
system

(Driven by a grid
computing infrastructure)

Probably written in Fortran

Computational chemistry application

Convert 'legacy' output to XML

Link against XML library written in C

Data management system

Knows about XML

Rewrite application in e.g. Java

Use a Fortran XML library
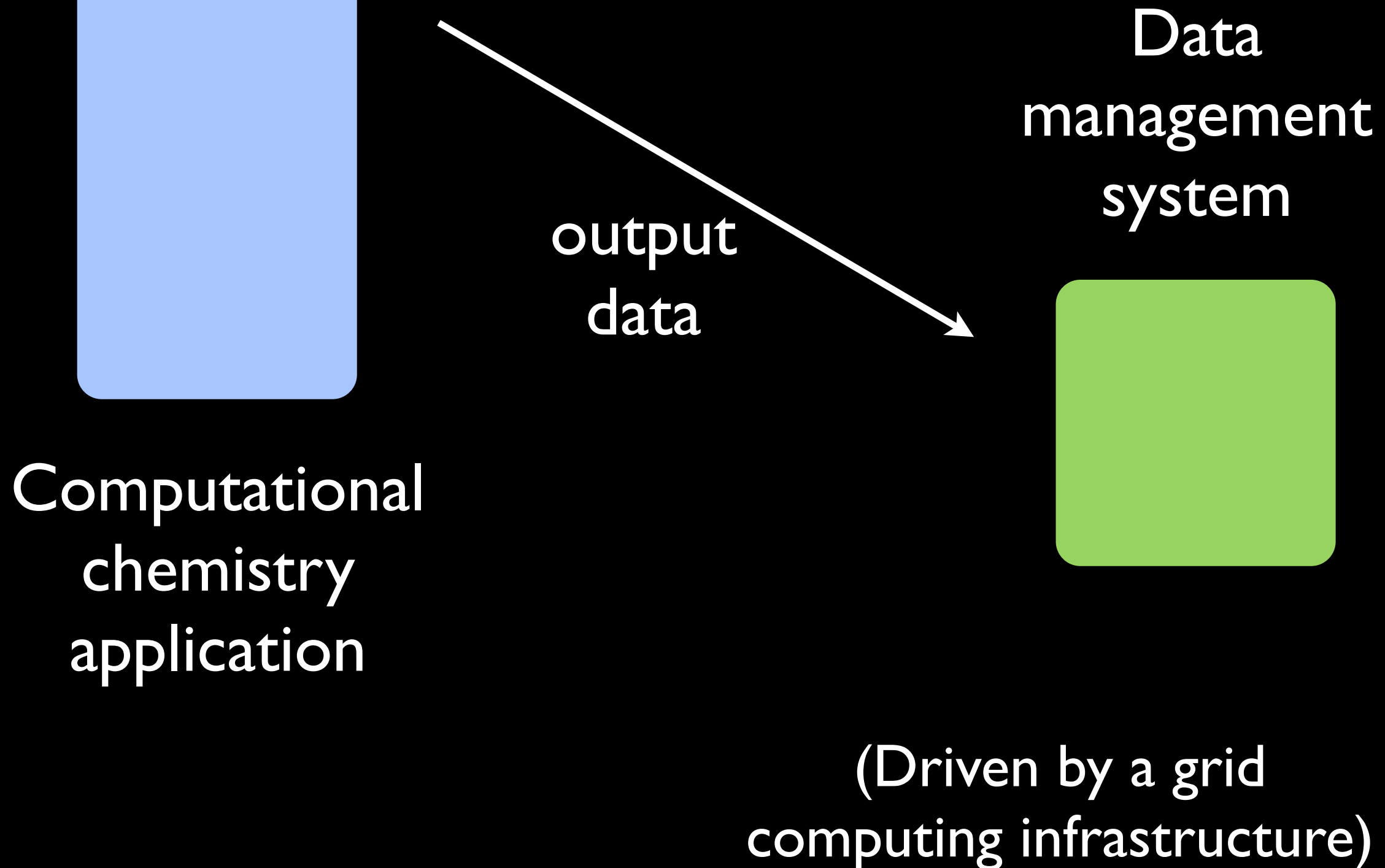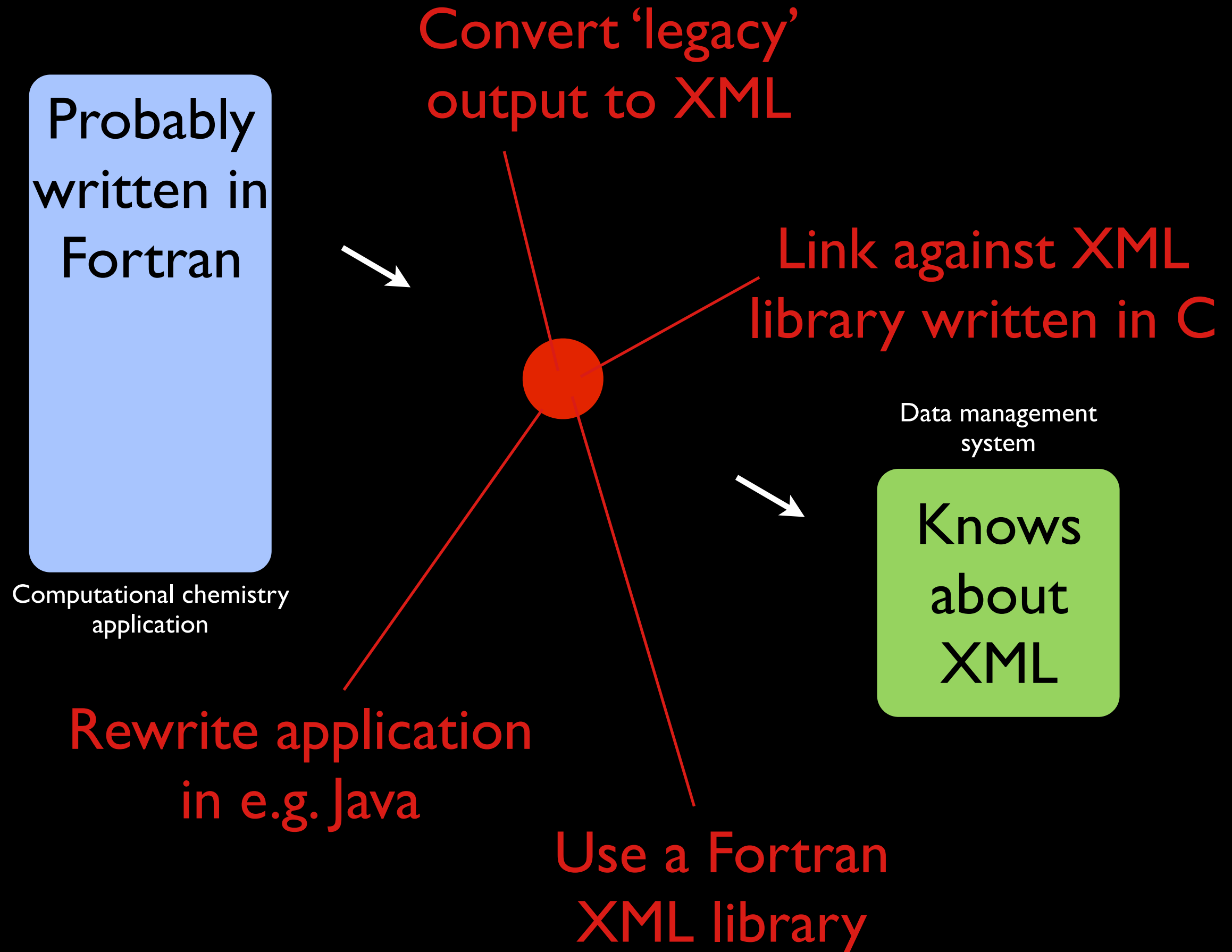
FoX: a pure Fortran 95 XML library

Fortran application

DOM
W3C document object model -
XML tree in memory.

SAX
Validating parser.
Uses callbacks.

utils
URIs etc.

fsys
Abort etc.

common
Numbers etc.

wxml
XML writer, enforces
well-formedness

XML document

- www1.gly.bris.ac.uk/~walker/fox
- Extensive tests and documentation
- Widely used (chemistry, geophysics, NHS, industry...)

- xmlf90: Alberto Garcia & Jon Wakelin
- FoX: Toby White 2003 - 2006
- FoX: Andrew Walker 2006 - 2012...

Not just any XML

XML specification defines the "punctuation" in our documents

Go from this:

```
#! /usr/bin/perl -w
$num = '[+-]?\d+\.?\d*';
while (<>) {
  if (/temperature = ($num)/) {
    $temp = $1;
  } elsif (/pressure = ($num)/) {
    $pres = $1;
  }
  ...
```

(if the format is kind)

To this:

```
#! /usr/bin/perl -w
use XML::XPath;
my $xp = XML::XPath ->
          new(filename => $ARGV[0]);
$temp = $xp -> find("//temperature");
$pres = $xp -> find("//pressure");
...
```

(which works if 'temperature' occurs many times in the document)

Not just any XML

XML specification defines the "punctuation" in our documents

Choose a language - CML: provides a menu of tags

At least, the way we used it

Not just any XML

XML specification defines the
"punctuation" in our documents

Choose a language - CML:
provides a menu of tags

Be more specific - CMLComp:
how to use the tags for computational chemistry

*Microformats and
document structure*

Not just any XML

XML specification defines the
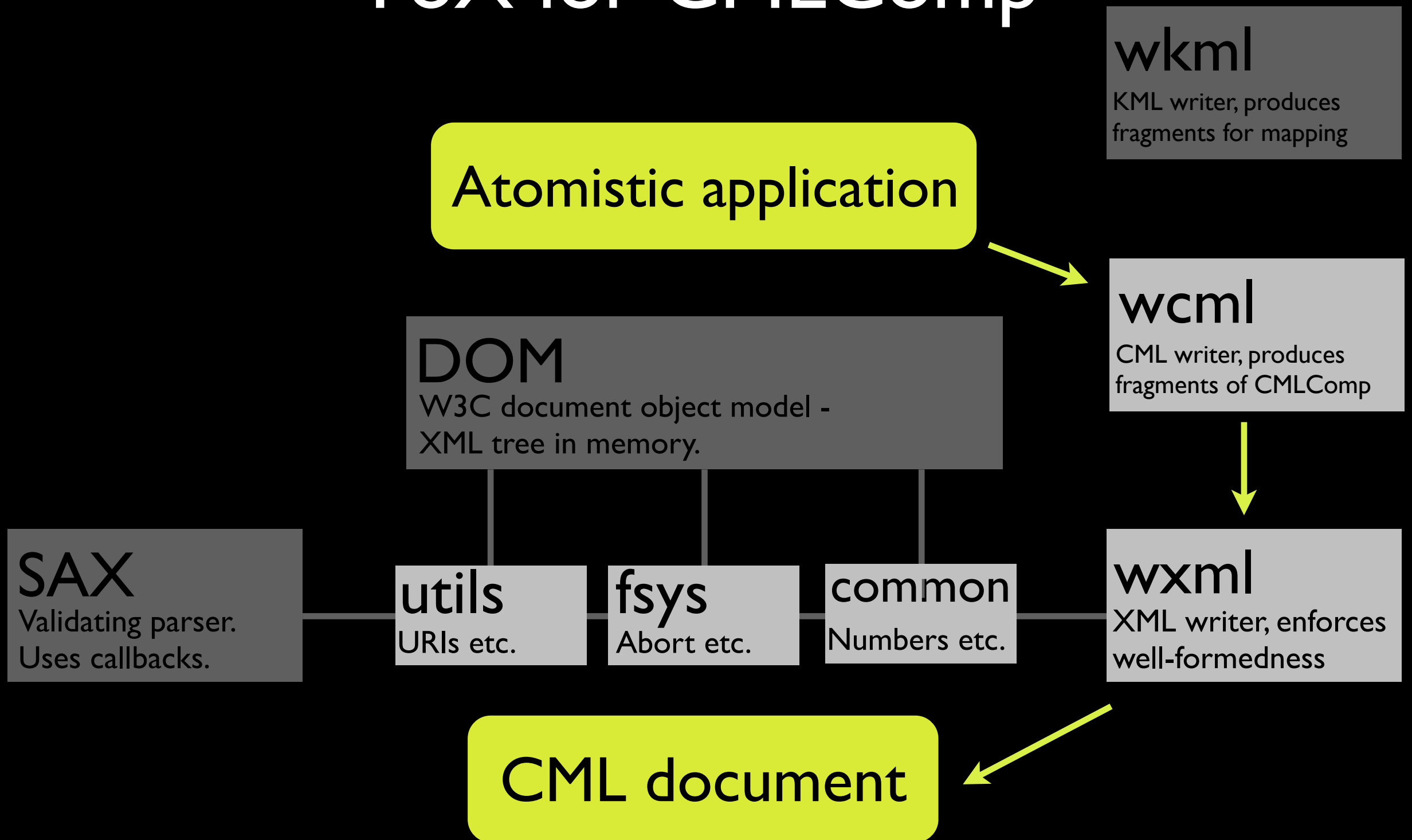"punctuation" in our documents

Choose a language - CML:
provides a menu of tags

Be more specific - CMLComp:
how to use the tags for computational chemistry

Semantics for each code is different:
use CML dictionaries

What is 'temperature'?

Not just any XML

XML specification defines the
"punctuation" in our documents

Choose a language - CML:
provides a menu of tags

Be more specific - CMLComp:
how to use the tags for computational chemistry

Semantics for each code is different:
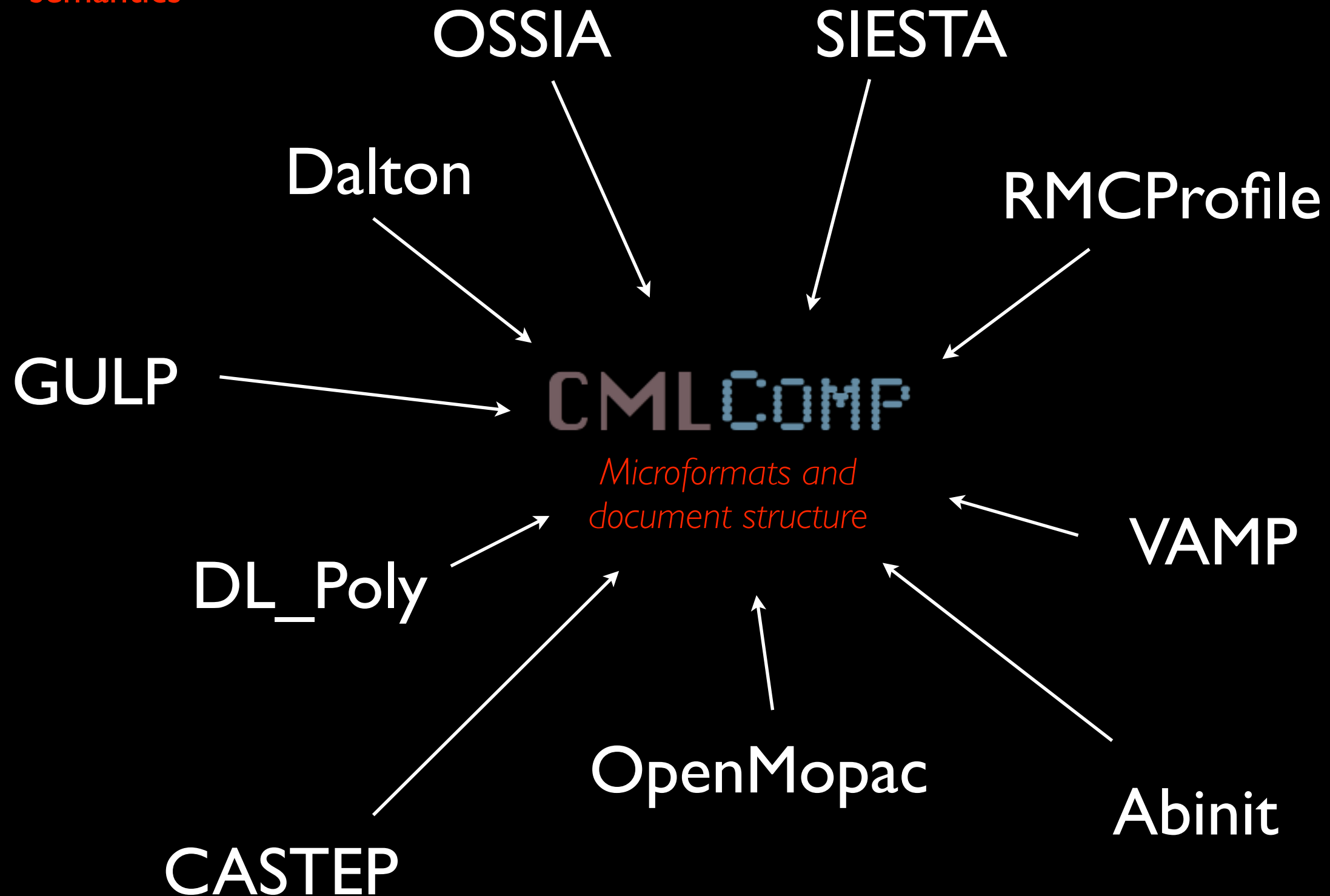use CML dictionaries

Build some tools

FoX for CMLComp

wkml
KML writer, produces fragments for mapping

Atomistic application

wcml
CML writer, produces fragments of CMLComp

DOM
W3C document object model - XML tree in memory.

SAX
Validating parser. Uses callbacks.

utils
URIs etc.

fsys
Abort etc.

common
Numbers etc.

wxml
XML writer, enforces well-formedness

CML document

- Simple API for producing CML
- Only supports a subset of CML used in specific ways
- Enforces many of the CMLComp constraints

- Cutdown distribution (wcml + deps.)
- Don't need to know much XML
- www1.gly.bris.ac.uk/~walker/CMLComp/

A subset of CML with more strongly defined syntax and semantics

OSSIA

SIESTA

Dalton

RMCProfile

GULP

CMLComp

Microformats and document structure

DL_Poly

VAMP

OpenMopac

CASTEP

Abinit

ccVis - demo

```xml
301
302    <entry id="temperature" term="Configuration temperature">
303      <annotation />
304      <definition>Configuration temperature</definition>
305      <description>
306      <h:p>The set temperature is used to control the Metropolis Monte
307      Carlo algorithm. Denoting <h:i>E</h:i> as the change in energy
308      associated with a proposed change in the configuration, the change
309      is accepted if <h:i>E</h:i> is negative. Otherwisee the proposed
310      change is accepted with probability given by
311      exp(-<h:i>E</h:i>/<h:i>T</h:i>), where <h:i>T</h:i> is the set temperature.
312      The units of temperature are the same as the units of the exchange
313      parameters, and OSSIA has no units-specific dependency.</h:p>
314      </description>
315        <metadataList>
316        <metadata name="dc:author" content="golem-kiln" />
317      </metadataList>
318      <golem:xpath>.//cml:parameterList/cml:parameter[@dictRef='ossia:temperature']</golem:xpath>
319      <golem:template call="scalar" role="getvalue" binding="pygolem_serialization" />
320      <golem:implements>value</golem:implements>
321      <golem:implements>relative</golem:implements>
322    </entry>
323
```

Dictionaries for: CASTEP, SIESTA, OSSIA.

```
301
302    <entry id="temperature" term="Configuration temperature">
303      <annotation />
304      <definition>Configuration temperature</definition>
305      <description>
306      <h:p>The set temperature is used to control the Metropolis Monte
307      Carlo algorithm. Denoting <h:i>E</h:i> as the change in energy
308      associated with a proposed change in the configuration, the change
309      is accepted if <h:i>E</h:i> is negative. Otherwisee the proposed
310      change is accepted with probability given by
311      exp(-<h:i>E</h:i>/<h:i>T</h:i>), where <h:i>T</h:i> is the set temperature.
312      The units of temperature are the same as the units of the exchange
313      parameters, and OSSIA has no units-specific dependency.</h:p>
314      </description>
315        <metadataList>
316        <metadata name="dc:author" content="golem-kiln" />
317      </metadataList>
318   ● <golem:xpath>.//cml:parameterList/cml:parameter[@dictRef='ossia:temperature']</golem:xpath>
319   ● <golem:template call="scalar" role="getvalue" binding="pygolem_serialization" />
320   ● <golem:implements>value</golem:implements>
321      <golem:implements>relative</golem:implements>
322    </entry>
323
```

Golem: tool to use and generate dictionaries

Put XPath in dictionaries to extract concepts

Use XSLT to transform CML to JASON

Some basic triples

www1.gly.bris.ac.uk/~walker/CMLComp/