Chemistry Add-In for Word

http://research.microsoft.com/chem4word/ http://chem4word.codeplex.com

University of Cambridge

Dr Joe Townsend

Collaborative R&D between the Unilever Centre for Molecular Science Informatics, University of Cambridge and Microsoft Research, Redmond.

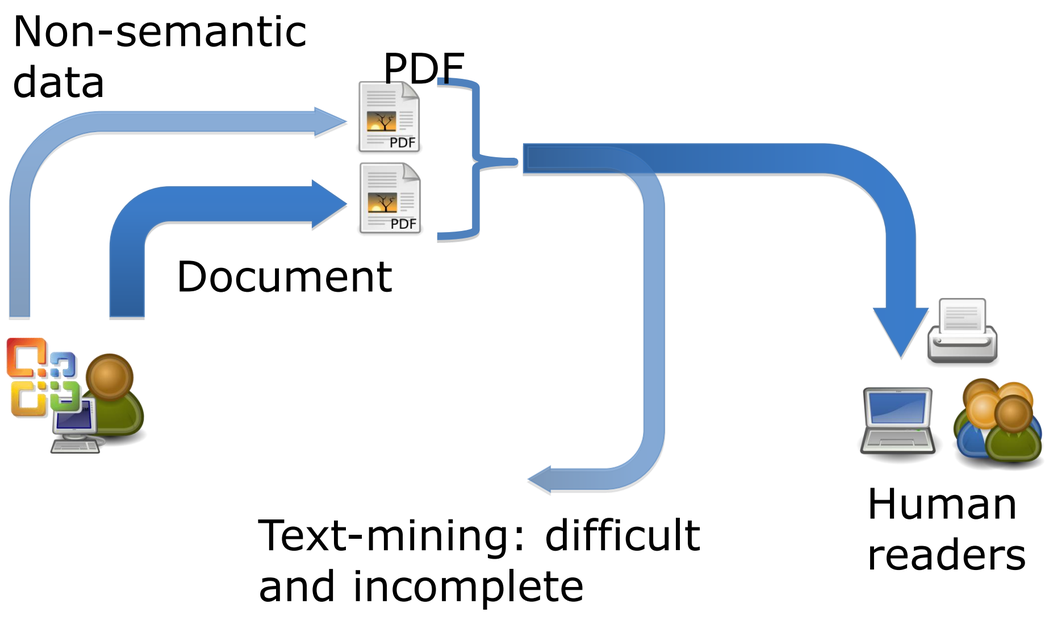
People *(UoC) Peter Murray-Rust, Joe Townsend, Jim Downing. (Microsoft) Lee Dirks, Alex Wade, Oscar Naim, Mike Galos, Tim Haughton.*

Semantically Editing Chemistry in Word

The Chemistry Add-In for Word, is an open source program that allows chemists to create, edit and manipulate chemistry (labels and 2D structures) in the Word environment. The on screen representation is backed by semantic data in Chemical Markup Language (CML). Combined with domain aware libraries we enable novel functionality in data checking during the authoring process, chemistry-centric article reading support and data-mining applications.

Current publishing practice

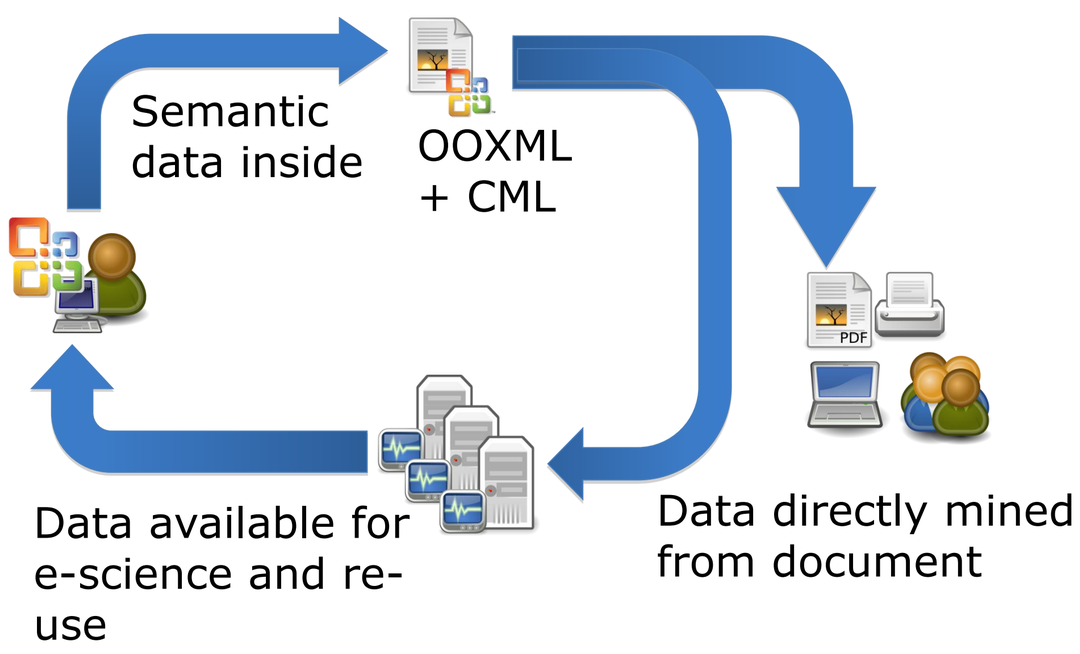
… is broken for data-rich science



- Data publication difficult and unsupported
- Insufficient data to fully support research

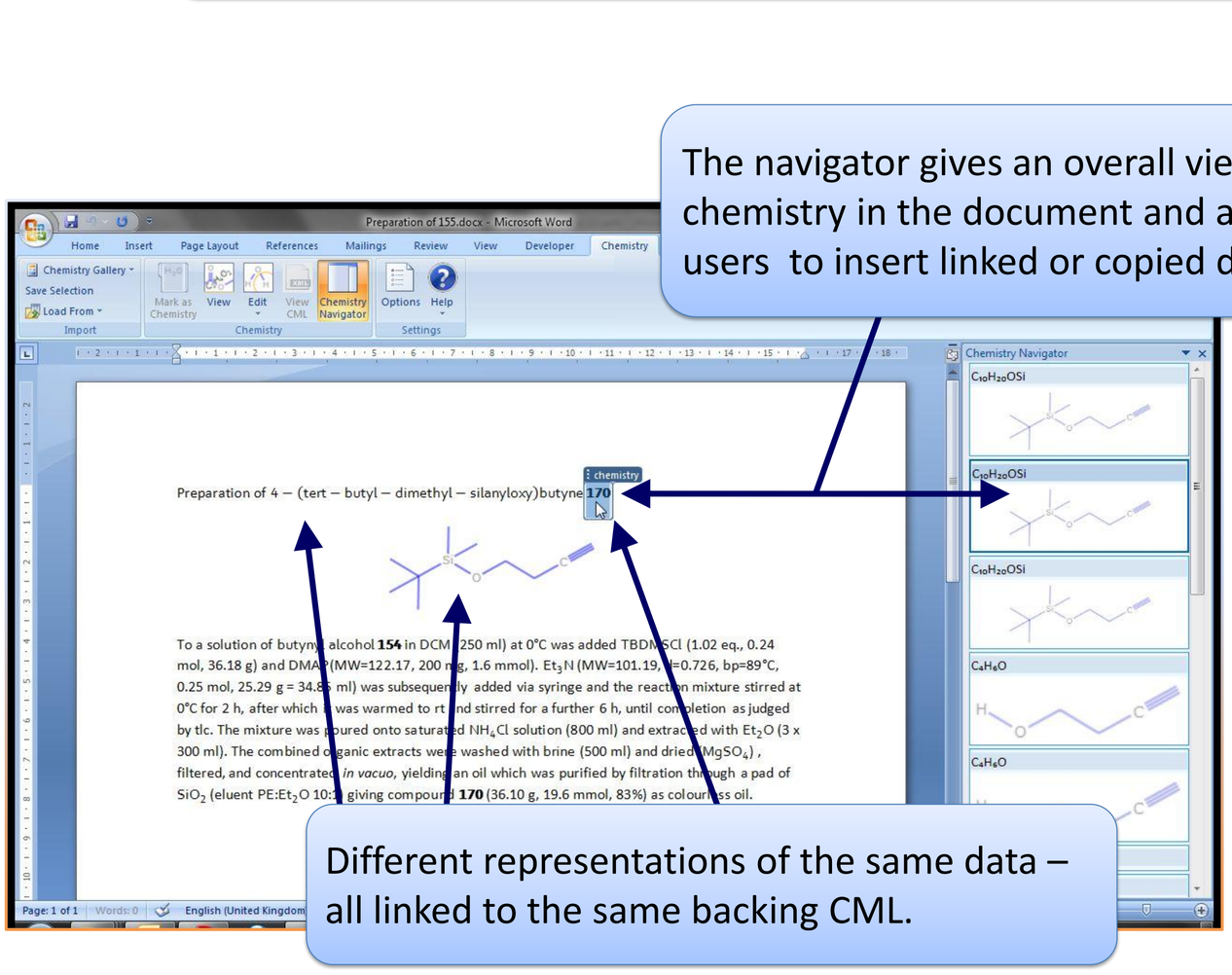
With Chemistry Add-In for Word

… the cycle is closed

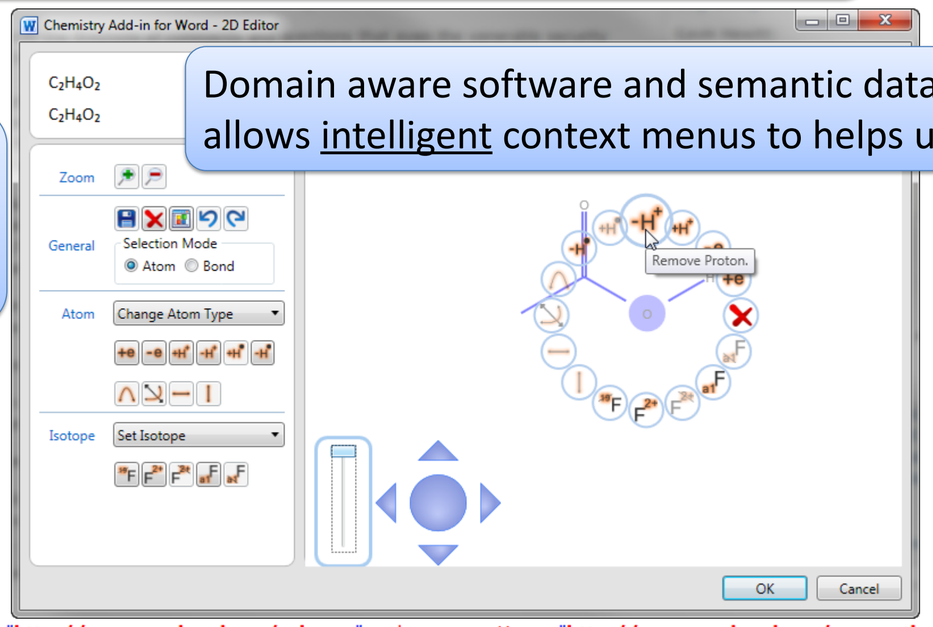


- Data preparation integrated into user workflow
- Open Standards promote Open Semantic Science

The Add-In (.NET) links a semantic chemistry engine (.NUMBO) through a command interface (CID) to a chemistry zone. Chemistry zones are textual or graphic renderings within a Word document (DOCX). All content and relationships are bound to CML in the DOCX package.



The navigator gives an overall view of the chemistry in the document and allows users to insert linked or copied data.



Domain aware software and semantic data allows intelligent context menus to helps users.

Different representations of the same data – all linked to the same backing CML.

```
- <cml xmlns="http://www.xml-cml.org/schema" xmlns:conventions="http://www.xml-cml.org/convention/"
    convention="conventions:molecular">
  - <molecule id="m1" formalCharge="-1">
      <formula concise="C 2 H 3 O 2 -1" />
    - <atomArray>
        <atom elementType="C" id="a1" x2="-16.6" y2="4.9" />
        <atom elementType="C" id="a2" x2="-15.3" y2="5.7" />
        <atom elementType="O" id="a3" x2="-13.9" y2="4.9" formalCharge="-1" />
        <atom elementType="O" id="a4" x2=
        <atom elementType="H" id="a6" x2=
        <atom elementType="H" id="a7" x2=
        <atom elementType="H" id="a8" x2=
      </atomArray>
    - <bondArray>
        <bond id="b1" atomRefs2="a1 a2" o
        <bond id="b2" atomRefs2="a2 a3" o
        <bond id="b3" atomRefs2="a2 a4" order="D" />
        <bond id="b4" atomRefs2="a1 a6" order="S" />
        <bond id="b5" atomRefs2="a1 a7" order="S" />
        <bond id="b6" atomRefs2="a1 a8" order="S" />
      </bondArray>
    </molecule>
  - <cidCommand instruction="RemoveProton" xmlns="http://www.xml-cml.org/dictionary/cmlx">
      <arg class="atom" name="atomPointer">a3</arg>
    </cidCommand>
</cml>
```

A complete record of the changes to the data is retained in the customXML.

UNIVERSITY OF CAMBRIDGE

Microsoft External Research Initiative

External Research

Microsoft Research