The Application of XML Languages for Integrating Molecular Resources

Content

1) The Application of XML Languages for Integrating Molecular Resources
2) Georgios V. Gkoutos,a Peter Murray-Rust,b Henry S. Rzepaa and Michael Wrighta
3) Abstract:
4) Keywords:
5) Introduction
6) Early Internet-based Integration of Molecular Resources
7) Molecular Integration based on XML
8) The ChemDig Project
9) Chemical Markup Language
10) The Importance of Separating Presentation from Content
11) Stylesheet-based Transformations
12) The Chimeral Project
13) Other Examples of XML Document Transformations
14) 1. Transformation to Acrobat Format
15) 2. Transformation to 2D/3D Molecular Representations
16) 3. Transformation for Editing and Annotation
17) Conclusions
18) Acknowledgements
19) References and Footnotes

The Application of XML Languages for Integrating Molecular Resources

Georgios V. Gkoutos,[a] Peter Murray-Rust,[b] Henry S. Rzepa[a] and Michael Wright[a]

[a]Department of Chemistry, Imperial College, London, SW7 2AY, England. [b] Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge.

Abstract:

We review recent developments which employ extensible markup languages (XML) as information descriptors providing an unparalled opportunity to achieve inter- and multi disciplinary integration of molecular resources. We illustrate these developments by reviewing two projects; the ChemDig molecular aggregator, and the Chimeral project which involves the use of, *inter alia*, CML (Chemical Markup Language).

Keywords:

Extensible-markup-language (XML), Chemical-markup-language (CML), Extensible-stylesheet-language-transformations (XSLT), XML Schemas, ChemDig, Chimeral, JUMBO, Semantic-Web.

Introduction

The impact of the Internet on how molecular resources are delivered to the human reader has undeniably been substantial. We argue in this review that hitherto the emphasis has

been on the delivery, that the degree of integration of these resources has been deceptively low, and that to a large degree achieving such integration still does require implied human intervention at virtually every stage. We propose that the next stage in the evolution of the Internet for molecular resource discovery must be towards what has been described as a semantic Web, [1] where not only humans but machines can make "chemically intelligent" decisions based on collections of authenticatable assertions about chemistry and molecular sciences. For the past six years, we have been working towards such a vision, and describe here some of the concepts, architectures and tools that have been developed to achieve this goal.

Early Internet-based Integration of Molecular Resources

We start by considering what the integration of molecular resources might have meant in the past. Classical integrations were achieved by organisations such as Chemical Abstracts, Beilstein or IUPAC, where human perception of molecular information was translated into a body of knowledge about the 30 or so million currently known chemical substances. Access to such integration is achieved through a single commercial portal, but integration across different portals remains largely a human activity by the researcher, necessitated in part because of their incompatible data structures. Individual scientists contribute freely to this body of knowledge via the primary publication process, which is also dominated by a small number of large publishers with traditional copyright rules that often meant transfer of all authors' rights to the publisher. It is also at this stage that much of the basic data and structure that defines a molecular resource is either uncaptured, or discarded, only to be again expensively and only partially and in an error-prone manner re-instated during the secondary and tertiary publication processes. [2]

We next consider how the advent of the Internet and the World-Wide Web from 1993 onwards has modified these classical processes. The Web-based document introduced some important new concepts. The first is that such a document is regarded as a container for a collection of information objects formally known as elements, and less formally as "markup", and expressed syntactically using a language such as HTML. The second concept was that these objects can be associated together by either being presented in serial form in the same document or computer file, or in a distributed manner using the so called "hyperlink". The use of such document links, which in fact take many syntactic forms, has resulted in what many people would now regard as the creation of new types of apparently integrated molecular information resources. Thus links could be established to molecular coordinate files (for example the very common Molfile or PDB format), to spectroscopic information (for example the JCAMP format), to interpretable scripts (e.g. the Rasmol script format, which has been used to assemble impressive integrated knowledge bases such as Protein Explorer [3]), to executable resources such as searchable queries to a remote database, and most ubiquitously, to graphical objects which took the form of bitmap images. In 1994, our first major effort to formalise how such molecular links were integrated was expressed by our proposed Chemical MIME standards, which were widely adopted and used by others to create some impressive collections of molecular information and knowledge. [4]

For the first time, on a global scale, the monopoly of the traditional molecular integrators had been extended to include a host of smaller specialist providers, ranging from dedicated individuals to medium sized organisations. From around 1996, the primary and secondary publishers also started to join the party by making primary research articles

available as Web documents, and now resources such as SciFinder or Science Direct offer apparently highly integrated environments. In such environments, [5] e.g. chemical sub-structure searches specified of the 30 million known chemical substance database can be integrated with delivery of the document corresponding to the primary published report, together with more specific properties such 3D coordinates or spectroscopic characterisation. Such integrated environments however literally come with a price, and with the important limitation that the integration process itself is proprietary, is not exposed publicly and its features are controlled entirely by single organisations and not by the original authors of the information. We set out to ask if a more open communal model of integration could be achieved by building on the standards first established by the World-Wide Web.

Molecular Integration based on XML

Our approach was two fold. Firstly, we investigated methods of how to integrate existing but separate resource collections into unified searchable collections where added molecular value could be automatically derived and included, and for which a degree of regularisation of the syntactic expression of the molecular resources could be achieved. We called this approach the ChemDig project, [6] and its basic features are described below. In parallel with this, we started work in 1995 on a better means of expressing molecular information than the combination of HTML and the "legacy" formats referred to via the Chemical MIME TYPES. This project we termed "Chemical Markup Language" or CML. [7] By 1996, it had become recognised that markup languages should be extensible into well defined subject areas, or "domains" and the World-Wide Web consortium (W3C) had defined an XML (eXtensible markup language) project as the foundation to achieve this. W3C recommended basic syntactic standards for implementation. [8] CML was one of the first domain specific examples of such an implementation of XML, another prominent examples being MathML (Mathematical markup language) [9] followed somewhat later by SVG (scalable vector graphics) [10] and others. Our two strands merged with the ChemDig project having the purpose of a tool for transforming classical web collections of documents into XML compliant form which can, *inter alia*, comprise XHTML (an XML conforming version of HTML) and CML. In the remainder of this article, after describing the essential features of ChemDig and CML, we will discuss how such an XML-based approach can achieve a higher level of molecular integration than more traditional approaches, and how the prospect of even higher level integration across subject domains becomes possible.

The ChemDig Project

ChemDig has been described in a series of articles, to which the reader is referred for full details. [6] Its basis is that of the Internet Robot, which makes use of the linking mechanisms available in HTML to automatically traverse a document collection. The basic robot is known as ht://Dig, the source for which is freely available for development, and which includes the facility to include so-called external parsers, in our case appropriate for capturing molecular resources. The operation of ChemDig comprises three stages (Figure 1).

- The first stage involves recognising that much of the HTML deployed on the estimated 3000 global collections of molecular resources can be syntactically not well formed, and where the inclusion of certain types of molecular resource such as e.g. Rasmol scripts can mean that the ht://Dig robot will not function correctly. A module

of ChemDig called JChemTidy functions to regularise all but the most badly formed HTML to the current specification of XHTML 1.0. These operations include normalising some of the more arcane ways of invoking links between documents and molecular resources into formal document metadata, using another module called JChemMeta. [11]

- ChemDig is now used to traverse a remote document collection, specified purely by the URL of the home or root page of the resource. When it encounters a molecular resource in the form of a Chemical MIME link, an external chemical metaparser is invoked to capture important metadata relevant to the file, and to return this information to the ChemDig robot in the form of a specific metadata schema, in our case based on the Dublin Core proposals. By default, a ChemDig database of bibiliographic keywords encountered during this process is created, which includes metadata information such as the presence in the collection of e.g a Molfile or spectroscopic file. Additionally, a molecular formula where it can be established and connectivity data such as a SMILES descriptor if it can be calculated is also included. The database can be subsequently searched against this information using standard Boolean-based keyword terms.
- ChemDig can now also be used to perform certain post-processing operations. These include diverting selected discovered chemical MIME resources into a specific chemical database. In our case, we created a Daylight database of molecular connectivity which allows chemical similarity searches to be integrated into the search operation. In parallel, a CML version of any molecules identified can be created, which together with the XHTML produced in the first stage can be written out into a new composite document. Now, both the bibliographic descriptions of the molecular resources originally expressed in HTML, together with the molecule-centred descriptions originally expressed in "legacy" formats, can be concatenated into a single XML-based document. Such an integration can be demonstrated with this article itself [12] using XML tools, a process described in more detail below.

Before describing the advantages of having such chemically integrated documents, we will discuss the basic features of CML itself.

Chemical Markup Language

As with ChemDig, the technical detail of CML has been discussed elsewhere, [6,13] and here we cover only the general concepts. When early prototypes of the CML language were first discussed in 1995, a common point of view we heard expressed was that it was "just yet another molecular file format", of which there more than 50 already! In fact however, CML is an expression of two ambitions. The first is to create an opportunity to avoid the fragmentation of the subject into "information silos" where the data structures and syntax are proprietary and non-interoperable with each other and other disciplines. CML follows a general set of recommendations established by the W3C for writing XML languages, and as such benefits from the considerable number of XML tools and operations derived from a communal effort from across many subject areas. [9] We emphasize that no prior existing molecular information formalism has ever been able to take advantage of the availability and inter-operability of such tools.

The second ambition was to take advantage of the extensible character of the XML specifications. The basic CML structures use a set of conventions which we term CoreCML, and which are specified more formally via a CMLDOM (CML document

object model) employing interfaces and methods constructed using a collection of about 300 Java classes. [14] A concrete example of this would be encoding the convention for an atom and the corresponding software specifications for what comprises a valid atomic number, specific isotopes and other behavioural characteristics of what the chemical community understands by the term atom. These can be extended by defined external conventions for specifying further molecular terms and behaviours, which themselves can be defined, standardised and extended by e.g. organisations such as e.g. IUPAC. CML thus provides a method for handling what is described as the ontology of chemistry, i.e. a formal and communal agreement in which the meaning of terms used in the subject domain can be incorporated into the knowledge base. Concrete examples of such ontology include concepts such as "aromaticity", "tautomerism" and other molecular properties that are often delegated for purely human perception, but which must also in the future be processable by software agents as well.

With CML therefore we specifically set out to capture molecule, atom, bond and electron-centric information according to specified or implied conventions, and to formally specify links between these objects via the citation of identifiers (IDs) associated with each object. As with HTML, it is possible therefore to address molecular information expressed using CML to both a very finely grained level and to a highly aggregated and integrated level. In this, CML has several characteristics of an object-oriented database.

The Importance of Separating Presentation from Content

At this point, we have to address another fundamental, and often very mis-understood concept associated with markup languages in general. This is the formal requirement to separate the information captured using the markup from the manner in which this information is subsequently used. This separation is also often referred to as one of style from content, or of presentation from data. Unfortunately, much confusion originated with HTML, which is often authored to create a presentational effect rather than to capture data and information. Take for example the element often found in HTML; <hr> When delivered to a Web browser, this will produce the effect of a horizontal line, serving to separate one section of the document from another. However, its *meaning* is far less clear, and open to many interpretations. Indeed,the meaning is often considered so obscure that whatever information it represents would not be processed by most tools. It is now generally recognised that where presentational effect is desired, a stylesheet should be used to achieve it rather than markup. The information elements are addressed via their (unique to the document) identifier from within the stylesheet. The concept of rigorous separation of style and data is strictly adhered to in CML as well. Whilst it is possible to display a CML document within a browser capable of handling XML as a so-called document tree view, in most cases the presentation of CML will involve either the application of a stylesheet within a browser, or the use of specific software capable of abstracting CML into an in-memory representation (the CMLDOM). The stylesheet can be specified within the document itself, it can be invoked as part of e.g. the browser preferences, or it can be applied as part of an external process.

Stylesheet-based Transformations

At this point, we must introduce to further XML formalisms. XSLT (eXtensible Stylesheet Language Transformations) were originally envisaged as an extension to presentational stylesheets (hence the name), but are now recognised as a powerful

object-oriented transformation language. These can be used in conjunction with XML Schemas, which provide a means for defining the structure, content and semantics of XML documents. Applied to CML, we envisage at least three separate but related and hence integrated applications of such stylesheets and schemas;

- Use of XSLT to transform an XML document for presentational effect within a browser. This was the original application we described in the Chimeral Project, where the original data carried in an XML document was converted to syntactic form capable of rendering using a combination of browser and Java display applets. [13]
- Use of a combination of XSLT components and Schemas to check the validity and chemical consistency of XML and CML documents. Schemas are used to define rules such as: a bond must be specified by at least two atoms; a (chemical) element must conform to that defined by the periodic table. XSLT components can be used to perform e.g. a valency check on the atom and bond specifications, or to normalise the hydrogen count on an atom.
- Use of XSLT to create additional information about a molecule specified using CML, as for example deriving a molecular formula or molecular mass, or to perform searches on the content of the CML document, including identifying separate molecules, finding specified sub-structures of molecules or answering queries such as "how many molecules associated with a particular toxicological effect contain chlorine".

It is important to appreciate that the outputs of all such operations can themselves be expressed in XML. Thus these operations correspond to structured annotations of the original document, and are seen as a core part of the process of handling XML/CML documents. Such a process does not have to be achieved by XSLT transforms alone, but could be handled by XML/CML aware editors based on the DOM model. We also note in this context that XSLT based operations can be assembled from XSLT libraries of core components.

In such a scenario, the integration of data with tools and resources to achieve specified transformations of the data becomes possible. The "document" of the original World-Wide Web now can be now instead regarded as an "Information utility" in which finely grained data can be directly captured and addressed, associated with the appropriate chemical ontology or dictionary terms, and transformed according to the reader's wishes. Such information utilities could be regarded as natural integrations of the concepts of separate documents and application software which are currently widely used.

Many interesting consequences follow. Information utilities should be capable of implicit integration, aggregation and extension, both within and across subject boundaries. Established mechanisms for handling copyright and intellectual property associated with a "document" will have to be fundamentally rethought. Journals would now consist of integrated collections of information utilities, but authors may become less willing to assign copyrights in such components to publishers, and the fundamental role of the publisher could well be challenged if the role of authentication, validation, integration, aggregation and distribution can be routinely achieved by software agents rather than only humans working for publishers.

The ultimate integration would be achieved if all data collections such as all journal articles, Web-pages, instrumental and computational outputs, books and dictionaries and ontologies of chemistry and other subjects were to be available in this form. A utopian

vision, but how might it work in practice? To start answering this question, we decided to develop some simple test-beds based on these principles. In recognition that we strove to integrate both chemical and non-chemical data with transform operations into new forms, we named the project Chimeral.

The Chimeral Project

This was conceived as a working test-bed example of the application of CML and its integration with other XML languages. Currently, two examples have been integrated using XML, and comprise what might be termed "smart" chemical journal articles. The first [15] makes use of the following open standard XML components;

- CML to carry molecules and reactions
- XHTML to carry hypertext
- A simple document language (DocML) created by us to express formal components of the article such as sections and headings, authors, and literature citations
- SVG (an XML language used to carry graphical representations and used as a structured replacement for bitmap images)
- SPECTRUM (an experimental XML language used as a proof-of-concept for expressing chemical spectral data derived from JCAMP files)
- XML schemas used to validate the XML components
- XSLT "translets" used to produce a display of the article within the Internet Explorer V5 or 6 browser (and shortly an anticipated implementation in the Netscape 6 browser)

When the XML root document defining the article is opened using the browser, the XSLT stylesheets are used to transform the various components into a form where existing software can be used to display it. Thus for example, SVG components were wrapped for display using the Adobe SVG browser plugin tool, CML was transformed and wrapped with XHTML for display using a variety of Java applets and plugins, and e.g. DocML components such as journal citations were wrapped with XHTML for display. We believe this article to be the first published XML and chemistry-based article.

The second article [16] illustrated how a further XML component can be used to add a digital signature (XML Signature) to particular objects within an article, which serves as an authentication mechanism for these components. The signature can be associated with a particular individual, an organisation, or an instrument or software agent, and its presence can be used as unambiguous proof that the data component has not been changed since it was created on a specified data, and also to assert intellectual or other ownership of that component. Furthermore two or more signed components can themselves be wrapped in an envelope, which can itself be signed. These mechanisms provide a way of establishing the provenance of any component, and indeed of transformations conducted on these components by others. An associated resource developed for the Chimeral project was a server-based system which accepts as input the URL of a conventional HTML based document or Molfile connection and coordinate file, converts these to XML/CML form, can be used digitally sign the conversion process, and then can return the aggregated and integrated document to the user (Figure 2) [17]

Other Examples of XML Document Transformations

1. Transformation to Acrobat Format

It is important to emphasize that XML information utilities can be processed by software other than Web browsers. The first example is a transform known as FOP (formatting object), which can be used to produce a high quality Acrobat file for printing, archival or legal purposes. The process was applied to this article itself as an illustration. Using ChemDig, the HTML, together with molecule data and diagrams in the form of SVG or JPG bitmaps conventionally in-lined using <embed> syntax, can be integrated into a single XML document. The molecule data, originally in Molfile format (Figure 3) is transformed to CML syntax, and the HTML into an XML-compliant version (XHTML), to allowed the integrated aggregation to be processed by XML-compatible software tools. Thus, the *resulting XML file* can be transformed into an Acrobat file using an especially written *XSLT stylesheet* and XML tools [18] . During this process, the CML molecule components (Atom, Bond) are transformed using XSLT to SVG vector components, and these scalable units are retained in the resulting *Acrobat file* , resulting in no loss of rendering resolution. The original JPG bitmaps of course cannot be so scaled.

2. Transformation to 2D/3D Molecular Representations

A major advantage of integration using XML is the facility to re-use XML in other contexts. We illustrate this by reading the XML/CML file corresponding to this article into the CML-aware browser JUMBO3. The alizarin molecule present is filtered and rendered using both 2D and 3D representations (Figure 4). Other browsers can handle other aspects of the document. For example, the Amaya browser can display both the XHTML and the SVG namespaces, and could also handle MathML components if they were to be present.

3. Transformation for Editing and Annotation

An XML document can be not merely rendered in a browser, but passed to an editing program as well. Amaya is a browser which also allows the XHTML and SVG to be edited. However, the use of a "hard-coded" executable program also results in some measure of inflexibility. It is however also possible to construct an XML-aware editor using XSLT transformations. Thus, the XML version of this article can be loaded into what we term an "eXtensible Annotating Chemical Editor" or XACE (Figure 5). XACE is constructed from a CML aware version of the JME structure editor (written as a Java applet) and a selection of XSLT components to filter and process the XML. The tool is rendered using an XML/XSLT compliant browser such as Internet Explorer. Such an environment can be used to *e.g.* edit the molecular structures, add new properties to the molecules, perform valency checks, compute molecular formulae, and add metadata in RDF format (another XML language).

Conclusions

The subject of this article has been to illustrate how chemical integration of information content at a molecular or sub molecular level of atoms, bonds and electrons can be achieved using combinations of XML languages. We suggest that adopting this metaphor on a wide and global scale could result in entirely new opportunities for the molecular science and other communities to construct a communal knowledge base founded on the information utility rather the printed article or electronic document. The Internet will then reveal its full potential, not merely of a delivery agent, but of a symbiote between machine and human capable ultimately of original scientific discovery.

Acknowledgements

We thank Merck Sharp and Dohme and the EPSRC for the award of a studentship (to GVG), Peter Ertl for help in creating a CML-aware version of JME and Allistair Crossley for similar help in creating a CML-aware version of his JMVS 3D molecule renderer.

References and Footnotes

[1] See articles by Harnad, S, *Nature*, 1999, **401 (6752)**, 423; Bachrach, S. M. *Quim. Nova* 1999, **22**, 273-276; Kircz, J. "New practices for electronic publishing: quality and integrity in a multimedia environment", UNESCO-ICSU Conference Electronic Publishing in Science, 2001; Rzepa, H. S. and Murray-Rust, P, *Learned Publishing*, 2001, July issue. These topics are currently being debated on forums such as the American Scientist Forum; *http://amsci-forum.amsci.org/archives/september98-forum.html* and at Chemistry pre-print sites such as *http://preprint.chemweb.com/* .

[2] Berners-Lee, T, Hendler, J, and Lassila, O, *http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html* ; Berners-Lee, T and Fischetti, M, "Weaving the Web: The Original Design and the Ultimate Destiny of the World-Wide Web", Orion Business Books, London, 1999. ISBN 0752820907.

[3] Martz, E., "Protein Explorer: Freeware for 3D visualization of macromolecular structure", *FASEB J.*, 2000, **14**, 22; *http://www.umass.edu/microbio/chime/explorer/pe.htm*

[4] Rzepa, H. S., Murray-Rust, P. and Whitaker, B. J., "The Application of Chemical Multipurpose Internet Mail Extensions (Chemical MIME) Internet Standards to Electronic Mail and World-Wide Web information exchange", *J. Chem. Inf. Comp. Sci.*, 1998, **38**, 976-982.

[5] For a Registry of Innovative E-Journal Features, Functionalities, and Content, see G. Mckiernan, *http://www.public.iastate.edu/~CYBERSTACKS/EJI.htm*

[6] Gkoutos, G. V., Rzepa, H. S., and Wright, M., "Hierarchical display of Chemical Data in Web Browsers", *Internet J. Chem.* , 2000, **3**, article 7; Gkoutos, G. V., Kenway, P., and Rzepa, H. S., "JChemTidy: A Tool for Converting Chemical Web Document Collections to an XHTML Representation", *J. Chem. Inf. Comp. Sci.*, 2001, **41**, 253-258; Gkoutos, G. V., Kenway, P. R., and Rzepa, H. S. "A robot-based resource discovery tool for adding chemical meta-information and value to web-based documents", *New. J. Chem.*, 2001, 635-638. 215.

[7] Murray-Rust, P. and Rzepa, H. S, *J. Chem. Inf. Comp. Sci.*, 1999, **39**, 928 and articles cited therein. See *http://www.xml-cml.org/*

[8] The definitive source of information about XML projects is available at the World-Wide Web Consortium site; *http://www.w3c.org/*

[9] See *http://www.w3c.org/Math/*

[10] SVG, see *http://www.w3c.org/Graphics/SVG/* ; PlotML, see *http://ptolemy.eecs.berkeley.edu/ptolemyII/ptII1.0/*

[11] The RDF specifications provide a lightweight metadata and ontology system to support the exchange of knowledge on the Web, see *http://www.w3c.org/RDF/*

[12] Gkoutos, G. V., Leach, C., and Rzepa, H. S., "ChemDig: New approaches to Chemically Significant Indexing and Searching of Distributed Web Collections", *New. J. Chem.*, 2001, submitted for publication.

[13] Murray-Rust, P, Rzepa, H. S, Wright, M. and Zara, S, "A Universal approach to Web-based Chemistry using XML and CML, *ChemComm*, 2000, 1471-1472

[14] Murray-Rust, P. and Rzepa, H. S. "Chemical Markup, XML and the World-Wide Web. Part II: Information Objects and the CMLDOM", *J. Chem. Inf. Comp. Sci.*, 2001, **41**, 1113

[15] Murray-Rust, P, Rzepa, H. S, Wright, M, "Development of Chemical Markup Language (CML) as a System for Handling Complex Chemical Content," *New J. Chem.*, 2001, 618-634. The full XML-based article can be seen at *http://www.rsc.org/suppdata/NJ/B0/B008780G/index.sht*

[16] Gkoutos, G. V., Murray-Rust, P., Rzepa, H. S., and Wright, M. "Chemical Markup, XML and the World-Wide Web. Part III: Towards a signed semantic Chemical Web of Trust", *J. Chem. Inf. Comp. Sci.*, 2001, **41**, 1124.

[17] Gkoutos, G. V., Murray-Rust, P., Rzepa, H. S., and Wright, M., "A Resource for Transforming HTML and Molfile Documents to XML Compliant Form", *Internet J. Chem.* , 2001, **4**, article 5.

[18] Details of this procedure can be found at *http://www.ch.ic.ac.uk/rzepa/xml/fop.html*

Figure 1. A schematic diagram of the operation of ChemDig, illustrating how the aggregated chemical information can be re-distributed through various output channels.

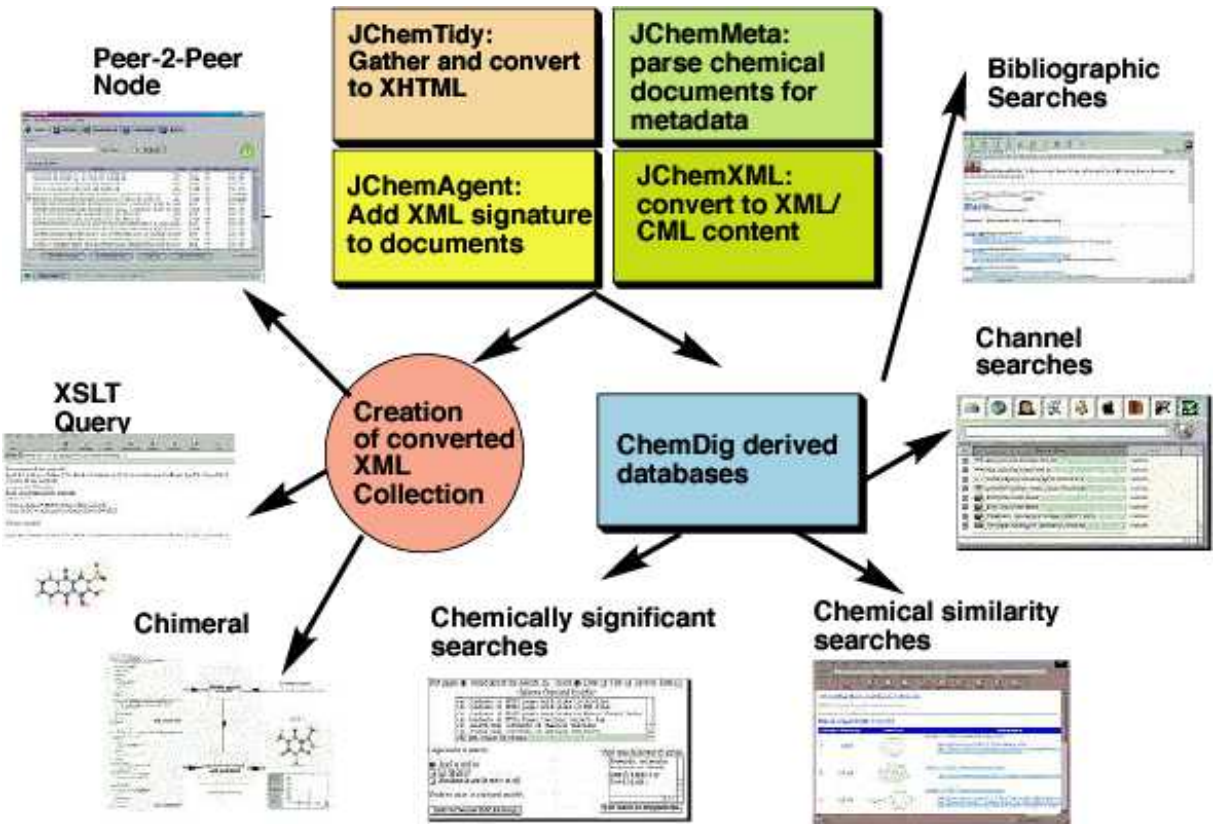

Figure 2. A schematic diagram showing how a molecular resource on the Internet expressed as a URI can be converted to CML form, annotated with additional metadata, digitally signed and the result viewed within a browser using XSLT transformations.

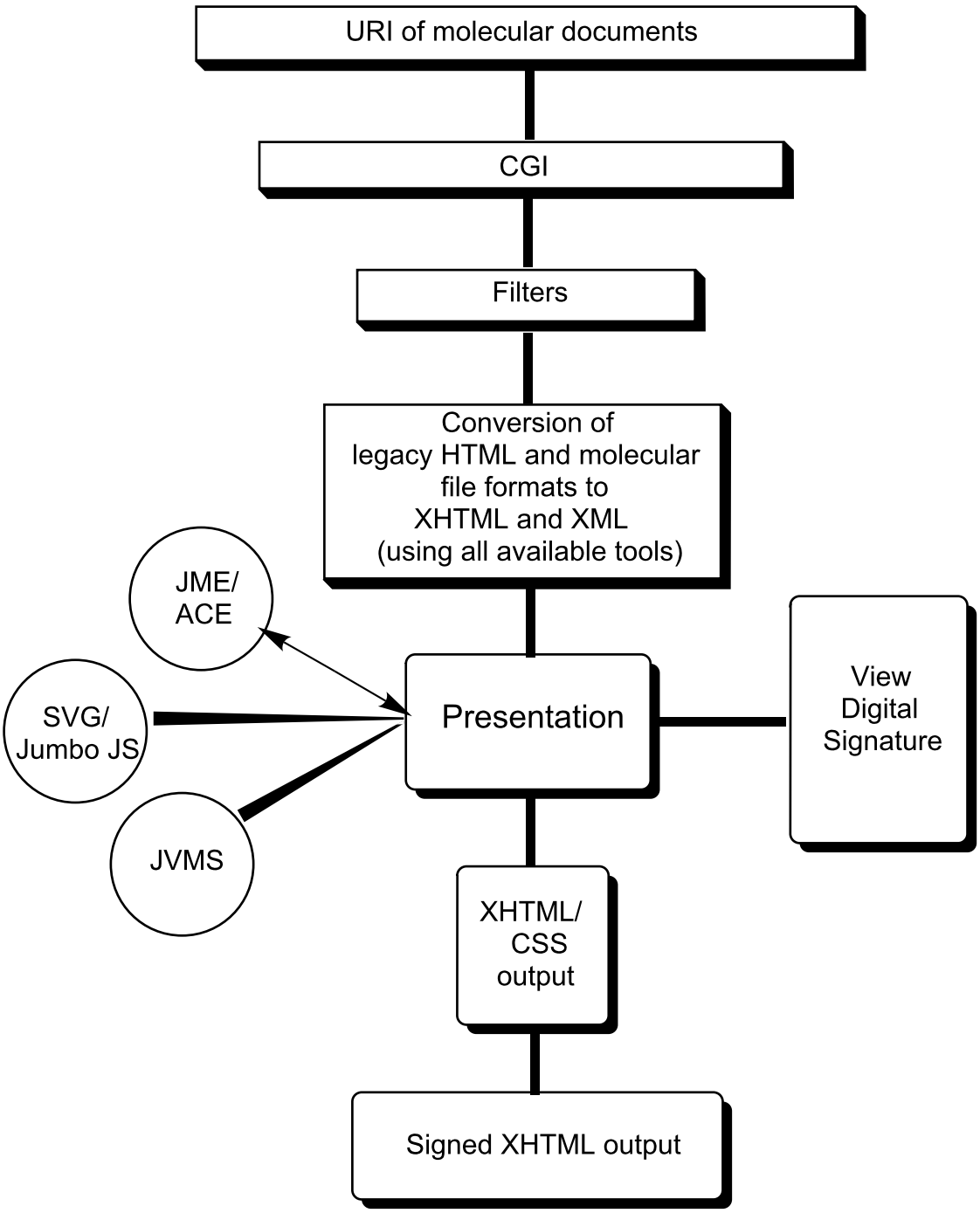

Figure 3. Molecular data in Molfile format, transcluded into the HTML version of this document using the <embed> syntax, and rendered using an appropriate plugin such as *Chime* .

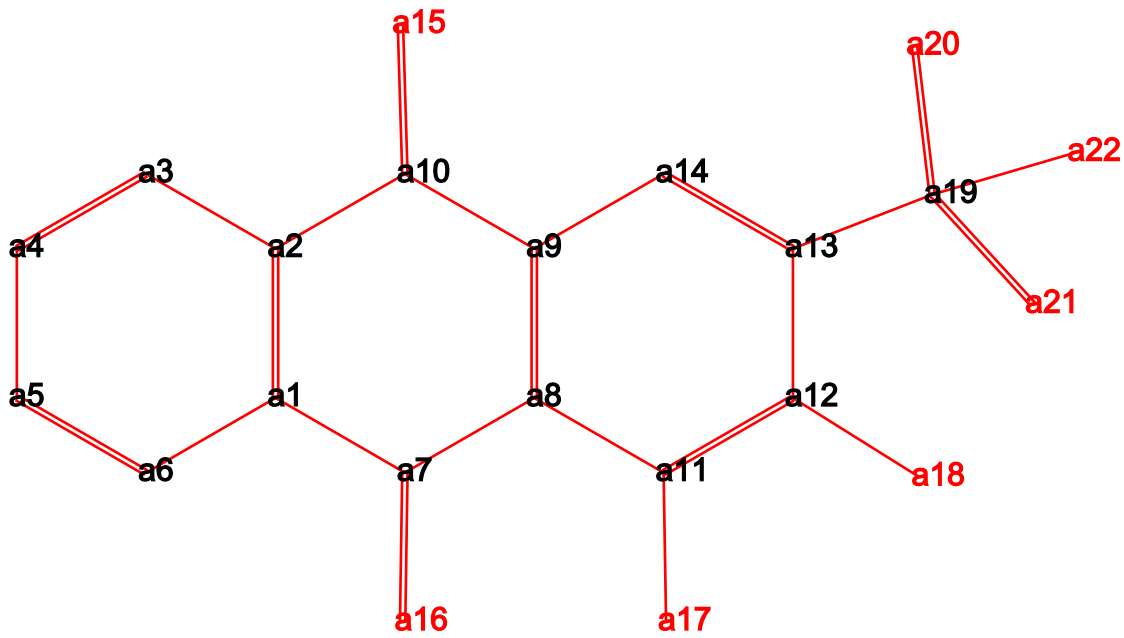

Figure 4. Illustration of the result of reading the XML document corresponding to this article into the JUMBO3 browser (available for download from *http://www.xml-cml.org/*)

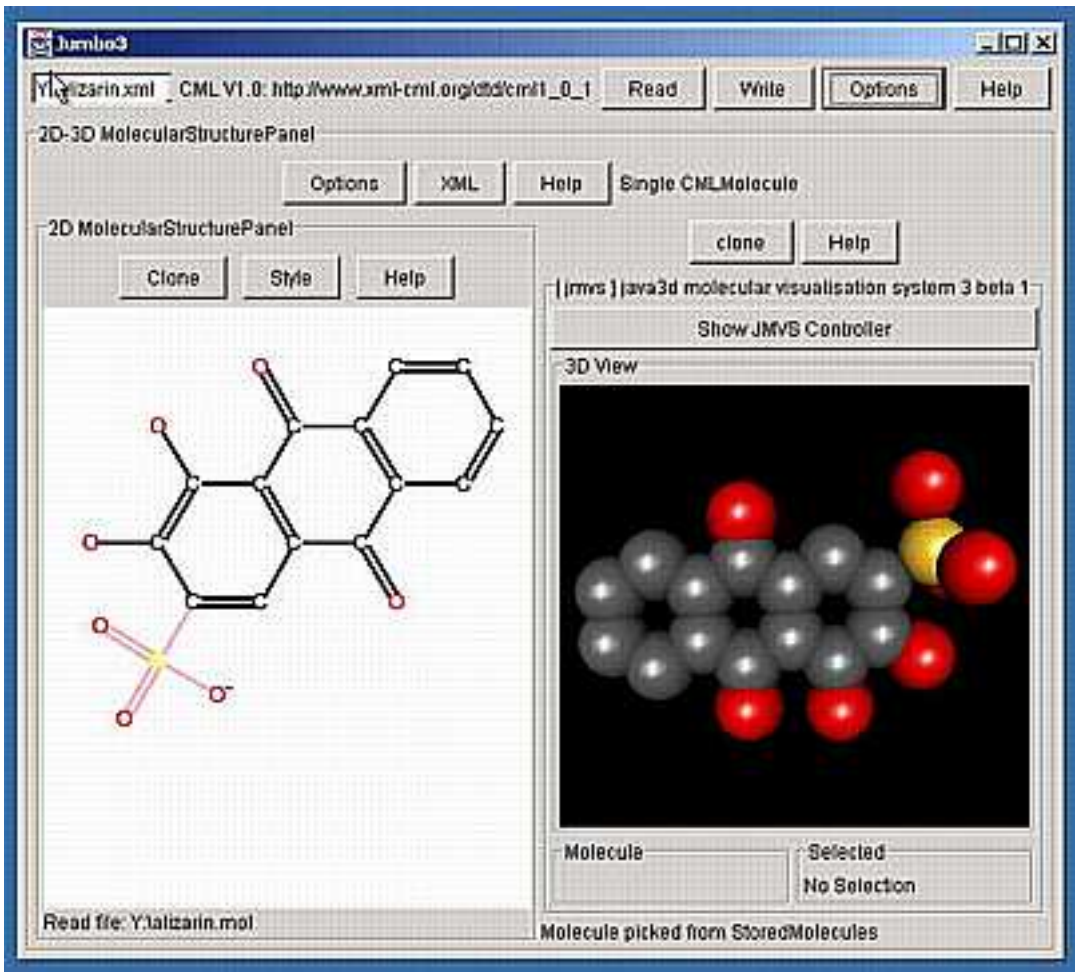

Figure 5. An eXtensible Annotating Chemical Editor (XACE) displayed within the Internet Explorer Browser and using a combination of XSLT defined transformations and the JME editor to display and edit metadata associated with the molecule. Some fields in the metadata are entered as annotations by the authors, others are quantities derived using the XSLT transformations. XACE is available at *http://www.xml-cml.org/*