Context Slicing the Chemical Aether

Jeremy G. Frey[1], David De Roure[2], monica m.c. schraefel[2],
Hugo Mills[2], Hongchen Fu[1], Sam Peppe[1],
Gareth Hughes[2], Graham Smith[2], Terry R. Payne[2]

[1] School of Chemistry, University of Southampton, UK
{ j.g.frey, h.fu, s.peppe}@soton.ac.uk
[2] School of Electronics and Computer Science, University of Southampton, UK
{dder, mc, hrm, gvh, gms, trp}@ecs.soton.ac.uk

Abstract

In this paper we discuss the information system requirements of an e-Science scenario, drawn from chemistry research, and describe our approach to a solution which employs notions of adaptive (and adaptable) hypermedia and utilises Semantic Web technologies.

Keywords Semantic Web, Semantic Grid, Adaptive Hypermedia

1. Introduction: The Chemical Aether

The typical chemist operates in a space that contains a rich variety of information, equipment and services many of which require interaction with other researchers. It is useful to consider this space as a pervasive environment, which we refer to as the "Chemical Aether". Like the conventional aether, which was the medium that was thought necessary to support the propagation of light waves, the chemical aether is a useful construct from which chemists in particular roles draw out the services and information they need to satisfy that particular role. That is, the information and services chosen and their appearance depends on the context.

The components of the chemical aether do not necessarily have to exist in the pervasive environment – they can have an independent existence. Placing them in the context of the aether, however, gives a whole which is greater than the sum of its parts. It can however lead to problems when we step outside the conventional discipline boundaries to consider multi and interdisciplinary research (which of course this paper is an example). This is a non-trivial observation as such multi and interdisciplinary research areas are currently some of the most prolific. Proper pursuit of these areas requires that different disciplines can share and understand not just "results" but frequently procedures and limitations. Thus any models we build to facilitate efficient recording of a chemical study (including for example the planning, procedures, analysis and publication) which needs to be achieved in such a way that broadens the contexts that can appreciate the material rather than restrict it further to only those with discipline specific knowledge. Clearly the presentation of the material to those experts even within the chemical context will need to differ from the presentation to less expert users. The important point is that within the chemical aether the necessary material has been captured and is available if required.

In this paper we discuss the requirements to undertake chemical research and consider the advantages of adopting an adaptive (or adaptable) hypermedia approach using semantic web technologies. In section 2 we introduce the chemistry context and discuss two scenarios which together provide a comprehensive set of requirements for this exercise. This is followed by a discussion of our hypermedia and semantic web approach in sections 3 and 4. In section 5 we provide a summary of issues.

2. The Chemistry Scenarios

2.1 Background

2.1.1 Nature of Chemical Information.

The information required by chemists in order to carry out their usual activities is very broad. It encompasses not only molecular structures and properties but the panoply of processes that have been developed to utilise and transform these materials (similar statements can be made about what is done with the data about the molecules). It differs from other disciplines. For example, in particle physics experiments, terabytes of data are generated daily but the data always has the same form and there are relatively few different types of experiment being undertaken at any one time. In contrast the chemist's data is highly variable, frequently multimedia and, with the rapid evolution of methods and experiments, the procedures that perhaps would normally be considered metadata are elevated almost to the level of data. This metadata is often more important to the wider community of chemists than the particular result obtained in any

one application as it aids in repeating the experiment.

2.1.2 Chemists.
Chemistry is a very broad discipline and not all chemists are the same.[1] Similarly, not all chemical investigations are the same. For the purposes of this article we present a number of archetypal chemists and activities. The synthetic chemists' main aim is to synthesise new compounds and to find new, more efficient routes to existing materials. In contrast, physical chemists are concerned with the properties of these materials. This leads to somewhat different styles of working in the laboratory. Both these groups have an intrinsic belief in the repeatability of their systems; though as every student knows the reproducibility increases remarkably with increased understanding of the system. The issue of reproducibility does tend to mark out another group, the biological chemists, who frequently work on systems which could be described as significantly more complex and intrinsically less reliable, which places a different load on the planning of data capture in the laboratory.

2.2. Example Contexts: The Laboratory Context and the Publication Context

As part of the Comb-e-Chem e-Science Pilot Project**,** several aspects of translating chemical practice into the chemical aether are being undertaken. We focus on two here:

1. The "Smart Tea Project" which relates to what is probably perceived by the general public as "what chemists do". This is looking both at how to support and how to capture the processes going on when a synthetic chemist, working at the bench, aims to make a compound.

2. The Second Harmonic Generation (SHG) project, which considers the process of publication for chemists, or more particularly how to store and retrieve archived material for subsequent analysis and research.

In each case, services are designed not only to aid access to data after the fact but to enhance the practice of that data's production via appropriate interaction, data and service design.

2.2.1 The Lab Context and the Smart Tea Project

The Smart Tea Project [20] is not to be confused with other research [1,3,18] efforts which have attempted to digitize the lab book, the main recording device of the traditional lab environment. By focusing on the book rather than the environment, and thereby attempting to replicate paper-based practices, these projects have often meant higher cost to the scientists for input with limited gain beyond some search/retrieval over the data. Learning from these problems, our approach has been to consider each part of the lab experience that generates or captures data and convert those parts to services. The data from these services can in turn be made available wherever and however in the lab it is appropriate for the chemist's current context.

2.2.2 The Publication Context.
Smart Tea provides us with a model of some of the initial phases in the generation of chemical information, though as will be apparent there is a significant amount of cross referencing, with the planning and safety stages drawing considerably on existing information and insight.

The SHG project takes the results from an existing Physical Chemistry experiment, (i.e. a set of much more numerically-focussed information) for which the analysis forms a sequence of increasing levels of abstraction. It derives information about a medium, and aims to hold this material in such a way as to make "Publication@Source" realisable [9]. The focus here is on the "Publication Context" or perhaps more generally the "Dissemination Context" to emphasise that the different methods of publication are simply different views, with different levels of detail, of the same underlying data.

The dissemination context requires active semantically rich hypermedia documents that are the modern rendering of a traditional scientific document. When interrogated by some users the context delivers the high level "journal type paper" but with a shift of context the document would be transformable to the highly detailed laboratory context (or the reverse). For example a new student might be attempting to understand the work in sufficient depth to repeat and extend the work. To be more specific, information at the higher levels of abstraction are linked via the analysis that generated them to the data used in this generation process, not simply within the document itself but to the archived material. Seen in this context the journal paper is simply a high level context looking on to the virtual world of the laboratory (and its occupants) and the analytical processes which generate, integrate and transform the data. Such a fully integrated representation that reflects this integration implies the chemical aether that can be sliced, on demand, to reflect the various contexts of

[1] *though they usually like beer and are quite good cooks*

the chemical cycle. A *smart laboratory* generates meta-data rich data viewed by *smart documents*.

2.2.3 Provenance.
The above integrated approach smoothly delivers much of the necessary provenance of the data and the processing of that data. This goes right back into the laboratory, as the smart laboratory context understands the issue of who is doing what, and indeed uses this to enable inference to associate materials, information, people and projects. The requirements for various applications of provenance information are not fully understood at this time and, together with general issues of trust, will be further investigated through this work.

2.3 Examples of People-Centred Contexts for Chemistry Scenarios
Chemists play different roles at different times. To make this discussion more concrete it is worthwhile to provide some example scenarios. Consider a research student in a synthetic chemistry laboratory about to embark on a new project. The stages that need to be covered might be the following:

- Proposal
- Planning
- Safety assessment
- Laboratory Synthesis
- Chemical analysis of products
- Integration of procedures and results
- Analysis & Report
- Discussion & Review

Each of these provides its own context and places different requirements on the type of material and information that is needed (from internal laboratory and external literature sources).

The research supervisor also has their own context for looking in a different way at the same experiment. The context is similar, with considerable involvement in the Planning and the Discussion and Review phases, but for an experienced student less involved with the details of the other contexts (but not their existence, especially the safety!) However should the results of the experiment be unexpected in anyway (success or failure) then the supervisor context will deepen with the need to review the details of the procedures. The adaptive and adaptable hypermedia system needs to be able to adapt to the following –

Student: "The procedure did not work as well as expected" or "I can't model all the results"
Supervisor: "Exactly which procedure did you follow" or "Which model did you apply" {high level context}
Student: "Smith's Procedure" "Langmuir model by Least Squares" {Still at high level context}
Supervisor: "Show me exactly what you have tried" {Switch to much more details context to see the laboratory record or the analysis workflow}

When investigating the literature, the level of details needed about the work depends again on the context. We can readily distinguish several contexts

- Referee of original work
- General interest for ideas
- Need to obtain data for personal analysis. Often the data important to you may not be an item that was considered to be important at the time the high level abstraction was originally generated and thus we see the need to be able to delve down. Frequently in paper publications a graph is presented that another user would like to use for an alternative analysis and he or she has resorted to scanning in the graph.
- Need to understand the methods to apply them to similar or different studies
- Possible re-analysis using new techniques
- In extremely unusual cases the need to check for fraud

It should be apparent that the type of information, the level of abstraction and detail that will be required in these different contexts could be quite different and variable.

2.4 Translating the Lab Process to the Chemical Aether
In the following section, we present in closer detail first what the chemist does in planning and carrying out an experiment. Then, we present how the Smart Tea project converts these phases into services which in turn translate the generated data into information that can be used on demand by other lab contexts.

The Experiment, from Plan to Execution

1. A chemist plans an experiment, which means determining the chemicals to be used to create a compound. In keeping with safety requirements, the plan must explain what chemicals are to be used, what processes will be performed with them, and whether or not there are any hazards associated with the chemicals/procedures being used. The safety plan must then be approved.

2. A form of this plan is used throughout the rest of the experiment – it acts as a reminder for selecting the predetermined set of chemicals

from stores, for specifying amounts to be measured, for determining the process to be followed in the lab, and for considering the success or failure of the result. The process itself may be informed by reference to previous experiments either within the lab or from the published literature.

3. With respect to potential intellectual property claims, on completion of the experiment, the results are to be signed off by a secondary source.

4. On completion of the experiment, analysis is performed on the results. The writeup of the experimental results and analysis can then form the basis of a range of publications, from a report to a supervisor to a journal publication.

Currently, the bulk of the above work – the safety form, the experimental plan design, the recording of data, musings on analysis – is carried out via paper: A safety form is written up and submitted for approval of the experiment; the relevant data from the safety plan is then re-recorded into the lab book where the rest of the results of the experiment are also recorded. The disadvantages of having data recorded only on paper are obvious: they cannot be readily accessed, shared or searched. Indeed, historically, the loss of access to or knowledge of previously performed work has caused bodies of work to need to be repeated.

As noted above, making data from experiments sharable and searchable has been the motivation for numerous "eLab Book" projects which have attempted to digitize the book. In looking at the problems encountered in these projects, our approach has been to break with the book, and instead to distribute the generating and recording of data among contextually appropriate services throughout the experimental process. The data from these services can then be combined on demand from the aether into whatever context is appropriate for the chemists.

As an example of how these distributed services work in Smart Tea, consider the following part of the experimental process listed in the above overview: the measuring of chemicals.

In our aether-aware lab environment, the chemist prepares an experimental plan. The plan has a list of chemicals to be used, along with their planned weights and their values in moles (this conversion from weight to moles has been provided by a conversion service from the aether). The plan also provides a template of the steps of the experiment that the chemist can refine as appropriate. The appropriate information from the plan is submitted automatically for safety approval to the appropriate authority; on approval, the information is available for reuse throughout the lab. One such context for reuse is the weigh station where the scales are connected to the lab aether via a scale service. The scales are also connected to a barcode scanner. Chemists can pass a container of the chemical to be used in front of a scanner. This brings up on the scale's display a list of planned experiments in the lab using that chemical. The chemist selects their experiment from the list[2]. The display then shows the chemist the part of the plan appropriate for the scales: the list and amounts of *dry* chemical to be measured for this experiment. Once the chemist measures the chemical, the amount is recorded both for the experiment, and for stores. When the chemist moves from the scales to the workbench area, the chemist selects the current experiment from the Bench station. There, the amounts of measured chemicals from the scales are displayed, along with the list of any wet chemicals to be measured at the bench. The process to be used in dealing with each is also articulated.

The above slice of the chemist's cycle – measuring out chemicals – shows the integration of context aware services. The scales service polls active plans in the lab for chemicals to be used. The selection of a particular experiment refines a context for that selection further. When the experiment is selected, it provides information to the chemist about other chemicals required and their planned amounts. It publishes the recorded amounts back to the plan so that these amounts can be reused at the bench where the experiment is carried out. Stores services listen for amounts published as well so that replacements can be ordered. This data can also be associated on demand with other services such as a lab history to see who was doing what when in the lab on a given date and what any of the ambient readings in the lab were at that time. Similarly, an unexpected result in an experiment can be traced back to a specific batch of a chemical which can then be followed up with the producer.

With the semantically informed lab aether, all of this information can be generated and captured transparently *without asking chemists to change how they work*. Right now, outside an aether-based lab, little of this information is being captured except for the name of the chemical and the amount weighed, and this only in the chemists' lab book. While digital lab and elab book projects attempt to

[2] Note, there are many ways of possibly identifying oneself to the scales or any other instrument in the lab – RFID tags, retinal scans, personal bar codes. The technique of identification of person to plan is incidental next to the fact that (a) it is possible to do this and (b) the cost of doing so is lightweight enough to make the effort worthwhile for the process as a whole.

make at least the data captured by a lab book sharable, the cost is high, with the result being that the scientist either duplicates information stored on a computer and pastes it into a lab book, or must scan material from the lab book into the computer system. Our break-the-book approach has meant data is simply captured or generated at source and is available for re-view in other associated contexts on demand.

The above scenario has brought into the foreground the role of services in the lab for data capture. In the following section, we consider another example scenario of the chemical aether, this one focusing more on its adaptive hypermedia aspects. In section 3, we then look in more detail at how these examples are informed by adaptive and adaptable hypermedia and semantic web services. We then present how we are proposing to build this lab Aether within the context of the Comb*e*chem/Smarttea projects.

2.5 Workflow and Second Harmonic Generation

The Second Harmonic Generation (SHG) experiment is being used as a test to understand how we can capture and use the workflow of a modelling process. Currently for the SHG we have the data (in a database) and the analysis steps are articulated essentially in excel spreadsheets, one sheet for each data analysis but only a few types of sheets each type for a different type of model. So we will link to the excel sheet from the database to provide the model specification for the current analysis. This happens to be the way the analysis has been done; in the future we would aim to use a more services pipeline approach. At each stage it is important to remember the links in/out to the literature that are important in specifying models, necessary additional data, assumed parameters. Currently this is not an easy thing to keep track of, and is another area in which placing the whole activity within the Chemical Aether will enable automatic tracking of the external links.

It is clear that we need a very powerful linking system to provide the links from the analysed data back to the original data and processes. It is unlikely that all these links can be generated at the same time the data is generated since the links needed depend heavily on the chosen context. As a context is produced and used links will be generated, some by inference others by the action of the user in drawing down materials and using services. In order to develop such a linking service, we are currently working through the following questions:

(a) How to represent these links, what underlying structure to hold the links. It appears we have a choice: database or triples?

(b) How to add the links,

(c) How to use the links? See below for the comments on URI's.

For the SHG experimental data we have implemented a database for the experimental results (thus allow provenance trace back to the data). Soon the database will hold results of various stages of the model building process (workflow). Our current method seems inflexible and it is not clear it can capture the Student Centred Context.

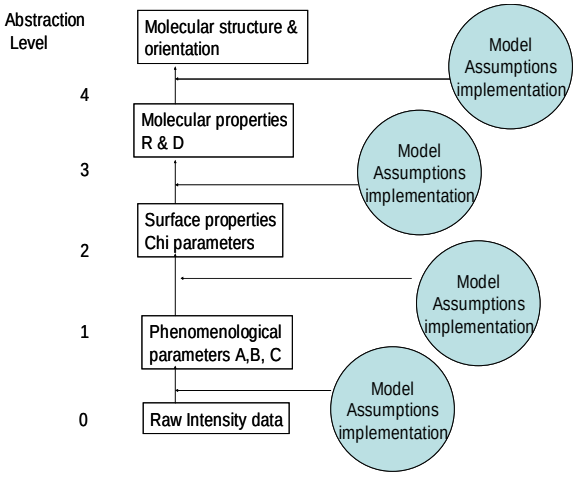

Figure 1. Chemical Metadata

2.5.1 Planning, database and schema issues

In our work on the SHG experiment we found that the rapid evolution of the experiment and inconsistencies in the data collection and format added significantly to the difficulty of the data base construction and use. This contrasted to the work associated with the National Crystallography Service (NCS). This EPSRC service takes in samples and determines (if possible) the structure by single crystal diffraction, returning a result in an internationally accepted standard. The service has a high turn-over but while considerable expertise is required to obtain results from difficult samples the procedures and the forms of the raw and analysed data are the same. This justifies detailed work on the construction of efficient databases for sample tracking and data archives (with all appropriate metadata some of which needs to come from the service users).

The apparent issue has been the ability to evolve the database design to keep up with the experiment. This may partially be exacerbated by applying this approach to an existing experiment with somewhat entrenched (as much as they were effective) methods. On reflection the problem lies more with the required evolution of the data schema since as long as an appropriate schema exists the construction of an adequate (not necessarily

efficient, but fit for initial purpose) data store is relatively automatic. Given the schema, an efficient alternative compatible with the initial system can subsequently be generated without fear of incompatibility. The design of the schema is, in reality, the issue rather than the actual database construction.

2.6 Approach to Experimental Methodology

If the approaches recommended here are to be followed it implies a need to shift viewpoint to the construction of a data schema prior to the design of the experiment. This is, of course, standard practice for many large scale projects, but not always for smaller scale, more readily repeatable experiments. Perhaps this contrasts with the social sciences where the inability to repeat the same experiment means that much greater explicit emphasis is given to methodology in developing experimental designs. However this should not be pushed too far. Chemical investigations are usually well planned and the necessity for safety considerations (with for example a legal drive from the Control of Substances Hazardous to Health – COSHH) mean that most investigations are only started after a detailed plan considering possible outcomes has been produced and approved.

Such a plan provides a very useful framework on which to provide a context for the synthetic chemist as they are undertaking their work. The context draws together a series of other contexts (such as The Bench, the Fume Hood, the weighing machine etc.). The lab example above shows how effective this approach can be while still allowing for flexible response to unusual observations.

The suggestion is that good practice in regard to process needs to be more formally extended to the consideration of the form as well as the content of the data obtained or derived. This fits in well with the view of statisticians that researchers should know how they intend to use the data before collecting it to ensure they know what data to collect. It also allows for much better "Design of Experiments". A statistical concept that enables the optimal design of which experiments to cover a range of parameters is necessary to obtain not only estimates of the responses but understanding of the uncertainties sand correlations involved.

2.7 Summary of Requirements

Based on these two scenarios, we have identified the following requirements:

1. Working with scientific data – generated and legacy data, rapid evolution of data schema.
2. Need to include data planning as part of the experiment plan and evolve the schema in concert.
3. Rich & variable data sources.
4. Annotation.
5. High context sensitivity for services and data reuse.
6. Aim to allow non-specialist users to understand and manipulate the information.
7. Planning & workflow prescriptive, supportive and retrospective (for discussion and repeatability).
8. Information is only generated if the meta data is attached thus whole range of metadata capture facilities are required.
9. Real time or concurrent generation of data & metadata.
10. A system must cope with no one run being the same as any other (which is different to industrial production).

3. BUILDING THE CHEMICAL AETHER: THE ROLES OF ADAPTIVE/ADAPTABLE HYPERMEDIA AND THE SEMANTIC WEB/GRID

3.1 Adaptive and Adaptable Hypermedia

Adaptive hypermedia systems such as AHA, FOHM and Linky [4,17] support context sensitive delivery of content. In the case of the weigh scales in the lab description, for instance, the context is first the scale: this determines which data from any plan is appropriate to have available at that site. The second context is the experiment itself. This determines which specific data of the given set is appropriate for display. The information represented to the user is determined by the contexts at play; the system *automatically* determines appropriate information for that context, providing links to the appropriate data. The adaptation is informed by a user- or domain- based model for the given context.

Adaptable hypermedia on the other hand [19] supports *user-determined*, rather than system-determined adaptation of context. This kind of adaptation can best be seen in the publication scenarios. In that case, an adaptable hypermedia representation of the experiment might support how much detail on a particular point is viewable. For instance, if the chemist says that "this process is described in paper X," a link may be available which lets the user decide if they wish to see the description of that experiment from paper X right at that location or simply open the paper itself. The reader may also wish to select a particular version of the paper: an executive summary, the complete document, the analysis only, and so on. Both

models of adaptation are necessary within the aether, depending on the context of use. The advantage of thinking about representation of information in terms of adaptive/adaptable hypermedia is that there is a rich interdisciplinary body of research – from artificial intelligence to user interaction to education – to draw upon, saving the chemistry community from reinventing the wheel, at least entirely.

3.2 Open Hypermedia, Link Services and Generic Links

Within the open hypermedia research there is also a well defined body of research and models of link services and generic links which enrich Adaptive Hypermedia. An adaptive link service [7] provides links to resources based on what the system determines as the user's current context. If the chemist has been looking at experiments relating to a particular compound, the link service would build a list of links that relate to resources on this subject. The chemist then has the opportunity to investigate associated resources without having to initiate a specific search. Similarly, in a more adaptable approach, the user can indicate to the system the specific concept of interest, and the system will generate appropriate links based on its knowledge of this concept. In flexible systems, if no links are available, chemists can author their own links to be maintained and associated with a given concept.

Several of these link service approaches have already been adapted within the semantic web service-based approach, such as COHSE [5] and Magpie [13].

3.3 Semantic Web Services

Hypermedia systems research has been largely focused on the representation of and association of documents to each other. The Semantic Grid [6] on the other hand is involved with the representation and association of *services*. This raises an interesting challenge: to what extent can adaptive hypermedia techniques be employed with services; conversely to what extent can Semantic grid techniques (such as service composition through workflow) be applied in the information systems context? These services are supported by metadata-rich information that allows machine reasoning over the service/data space.

The DAML Service ontology (DAML-S) [1] is emerging as the *de facto* standard for describing *Semantic Web Services*. It is based on DAML+OIL (which extends RDF(S)) to provide an expressive means of encoding semantics across a distributed, open environment, i.e. the semantic web, and describes services from the perspective of discovery (*Service Profiles*), workflow choreography (*Process Models*) and invocation (*Grounding*). Semantic Web Services are defined in terms of their capability descriptions (i.e. interface signatures that define its inputs and outputs, and AI based statements relating to preconditions, and effects), and extensible metadata representing concepts such as service category, security and Quality of Service parameters.

As well as supporting service discovery, the *Profile* provides the initial metadata that is used to negotiate Grid-based service level agreements [13] or performance parameters prior to invocation. Services advertise themselves by means of a service repository [16,22], which may provide several query based mechanisms such as yellow or white page lookup, or capability search. The *Process Model* provides a description of the process decomposition of a service. This decomposition consists of several process, each corresponding to an invocable process (e.g. Web Service operation), a real-world action (performed by a human user and mediated by a user-interface), or an abstract process description (representing another service that should discovered at run-time through a service repository).

4. SEMANTIC WEB APPROACH

4.1 Philosophy

We have adopted a Semantic Web approach based on the following principles:

1. Use of the Resource Description Framework (RDF) to represent associations between entities (e.g. resources and services), as opposed to (but informed by) other link implementations which have been established in previous open hypermedia projects.
2. Adoption of shared, unique URIs, together with the adoption of RDF. In this way, the various forms of metadata are effectively linked together by the entities that are described.
3. The use of DAML-S Profiles for service descriptions, for locating services or data repositories.
4. The use of DAML-S Process Models for describing experimental workflows in terms of actions and services

4.2 Progress

The Smart Tea Project has begun to model and implement the lab scenario described above. To this end, we made tea – literally – as a chemistry experiment. This has helped especially the non-chemists on the project first, to understand the

experimental space itself by analogy to a well-understood (if highly debated[3]) process, and then to reason about both the services (defined as semantic web services) and their representations (modelled as hypermedia documents) that we could define for this space. From this, we have worked with chemists to determine the services we would require for an experimental planner and the various stations/services that would be needed throughout the lab. We validated the walk though of this model with chemists in a design review. We then validated our tea-based model by moving from the tea-analog to a published chemistry experiment for making aspirin [21]. The resulting RDF graph [1] from the aspirin process successfully matched with our experimental model for tea. From this result, we have begun to build up our lab aether ontology for use by the lab services.

We have conducted a study of the information system requirements of the Second Harmonic Generation experiment by prototyping the system using established database technology within the chemistry department. Based on this we are in the process of remodelling the system using an RDF triplestore (Jena).

We have already built some preliminary services that will support the experimental planner. These services provide computational tools or integrate with auxiliary systems. For instance, we have converted an existing simple inventory database used by the chemists into a web service. Coupled with technologies such as barcode scanners, we can begin to provide integrated inventory management support. Another service provides many useful standard pieces of information about a chemical and a simple service provides all of the data of a normal periodic table. All are used by a grammes–to-moles chemical calculator. During our planning phase we found that no such systems were available from national services or major chemical suppliers. We are now in the process of using these building blocks to construct the higher level scales interface system and planner application.

The above phase of the work is largely focused on making services available to support carrying out the experiment. As such, this first phase of the project is oriented more towards the semantic web and user interaction challenges of the project, some of which are further articulated in section 5, below. In the next phase of the project, we will be engaging more of the adaptive/adaptable questions in the project, as outlined above in the publication scenarios.

[3] See smarttea.org for a list of links to documents debating when milk is to be added or the type of vessel to be used.

5. CONCLUSIONS AND FUTURE WORK

These are some of the challenges that we have identified through the work so far:

- Automated authoring and automated adaption for user determined rendering of the information. We have some experience of this with tools such as Linky but they have not been tested on large scale repositories or for dynamically determined interaction.
- The form and interface of the experiment planner.
- The need for distributed triplestores for scalability and interoperability.
- Timestamping of events and time-dependent views of data.
- The design of messages that will transport data from interface systems to triplestores.
- The generation of unique, shared URIs for the chemical aether.
- Triplestore technology in the context of rapidly evolving data.
- Discovering and choreographing chemistry-based services.
- Trust and security issues.

Acknowledgements

This work is supported by the Comb*e*chem project (*Structure-Property Mapping: Combinatorial Chemistry and the Grid*), grant GR/R67729/01 from the UK Engineering and Physical Sciences Research Council. We wish also to thank the chemists from the synthesis lab of Dr Martin Grossel.

References

1. Ankolekar, A., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., Martin, D., McIlraith, S., Narayanan, S., Paolucci, M. and Payne, T. and Sycara, K. DAML-S: Web Service Description for the Semantic Web. First International Semantic Web Conference (ISWC) Proceedings, pp 348-363, 2002.

2. Arnstein, L., Borriello, G., Consolvo, S., Hung, C., Su, J. 2002. Labscape: A Smart Environment for the Cell Biology Laboratory, IEEE Pervasive Computing Magazine, vol. 1, no. 3, July-September 2002, IEEE Computer Society, NY, NY

3. Arnstein, L. F., Grimm, R., Hung, C, Hee, J., LaMarca, A., Sigurdsson, S. B., Su, J., Borriello, G., Systems Support for Ubiquitous Computing: A Case Study of Two Implementations of Labscape, Proceedings of the First International Conference on Pervasive Computing, 2002, Springer-Verlag, Germany

4. Bailey, C., Hall, W., Millard, D.E. and Weal, M.J. Towards Open Adaptive Hypermedia. In De Bra, Paul and Brusilovsky, Peter and Conejo, Ricardo, Eds. Proceedings 2nd International Conference, Adaptive Hypermedia and Adaptive Web-Based Systems 2347(-), pages 36-46, Málaga, Spain, 2002.

5. Carr , Leslie, Wendy Hall, Sean Bechhofer, Carole Goble Conceptual linking: ontology-based open hypermedia. Proceedings of the tenth international conference on World Wide Web, April 2001, 334 – 342.

6. D. De Roure, N. Jennings, and N. Shadbolt. "Research Agenda for the Semantic Grid: A Future e-Science Infrastructure". Technical report UKeS-2002-02, UK e-Science Technical Report Series, National e-Science Centre, Edinburgh, UK. December 2001.

7. Deconstructing Tea. http://www.smarttea.org/aspirin.html

8. El-Beltagy, S.R., Hall, W., Roure, D.D. and Carr, L., Linking in Context. In *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, (2001), 151-160.

9. Frey, J. G. and De Roure, D. and Carr, L. A. Publication At Source: Scientific Communication from a Publication Web to a Data Grid. EuroWeb 2002 (The Web and the Grid: from e-science to e-business), Oxford, UK. 2002.

10. Goble, C. and De Roure, D. The Grid: An Application of the Semantic Web. ACM SIGMOD Record, Volume 31, Number 4, December 2002.

11. Harter, J., Y. Zhang, R.C. Glen, and P. Murray-Rust, The World Wide Molecular Matrix - a peer-to-peer XML repository for molecules and properties. EuroWeb 2002 (The Web and the Grid: from e-science to e-business): p. 163-164, Oxford, UK. 2002.

12. Hendler, J. Science and the Semantic Web, Science 2003 299: 520-521

13. Lawley, R., Decker, K., Luck, M., Payne, T.R. and Moreau, L. Automated Negotiation for Grid Notification Services. International Conference on Parallel and Distributed Computing , 2003

14. The Magpie Semantic Filter, Open University, UK. http://kmi.open.ac.uk/projects/magpie/.

15. Michaelides, D.T., Millard, D.E., Weal, M.J. and De Roure, D. Auld Leaky: A Contextual Open Hypermedia Link Server. In Proceedings The 7th Workshop on Open Hypermedia Systems, ACM Hypertext 2001 Conference, Aarhus, Denmark. 2001.

16. Miles, S., Papy, J., Dialani, V, Luck, M., Decker, K., Payne, T., and Moreau, L. Personalised Grid Service Discovery, Nineteenth Annual UK Performance Engineering Workshop (UKPEW'03), University of Warwick, Conventry, England, July 2003.

17. Millard, D., Aben, K.O., Davis, H.C., De Bra, P. and Weal, M. AHA! meets Auld Linky : Integrating Designed and Free-form Hypertext Systems. Hypertext 03, Nottingham, UK. To appear. 2003.

18. Myers, J.D. 2003. Collaborative Electronic Notebooks as Electronic Records: Design Issues for the Secure Electronic Laboratory Notebook (ELN). In Proceedings of the 2003 International Symposium On Collaborative Technologies and Systems (CTS'03), January 19-24, 2003, Orlando, Florida

19. schraefel, m.c., Contexts: Adaptable Hypermedia. In *Adaptive Hypermedia and Adaptive Web-Based Systems International Conference, AH 2000*, (Trento, Italy, 2000), Lecture Notes in Computer Science, 1892, 369-375.

20. Smart Tea Project web site. http://www.smarttea.org.

21. The Synthesis of Aspirin. Chemistry Department, California State Universit at Stanislaus http://wwwchem.csustan.edu/chem1002/aspirin.htm

22. Wrore, C., Stevens, R., Goble, C., Roberts, A. and Greenwood, M. A Suite of DAML+OIL Ontologies to describe Bioinformatics Web Services and Data. International Journal of Cooperative Information Systems, Vol. 12, No. 2 (2003) 197-224.