Towards A Rich-Context Participatory Cyberenvironment

Yong Liu, Robert E. McGrath, James D. Myers, Joe Futrelle
National Center for Supercomputing Applications
1205 West Clark Street
Urbana, Illinois 61801
1-217-265-7640
{yongliu, mcgrath, jimmyers, futrelle}@ncsa.uiuc.edu

ABSTRACT

To enable and support innovative research in science and engineering, the next generation Cyberinfrastructure must be able to support collaboration across disciplines and conceptual contexts. At NCSA, we are building Cyberenvironments which support "architecture of participation" where user-driven innovation is empowered. In this paper, we will first describe the Cyberenvironment and Web 2.0/Where 2.0 concepts, and present our definition of a participatory Cyberenvironment and the roles of contexts for building such Cyberinfrastructure. We then present our arguments of the importance of supporting the full range of social, geospatial, causal and conceptual contexts. We will describe the foundation work that we have built so far, the CyberCollaboratory (a collaborative portal) and Tupelo (a semantic content repository), and then provide the vision for the path towards a rich context participatory Cyberenvironment with potential impact on scientific communities such as distributed environment observatory networks.

Categories and Subject Descriptors

D.4.7 [***Operating Systems***] Organization and Design - *Distributed systems*
H.5.3 [***Information Interfaces and Presentation***] Group and Organization Interfaces - *Collaborative computing*
K.3.1 [***Computer Uses in Education***] *Collaborative learning*

General Terms

Design

Keywords

Cyberenvironments, Cyberinfrastructure, Collaboratory, Context, Geospatial Context, Social Context, Causal Context, Conceptual Context, Participatory, Architecture, Web 2.0, Where 2.0, Collaboration, Virtual Organization

1. INTRODUCTION

The ability to provide community scale infrastructure while enabling innovation by individual researchers is a central challenge for Cyberinfrastructure and e-science efforts. The very nature of a scientific community is increasingly tied to collaborations that span disciplines, laboratories, organizations and national boundaries. Such activities involve creating and consuming digital artifacts using complex processes. Effectively working with heterogeneous resources such as sensors, software components, databases, scientific instruments, networks and people requires substantial contextual information which represents knowledge about the Cyberenvironment and its users.

The National Center of Supercomputing Applications (NCSA) has initiated efforts in building end-to-end Cyberenvironments that provide flexible middleware with semantic contexts [37]. Examples of such semantic contexts include how, when, where, why the scientific data were generated and used and who is related to or responsible for these processes/activities. These are metadata which describe the conceptual relationships among different artifacts (e.g., provenance describes the causal relationships, which answer the "why"-questions.).

The CyberCollaboratory [44], building on a traditional collaboration portal, is evolving these concepts to develop a *participatory Cyberenvironment*, inspired by and built on the Web 2.0, with a focus on supporting distributed environmental observatory networks. Traditional science gateways focus mainly on providing user access to data and computing resources such as the TeraGrid User Portal ([55]), which are built with relatively difficult to use interfaces (e.g., the OGSI [57] and WSRF [1]), and usually do not provide extensive social networking interaction or social context. Science collaboratory projects, such as [2, 8, 44] have shown that collaboration on contexts can be quite useful. Portals such as Nanohub [42] show that integrating social networking/tagging capabilities into Cyberinfrastructure [10, 59] strengthens scientific collaboration and promotes sharing of the rich knowledge networks. The NCSA CyberCollaboratory combines these features, along with Web 2.0-style APIs and formats, to create a new, participatory Cyberenvironment, which can be extended and used by many communities.

This paper is organized as follows: we will first describe the Cyberenvironment and Web 2.0/Where 2.0 concepts, our definition of a participatory Cyberenvironment, and the roles of contexts for building such Cyberinfrastructure. We will then present our arguments of the importance of supporting the full range of social, geospatial, causal and conceptual contexts of contents. We will show the foundation work that we have built so far with implementations in our CyberCollaboratory and Tupelo [58] (a semantic content repository) and then provide the vision for the path towards full contextualized Cyberenvironments.

Yong Liu, Robert E. McGrath, James D. Myers and Joe Futrelle, Towards A Rich-Context Participatory Cyberenvironment, International Workshop on Grid Computing Environments 2007, Nov., 11-12, 2007, Reno, NV, USA. An electronic version of this document will be available at http://casci.rit.edu/proceedings/gce2007

2. WEB 2.0/WHERE 2.0, CYBERENVIRONMENT AND CONTEXTS

2.1 Web 2.0 and Where 2.0

The "Web 2.0" [48] has become a popular buzz word, and a few recent Cyberinfrastructure articles have reviewed a number of Web 2.0 concepts and its relationship with scientific communities [22, 49]. The Web 2.0 is seen to promote an "architecture of participation" ([46]) enabling and encouraging different levels of participation, for both people and software components (e.g., MySpace.com, Facebook.com and YouTube.com, etc., provide novel ways for ordinary users to organize data and communities.). To describe this phenomenon, we consider the full cycle of data and information: publish, transport, aggregate (a "mashup" in the popular terminology [48]) and consume. For publishing, different users not only can publish different contents in different ways such as wikis, blogs, etc., but also can publish open APIs (e.g. Flickr's APIs [62]). In either case, other users/applications may consume such contents or APIs through simple subscriptions such as RSS feeds or remix through mashups to produce new content or new APIs. Open standards and de-facto standards promote such software and encourage participation by developers and users. For example, metadata formats such as FOAF ("Friend of a friend") [7], GeoRSS [15], KML [18] and other "microformats" ([34]) are de-facto standards to carry social and geographical information from provider to consumer.

The "Where 2.0" [61] is a term coined to describe the increasing importance and widespread use of geospatial context and location-based services. The combination of lightweight, open APIs and services (such as Google Maps API [5], GeoNames service [14], GeoTruc service [16], and GeoIQ API [13] etc.) with web-friendly protocols and simple, web-friendly formats (such as KML or environmental observation-oriented ObsKML [45]) have made it relatively easy to write programs to consume, mix and share data from multiple sources which have significant geospatial context. For example, one user can publish geo-tagged sensor reports and his own digital photographs of the sensing location, and a second user may use Google Map APIs to display both in Google Maps and publish and share the result as a KML file.

2.2 Definition of a Participatory Cyberenvironment

We have coined the term "Cyberenvironment" (CE) to describe the next generation Cyberinfrastructure to support 21st century scientific research and discovery [39]. Like other problem solving suites and portals, Cyberenvironments provide an interface to local and shared instruments and sensor networks, data stores, computational resources and capabilities, and analysis and visualization services within a secure framework, combined with capabilities to enable the management of complex projects, development and automation of processes, and group and community-scale collaboration and coordination with distributed colleagues. However, rather than focusing solely on *access to* advanced resources, Cyberenvironments emphasize the integration of resources into end-to-end scientific processes, integration across Cyberenvironments, and the continuing development and dissemination of new resources and new knowledge. The challenge in creating Cyberenvironments is to separate scientific concerns from the basic Cyberinfrastructure coordination mechanisms and to open, participatory use and reuse of shared resources, leading to a robust infrastructure for scientific practice that can harness the creativity of individuals to quickly evolve as needed for next-generation research.

NCSA is evolving these concepts to develop a *participatory Cyberenvironment*, inspired by and built on the Web 2.0 and Where 2.0 patterns [31]. As part of NCSA's efforts to build national Cyberinfrastructure to support collaborative research in environmental engineering and hydrological sciences, the NCSA's CyberCollaboratory, which is built on top of Liferay portal framework ([25, 29]), has been evolving towards a participatory Cyberenvironment [30, 36].

We think the characteristics of a participatory Cyberenvironment can be described as follows:

1) *An architecture of participation for scientific activity*: This refers to both human and software participation. Scientists and engineers work in groups. Such social contexts have important implications since their scientific activities usually involve sharing various contents in their groups. Users should be able to freely create new groups (virtual organizations), invite collaborators to join their groups, and form social networks dynamically. Promoting human participation and fostering social networking among collaborators thus become important. In addition to sharing community-built tools, commercial off-the-shelf the tools such as Matlab, Excel and open source codes such as Liferay should be leveraged and allowed to enable scientists and engineers efficiently use those tools towards discovery.
2) *An open service platform*: reusable and standard-compliant service components must be built. The Liferay portal framework was picked mainly because of its JSR-168 standards compliant, but also because of its capability to expose portal-wide services to external usage through SOAP, JSON or REST style APIs ([49]), which allows other applications to use them for mashups. In addition, the platform itself must be extensible for building additional capabilities [29].
3) *An integration and presentation platform for knowledge network*: scientific process is increasingly involved with many interconnected objects: sensor, data, model, workflow, people, publication, computing resources etc. We have been using NCSA's semantic content repository middleware Tupelo [58] to capture events happening in the CyberCollaboratory and store them in RDF triples. Such information can be used to provide knowledge network and provenance tracking. For example, NCSA's CyberIntegrator [32], an exploratory workflow tool, can store or publish workflow templates and annotations of the workflow as RDF triples in one or more stores. Queries to the RDF enables discovery of relations among data, processes, and people. For example, the CyberCollaboratory can use this and other metadata from the RDF stores to make recommendations to users, such as what tools are typically used to answer certain kinds of questions, or with certain types of dataset.

2.3 The Role of Contexts

In order to build such participatory Cyberenvironment, contexts play a very important role. In this section we discuss four contexts that are important for collaborative Cyberenvironments:

1) The social context, human relations, interactions, and status (Who)
2) The geospatial context, location or spatially referenced information (Where)
3) Provenance, history and causal relations (Why)
4) Semantic or conceptual context, domain-specific relations, ontologies, etc. (What)

Several existing projects are producing Cyberinfrastructure for specific communities, such as CI-Shell [24], Nanohub [42] and Comb-e-Chem [9]. These projects demonstrate the value of context for communities. We believe that supporting a range of contexts will make the participatory Cyberenvironment more useful and sustainable.

There are currently a number of initiatives in the Cyberinfrastructure and e-science domain that promote user participation and *social context*. For example, MyExperiment.org [41] puts workflows as the central objects in scientific activity, and aims to provide a social networking environment for users to upload, tag, find, share, annotate, and reuse workflows. SciLink.com [451] presents a "family tree of science", which allows users to find and connect with their peers through intuitive "genealogy"-type of structure. Nature Network [43] is another social networking site to promote user participation through blogging and forums on scientific and technical topics. We believe that these and similar efforts will show that such social networking capability provided by Cyberinfrastructure has significant impact on collaboration and discovery if used properly.

Geospatial context is important because it is commonly believed that 80% of all data and information either directly or indirectly are related to physical locations [23]. Common location components thus become one important integration vehicle to link diverse information across different domains. This is particularly true for the environmental or earth observatory networks where sensors, people, data etc. usually are associated with particular geospatial contexts. For example, NASA's web-based ScienceOrganizer portal [5] demonstrated that geospatial contexts can be used to integrate remote-sensing images and scientific survey data and generate "context maps" illustrating the geospatial paths of survey actors and the sequence and types of data collected during simulated surface "extra-vehicular activities" at the Mars Desert Research station. The remotely located scientific team found such context maps were extremely valuable for scientific decision making for activity planning and execution. With the advent of the geospatial web or "geoweb" [19] and community efforts such as Open Geospatial Consortium [56] to promote geo-information and solution interoperability, geospatial context is playing an increasingly important role in knowledge network presentation and integration, which is critical for a participatory Cyberenvironment.

A third special context is *provenance*, which describes the causal relationships among artifacts (e.g., data, people, instruments, publications, etc.) and events (e.g., processing steps, accession, custody) in a complex work process. One particular usage of the provenance is to validate e-science experiments [63]. The availability of provenance will empower other scientists to correctly interpret and validate their peers' work, as well as facilitating user participation. An example of the use of provenance in Cyberinfrastructure is the Collaboratory for Multiscale Chemical Science [40, 52], which is built upon a content management abstraction [53] and supports automated metadata extraction, content translation, and provenance browsing.

Lastly, semantic relationships generated by users or inferred from work processes can be also used to build domain-specific *ontologies* and metadata in a participatory Cyberenvironment, enriching the shared knowledge base and enhancing search, browsing, and analysis capabilities. An example of domain-specific ontology for the environmental observatory is CUAHSI Observation Data Model (ODM) [11]. In addition to relatively formal community standards, tagging and other folksonomy-style ontologies have already showed the power of user-generated metadata in Web 2.0 [17, 21]. These relatively simple mechanisms can be improved using ontologies, which can be combined and evolved through a collaborative, participative Web 2.0 approach [6].

3. PRELIMINARY IMPLEMENTATIONS AND DESIGN OF BUILDING CONTEXTS

To illustrate how the above contexts can play together, we describe an end-to-end scenario where all the above contexts can be leveraged to support observatory-centric scientific research. For example, a use case from one of the WATERS (Water and Environmental Research Systems Network: [60]) observatory testbed projects in Corpus Christi Bay of Texas requires support for the full lifecycle of scientific research [35]. These researchers are working to apply sophisticated models to streaming sensor data to identify sensor anomalies and to forecast conditions such as low dissolved oxygen (also known as "hypoxia"). This requires the ability for researchers to apply models built as workflows to the data streams and to publish their derived results as new streams available to the community. Researchers can create a group in the CyberCollaboratory (*social context*), deploy a Google Map-based sensor map *(geospatial context, mashup)* with this group page and allow anyone in this group to subscribe to either raw real-time data streams or derived data streams such as detected anomalies. If a user gets notified either by emails or a desktop-based alerting tool, she can login into the CyberCollaboratory by clicking the link in the email or the desktop tool and go to the sensor monitor page, where she can see real-time sensor data and anomalies based on her subscriptions plotted in an AJAX (Asynchronous Javascripts And XML)-based sensor data monitoring window along with a Google map-based sensor map and a clickable graphical knowledge network generated by *provenance and ontologies context* data showing related persons, publications, workflows and sensors etc. If she finds the existing anomaly algorithm is too sensitive, she can locate another anomaly algorithm from another observatory in the knowledge network and change the workflow parameters on-the-fly and publish the new workflow to a server which can then produce new data streams for community use.

A preliminary demo of such end-to-end system has been shown in an earlier work in SC06 [36]. Although some parts of this previous demo are still in early stage development such as the real-time streaming data management, dynamic knowledge network generation, ubiquitous provenance service, advanced social and geospatial contexts management and visualizations, it has been shown the power of such rich-context participatory Cyberenvironment, which has far more flexibility in terms of providing collaboration, coordination, community-scale customizations and user participation. We believe that the participatory Cyberenvironment is essentially a Web 2.0 approach for science and engineering.

The CyberCollaboratory has been undergoing redesign and new implementation since the beginning of 2007. In this section, we describe our foundation work of continuing moving towards a participatory Cyberenvironment.

3.1 Promoting User Participation

A key goal for the CyberCollaboratory is to facilitate user participation. The Cyberinfrastructure must lower barriers to participation and collaboration. For example, it should be possible to use resources with little effort and to join with minimal inconvenience. .Although collaboratories for various communities have been built in the past, such as [2, 8, 46], dynamically building and using social context is an important advance. The CyberCollaboratory allows any registered user to create a new group and invite both registered and non-registered people to join the group by emails. This simple and easy to use functionality lower the adoption barrier of using the CyberCollaboratory and lays the foundation for using many Web 2.0 technologies such as FOAF and other web-friendly metadata formats and protocols, which we describe in section 3.3. The Collaboratory should provide useful services for registered users, such as searches enhanced with social network and provenance relations.

3.2 Promoting Software Participation

In the evolution of the CyberCollaboratory we draw inspiration from the Web 2.0 mashup, which enables "mass personalization" of web content. We would like to enable a broad capability to share and reuse software and data, analogous to Web mashups, which we call this "Software Participation." For the sake of discussion in this paper, we classify the mashups into two categories: API-based and content-based.

API-based mashup is based on published/open accessible Application Programming Interface (e.g. Google Map APIs). By leveraging Liferay's open service APIs, the CyberCollaboratory has already enabled other non-portal software such as the CyberIntegrator [32], a desktop application, to use the portal API to gather group information and social context for individual investigator, as well as to publish and share workflow templates into the document library and JCR (Java Content Repository: [26]) store in the portal backend.

Content-based mashup entails the use of lightweight, extensible means of producing and sharing metadata, so that independently-produced metadata can be merged using a small set of generic facilities. We use the Resource Description Framework (RDF [3]) to represent descriptions generated by Cyberenvironment tools, and use standard RDF tooling (e.g., Sesame [54], Jena [27]) to build a shared knowledge base combining descriptions from multiple sources. RDF's global naming scheme (i.e., Universal Resource Identifiers (URIs) [4]) and the abstract querying and transformation operations provided by Tupelo ([58]) enable us to infer relationships between independently-produced descriptions and publish inferred information back to the shared RDF store to enable distributed applications to browse and search the enriched knowledge base.

For example, the CyberCollaboratory has been instrumented so that user-generated events such as joining a group or posting to a message board produce RDF event descriptions that we can use to track social relationships, associate authors with content, and link similar resources together even if they were produced using different CyberCollaboratory tools. This kind of information is especially valuable for analyzing the provenance of an artifact such as a document or scientific dataset, because users' interactions with it during various times in its lifecycle can be related to one another through RDF descriptions.

This strategy is not limited to events in the CyberCollaboratory but extends to the desktop as well. For example, the CyberIntegrator application records users' data analysis and processing activities as workflow descriptions, associating that provenance information with data in the Tupelo content repository, allowing applications to for instance trace the provenance of a dataset uploaded to the CyberCollaboratory back to the CyberIntegrator workflow that produced it, including all the steps in that workflow.

RDF can be used to represent and annotate a variety of existing metadata acquired from tools and sources outside the CyberCollaboratory, without requiring significant structural transformation. For example, many applications generate FOAF and Dublin Core ([12]) records which can be represented directly in RDF. Other formats including log files, newsgroup posts, and RSS feeds can be translated into triples. The "open world" semantics of RDF means that representing, storing, and retrieving these disparate sources as triples does not require specific schemas or agreement on data models. Users can employ alternative views of the conceptually global set of triples.

The semantic content repository thus serves as a kind of "semantic network" linking descriptions of distributed activities and information together, providing a rich context in which users can more easily locate information and integrate their work processes across heterogeneous tools.

3.3 Ongoing and Future Efforts

In this section, we will describe our ongoing work which applies the patterns and spirit of Web 2.0 to the CyberCollaboratory. One example is to open up the context used by collaborating groups, which is done through group pages and personal profiles within the CyberCollaboratory. It is straightforward to "expose" the individual group page and personal profile through microformats ([34]). For example, the current version of Liferay has implemented iCal standard for single calendar event.

The Liferay calendar portlet illustrates an an example of a microformats describing events:

```
BEGIN:VCALENDAR
PRODID:-//Liferay Inc//Liferay
Portal 4.3.0//EN
VERSION:2.0
CALSCALE:GREGORIAN
BEGIN:VEVENT
DTSTAMP:20070921T153018Z
UID:8f1bcec0-6857-11dc-bbd6-
223344556677
DTSTART:20070921T153000Z
```

```
DURATION:PT1H1M
SUMMARY:New Event
DESCRIPTION:weekly project meeting
COMMENT:meeting
END:VEVENT
END:VCALENDAR
```

This data can be made available to any authorized user through Web 2.0 style feeds and APIs. This mechanism not only enables conventional calendar features, it can be used for group-defined notifications, such as the availability of new documents, data, or messages.

A second example is to extend the CyberCollaboratory to support FOAF profile for each registered user augmented with physical location of individual user. The FOAF defines an RDF syntax for describing social relations ("A knows B"), which can be augmented with RDF triples identifying the location of the people. This implementation would allow users to find nearby users and create groups for those users. Our initial design is to leverage the user's zip code to find the longitude and latitude of the user location. We can use externally-available web service to do distance calculations between two zip codes [64].

A third example would be generating a geo-referenced provenance causal-relationship context map, which records data use, data creation and processing steps. The provenance provides a validation tool for users to track sensor data Quality Assurance, workflows, investigators and publications. The emerging Open Provenance Model defines a simple, standard vocabulary for exchanging provenance [37, 38]. By combing both geospatial and causal-relationships, users would be able to understand why these steps have occurred as well as where they come from. This will help foster community understanding and trust in data produced by many users in distributed environmental observatory networks.

4. CONCLUSIONS

This paper has described ongoing development of the next generation CyberCollaboratory and our strategy for moving towards a rich-context participatory Cyberenvironment, which enable reuse of software, data, and knowledge.

While the Web 2.0 has fostered mass personalization and personal socializing, we seek to foster and encourage knowledge intensive collaborations through dynamic, context-rich Cyberenvironments for scientific communities. Building participatory Cyberenvironment will facilitate user participation and innovation by opening up interfaces and data to allow customization and reuse. In the business world, building context is the key towards intelligent enterprise knowledge system [28]. In this paper, we have argued that the contextualized information for scientific innovation and discovery is equally critical for successfully managing complex investigations, for forming and sustaining dynamic teams, and for capturing, retaining, and disseminating knowledge.

This work builds on ideas and practices from science collaboratory projects, which have shown the promise of integrating tagging and social networking on top of Cyberinfrastructure. Key features include annotation and tagging ([42]), sharing user data and workflows ([40, 41]), provenance ([63]), and social networking ([43,51]).

To achieve these goals, we apply key patterns from the emerging Web 2.0: including APIs and microformats to foster "software participation", and social context to foster human participation [22, 49]. The next generation NCSA CyberCollaboratory provides generic social, geospatial, provenance, and conceptual contexts, as well as open service APIs, which can be customized and extended to create community-specific collaborative environments. As in the Web 2.0, "the intelligence is at the edge": communities, teams, and virtual organizations will use these mechanisms to produce and consume information for their own problem solving environments.

We envision that researchers involved with distributed environmental observatory networks will benefit from such participatory Cyberenvironment. While traditional portals organize the documents, data, and processes of a single community, the next generation CyberCollaboratory will enable sharing, reuse and promoting system science level study of the observed earth environment. For example, MetaCarta [33] style geo-referenced sensors, data (both real-time and model forecast output), documents, visualizations and publications can be directly integrated on to an observatory map, building a "rich context" knowledge map for a digital observatory, permitting multidisciplinary analysis and synthesis, and ultimately providing a pathway to approach geoscience problems and processes from an Earth system science perspective [50]. Distributed observatory networks will also benefit from the social context that allows individual observatory to set up their own groups, from small teams through global collaborations.

5. ACKNOWLEDGMENTS

We acknowledge the entire NCSA ECID project team for the CI development efforts. This material is based upon work supported by the National Science Foundation (NSF) under Award No. BES-0414259, BES-0533513, and SCI-0525308 and the Office of Naval Research (ONR) under Award No. N00014-04-1-0437. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of NSF and ONR..

6. REFERENCES

[1] Banks, T., 2006. Web Services Resource Framework (WSRF) – Primer v1.2. OASIS, Committee Draft 02, 23 May 2006, http://docs.oasis-open.org/wsrf/wsrf-primer-1.2-primer-cd-02.pdf

[2] Bebout, B.M., Carpenter, S.P., Des Marais, D.J., Discipulo, M., Embaye, T., Garcia-Pichel, F., Hoehler, T.M., Hogan, M., Jahnke, L.L., Keller, R.M., Miller, S.R., Prufert-Bebout, L.E., Raleigh, C., Rothrock, M. and Turk, K., 2002. Long Term Manipulations of Intact Microbial Mat Communities in a Greenhouse Collaboratory: Simulating Earth's Present and Past Field Environments. Astrobiology, 2: 383–402.

[3] Beckett, D., 2004. RDF/XML Syntax Specification (Revised)" W3C, Recommendation 10 February 2004.

[4] Berners-Lee, T., Fielding, R. and Masinter, L., 1998. Uniform Resource Identifiers (URI): Generic Syntax. IETF, RFC 2396, http://www.ietf.org/rfc2398.txt.

[5] Berrios, D.C., Sierhuis, M. and Keller, R. M.., 2007. Geospatial Information Integration for Science Activity Planning at the Mars Desert Research Station. In The Geospatial Web; How Geobrowsers, Social Software and the

Web 2.0 are Shaping the Network Society, Scharl, A. and Tochtermann, K., eds., Springer: London.

[6] Braun, S., Schmidt, A. and Walter, A., 2007. Ontology Maturing: A Collaborative Web 2.0 Approach to Ontology Engineering. 16th International World Wide Web Conference, May 8-12.

[7] Brickley, D. and Miller, L., 2007. FOAF Vocabulary Specification 0.9. http://xmlns.com/foaf/spec/.

[8] Chin, G.J. and Lansing, C.S., 2004. Capturing and Supporting Contexts for Scientific Data Sharing via the Biological Sciences Collaboratory, ACM Conference on Computer Supported Cooperative Work (CSCW-2004). Chicago: ACM Press*,* 409–418.

[9] Comb-e-Chem: http://www.combechem.org

[10] Contractor, N., 2005. The role of social network analysis in enabling cyberinfrastructure and the role of cyberinfrastructure in enabling social network analysis. White paper prepared for the National Science Foundation workshop on cyberinfrastructure for the Social Sciences.

[11] CUAHSI ODM: http://www.cuahsi.org/his/odm.html

[12] Dublin Core Metadata Initiative, 2006. Dublin Core Metadata Element Set, Version 1.1. http://dublincore.org/documents/dces/

[13] GeoIQ service:http://www.geoiq.com/

[14] GeoNames service: http://www.geonames.org

[15] GeoRSS: Geographically Encoded Objects for RSS feeds. 2007, http://georss.org.

[16] GeoTruc service: http://www.geotruc.net/

[17] Golder, W. A. and Huberman, B. A., 2005. The Structure of Collaborative Tagging Systems. Information Dynamics Lab, HP Labs.

[18] Google, 2007. KML Overview. http://code.google.com/apis/kml/documentation/.

[19] Google, 2007. Google LatLong: A new world unfolding. http://google-latlong.blogspot.com/2007/05/new-world-unfolding.html

[20] Google, 2007. Google Maps API: http://www.google.com/apis/maps/index.

[21] Hammon, T., Hannay, T., Lund, T. and Scott, J., 2005. Social Bookmarking Tools (I). D-Lib Magazine, vol. 11.

[22] Hannay, T., 2007. Web 2.0 in Science. CTWatch Quarterly*,* Volume 3, Number 3, August 2007. http://www.ctwatch.org/quarterly/articles/2007/08/web-20-in-science/

[23] Hart, G. and Dolbear, C., 2007. What's so Special about Spatial?. In. The Geospatial Web; How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society, Scharl, A. and Tochtermann, K., eds., Springer: London.

[24] Herr, B. W., Huang, W., Penumarthy, S. and Börner, K., 2007. Designing Highly Flexible and Usable Cyberinfrastructures for Convergenc*e.* Annals of the New York Academy of Sciences, *1093:161-179*.

[25] Infoworld, 2007. Best Open Source Portal: http://www.infoworld.com/article/07/09/10/37FE-boss-applications_2.html

[26] Java Content Repository: http://jcp.org/aboutJava/communityprocess/final/jsr170/index.html

[27] Jena: http://www.hpl.hp.com/semweb/jena.html

[28] Jonas, J., 2006. Introducing Perpetual Analytics. Invited talk at Industrial and Government Applications Track, Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[29] Liferay portal: http://www.liferay.com

[30] Liu, Y., Minsker, B. and Hill, D., 2007. Cyberinfrastructure Technologies To Support QA/QC and Event-Driven Analysis Of Distributed Sensing Data. International Workshop on Advances in Hydroinformatics, 4 – 7 June 2007,Niagara Falls Canada

[31] Liu, Y, Myers, J., Minsker, B. and Futrelle, J., 2007. Leveraging Web 2.0 technologies in a Cyberenvironment for Observatory-centric Environmental Research. OGF-19 Semantic Web 2.0 and Grid Workshop, Jan. 29, 2007, the 19th Open Grid Forum, Chapel Hill, North Carolina, USA

[32] Marini, L., Minsker, B., Kooper, R., Myers, J., and Bajcsy, P., 2006. CyberIntegrator: A Highly Interactive Problem Solving Environment to Support Environmental Observatories, Eos Trans. AGU 87(52), Fall Meet. Suppl.

[33] MetaCarta Web Services: http://developers.metacarta.com/

[34] Microformats: http://microformats.org/

[35] Minsker, B., Coopersmith, E., Hodges, B., Maidment, D., Bonner, J. and Montagna. P. 2006. An Environmental Information System for Hypoxia in Corpus Christi Bay: A WATERS Network Testbed. Presented at AGU 2006 Fall Meeting.

[36] Minsker, B., Myers, J., Marikos, M., Wentling, T., Downey, S., Liu, Y., Bajcsy, P., Kooper, R., Marini, L., Contractor, N., Green, H. D., and Futrelle, J. 2006. NCSA environmental cyberinfrastructure demonstration project: creating cyberenvironments for environmental engineering and hydrological science communities. In Proceedings of the 2006 ACM/IEEE Conference on Supercomputing (Tampa, Florida, November 11 - 17, 2006). SC '06. ACM Press, New York, NY, 161. DOI= http://doi.acm.org/10.1145

[37] Moreau, L. and Foster, I.., 2006. Provenance and Annotation of Data (IPAW 2006). Vol. 4145: Springer.

[38] Moreau, L., Freire, J., Myers, J., Futrelle, J., and Paulson, P., 2007. The Open Provenance Model.

[39] Myers, J. and Dunning, T. 2006. Cyberenvironments and Cyberinfrastructure: Powering Cyber-research in the 21st Century. Proceedings of the Foundations of Molecular Modeling and Simulation, Blaine, WA, July 9-14, 2006.

[40] Myers, J. D., Allison, T. C., Bittner, S., Didier, B., Michael Frenklach, M., Green, W. H., Ho, Y., Hewson, J., Koegler, W. S., Lansing, C., Leahy, D., Lee, M., McCoy, R., Minkoff, M., Nijsure, S., von Laszewski, G., Montoya, D., Oluwole, L., Pancerella, C., Pinzon, R., Pitz, W., Rahn, L. A., Ruscic,

B., Schuchardt, K., Stephan, E., Wagner, A., Windus, T., and Yang, C., 2005. A Collaborative Informatics Infrastructure for Multi-scale Science. Cluster Computing, *8 (4):243-253*.

[41] MyExperiment.org: http://myexperiment.org

[42] Nanohub: http://www.nanohub.org

[43] Nature Network: http://network.nature.com/

[44] NCSA CyberCollaboratory. http://ecid.ncsa.uiuc.edu/cybercollab/

[45] ObsKML: http://nautilus.baruch.sc.edu/twiki_dmcc/bin/view/Main/ObsKML

[46] Olson, G.M., Atkins, D.E., Clauer, R., Finholt, T.A., Jahanian, F., Killeen, T.L., Prakash, A. and Weymouth, T., 1998. "The Upper Atmospheric Research Collaboratory." *Interactions*, 5: 48–55.

[47] O'Reilly, T., 2004. What Is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software. San Francisco, CA: O'Reilly. http://www.oreillynet.com/pub/a/oreilly/tim/articles/architecture_of_participation.html

[48] O'Reilly, T., 2005. What Is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software. San Francisco, CA: O'Reilly. http://www.oreilly-net.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[49] Pierce, M. E., Fox, G., Yuan, H. and Deng, Y., 2007. Cyberinfrastructure and Web 2.0, International Advanced Research Workshop on High Performance Computing and Grids, Cetraro (Italy).

[50] Ramamurthy, M. K., 2006. A new generation of cyberinfrasturcture and data services for earth system science education and research. Advances in Geosciences, 8, 69-78, 2006

[51] SciLink.com: http://sciLink.com

[52] Schuchardt, K., Pancerella, C., Rahn, L. A., Didier, B., Kodeboyina, D., Leahy, D., Myers, J. D., Olumwole Oluwole, O., Pitz, W., Ruscic, B., Song, J., von Laszewski, G., and Yang, C., 2005. Portal-based Knowledge Environment for Collaborative Science. In GCE 2005: Workshop on Grid Computing Portals.

[53] Schwidder, J., Talbott, T. D., and Myers, J. D., 2005. Bootstrapping to a Semantic Grid. In Semantic Infrastructure for Grid Computing Applications Workshop, at the IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2005)

[54] Sesame: http://www.openrdf.org

[55] TeraGrid User Portal: https://portal.teragrid.org/.

[56] The Open Geospatial Consortium, Inc (OGC): http://www.opengeospatial.org/ogc

[57] Tuecke, S., Czajkowski, K., Foster, I., Frey, J., Graham, S., Kesselman, C., Maquire, T., Sandholm, T., Snelling, D. and Vanderbilt, P., 2003. Open Grid Service Infrastructure (OGSI). Global Grid Forum, Proposed Recommendation GFD-R-P.15, June 27 2003.

[58] Tupelo project: http://tupeloproject.org

[59] Wasserman, S. and Faust, K., 1994. Social Network Analysis. Cambridge: Cambridge University Press.

[60] Water and Environmental Research Systems Network: http://www.watersnet.org

[61] Where 2.0 conference: http://conferences.oreillynet.com/where2007/

[62] Wilkinson, D., 2007. Flikr Mashup*s*. Indianapolis: Wiley.

[63] Wong, S. C., Miles, S., Fang, W., Groth, P., and Moreau. Luc. 2005. Validation of E-Science Experiments using a Provenance-based Approach. UK eScience All Hands Meeting Proceedings.

[64] Zip Distance Calculator Web Service: http://webservices.imacination.com/distance/