A Web Collaboration Architecture

Antonio Tapiador, Antonio Fumero, Joaquín Salvachúa, Sandra Aguirre
Department of Telematic Systems Engineering
Universidad Politécnica de Madrid (UPM)
28040 Madrid, Spain
Email: {atapiador, fumero, jsalvachua, saguirre}@dit.upm.es

***Abstract*— Computer Supported Collaborative Work (CSCW) has typically been bound within organizations. Groupware applications are designed with organizational structures in mind, in a top-down approach that predicts communication and collaboration interactions between people. With Internet adoption in society, CSCW overflows organizational borders towards Internet Supported Collaborative Work (ISCW), where the organization of work becomes decentralized, centered on individual interests and pretty dynamic. This paper proposes a Web Collaboration Architecture for the new web environment. The article reviews common groupware problems and reflects recent organizational changes. It presents Web 2.0 features and technical state-of-the-art. Then, it describes the architecture proposed, to finish with conclusions and future research lines.**

I. Introduction

Since its beginnings in the 1980s, Computer Supported Collaborative Work (CSCW) has been focused on formal organizations. CSCW emerged as a multidisciplinary field, involving technologists, economists, social psychologists. anthropologists, organizational theorists, educators, and anyone else who could shed light on group activity. [13]. It was the time of corporate and academic networks, and the collaboration supported by computers only made sense in the framework of big organizations.

Subsequently, network connections and PCs become more popular. CSCW field grew covering more organizational aspects, ranging from the individual to organization [13].

Many groupware solutions have been developed for supporting these organization-centered scenarios. Nevertheless, they are far away from being completely successful. Some authors show groupware and intranets drawbacks [19], which include: difficulties finding relevant information, lack of coherent design and structure, inconsistent vocabulary, unclear ownership, becoming one-way communications channels for corporate information instead of sharing knowledge tools on a peer-to-peer level.

Groupware tools have been designed for "tree organizations". But organization environment has become dynamic, organizations themselves need to adapt to changes. Organic structures emerge, and the networks science approaches are gaining momentum within the management science, changing the way we understand the organization itself. Hierarchies don't apply no more; we must consider a networked approach for the modern organization [3].

As organizations flatten, become more organic, collaboration patterns turn semilattices. In these new environments, centered communication becomes a key factor [7]. Users must be able to customize presence and operation to suit individual needs, represent themselves as unique individuals and select and control the medium and manner in which they access and participate in the environment. The 'new organization' must be tackled as a series of virtual communities interacting for achieving a set of well-defined shared objectives and goals. We can see such a scenario change in the real world; In considering the species populating the digital economy landscape, we are shifting from "corporation man" to "networked (knowledge) professional" [5]. Such a change directly impacts the traditional top-down view which usually has shaped the CSCW systems design and deployment.

II. The new web environment

In the last few years, we have heard about the next World Wide Web generation, Web 2.0. The term "Web 2.0" is a controversial one. There isn't full consensus about what the "Web 2.0" is. We will adopt here the Wikipedia [21] definition, as the Wikipedia is one of its paradigmatic products:

> Web 2.0 generally refers to a second generation of services available on the World Wide Web that let people collaborate, and share information on line. In contrast to the first generation, Web 2.0 gives users an experience closer to desktop applications than the traditional static Web pages.

The qualitative leap from Web 1.0 is the ability of users to *easily change remote content*. The implications of this new ability are crucial, since communication channels become bidirectional. Everybody has the potential to participate on *"anything"* on the web.

A. Next generation web features

Some papers try to collect this new framework characteristics [20] [17]. From our point of view, there are four main features that characterize Web 2.0. The first two of them coming from the original Web:

- **Decentralization**. As a consequence of being a real network. Every node has the ability to act as emitter and receptor of information.
- **Openness**. Using standards in communication, free licenses on content, promotes collaboration.
- **Dynamic**. Applications are developed and deployed quickly. User suggestions are attended and supported.

1-4244-0429-0/06/$20.00 ©2006 IEEE

- **User orientation**. Easier and better user interfaces facilitates participation.

These four features just match with social software application problems, raised by Eric Gradman in [12]. As they are issues don't fully solved yet, they keep applications for being "Web 2.0 compliant".

B. Actions

Web 2.0 features (decentralization, openness, dynamic, user orientation) characterize the new framework where users manage online content.

This content can be original on the web, like new pictures or songs. It can be information related to users or organizations, like mail addresses or lists of contacts. Blogs are another example of content related to users, they provide context about their authors. Content can be also related to other content, like blog comments, user ratings or tags. Content can be composed by other content, like web pages including pictures.

We have identified three main actions in the new web:

1) Publish: Users publish information in several formats: news, events, pictures, audio, video, documents, etc.. Each data format is suffering a parallel evolution in its management. Different content publication sites appear: blogs for news, *Flickr*[1] for images, *Youtube*[2] for video publishing, and *del.icio.us*[3] in the case of bookmarks sharing are some examples.

2) Subscribe: The second action is content subscription. Microformats popularity with REST architecture allow users automatic harvesting of changes on their favorite web sites. Last blog news, new songs from a music group or new videos tagged with certain word, are automatically collected without the need of revisiting each specific site. Subscription content is multimedia, and web sites offer subscription services based on multiple patterns (posts, comments, tags, etc..)

3) Search: Content search has usually been performed through web search engines, web sites owning search robots that exhaustively follow web links, indexing the content. As users publish multimedia material in different content publication sites, searches must be performed according to content type on publishing specific sites. With tags appearance, users are able to add meta-information to published contents. Web sites providing content tagging capabilities also allow searching on these key words.

4) Site Collaboration: This group collects other type of actions performed in the web. These are site specific, depend on the web site and represent the tendencies toward collaboration. The paradigm of this group are wikis.

C. The Identity Problem

Since the Web 2.0 is giving people the opportunity of taking several actions based on the individual (publication of news, photos, comments, personal information, etc...), the trustability, and credibility emerge as key issues, not only from a sociological, but from a technological point-of-view. These issues translate into the identity key issue. Up to now, there isn't a solid identity model. Users have to register themselves every time they access a web site. This is awkward and not much handily. An identity mechanism should be build, allowing users log in using one single URI.

Some initiatives try to tackle the identity problem within the Web 2.0 framework, including Identity 2.0 [14], mIDm [4] and OpenID [9]. While OpenID is spreading, a consistent representation of identity in the Web is missing today. When the identity problem is solved, organizations and individuals will define and consolidate their web presence. Some proposals like PIF [12] point in this way; beyond the traditional web page, organizations and individuals web sites will gather news and media publications, subscription to several services, just as their own content search services, which will be also based on tags.

On the other hand, web identity on the client side will allow easy access to personal and organizational content and services.

D. Social Networks

All these interactions entail relations between people. Communication boost, channels like *Instant Messaging* become very popular. Social networks arise, as well as new web platforms dedicated to social interactions, like *Tribe*[4] - where people share their interest and search for similar people, or *43 Things*[5], where people are able to share their personal life's projects for receiving advices and feedback from different users. Social Networking tools are populating the Internet penetrating even the enterprise and corporate world. Social Capital is being used for assessing the value of networks, and Social Networks Analysis (SNA) techniques are being added to the corporate human resources toolkits because of the importance of "worthy networks" for the business models sustainability in a digital economy [16].

III. Web Puzzle Pieces

WWW architecture is constantly involving. Opposite to other fields of software architecture, the Web is a scenario with many actors participating on it. It changes every day. "De-facto" standards become main-stream due to their wide adoption among users. They are not the product of planned design and discussed consensus, but rather a kind of evolutionary product where many factors have been relevant. This is the fate of Web 2.0 architecture, which is probably becoming outdated as we are writing this paper.

A. The Browser as Rich Client

From the user point of view, the browser is the main gateway to the Web as we know it. Other desktop applications may use web protocols in order to gather information, but the main access to web resources is made through the web browser as a rich client. This fact turns the browser the best target for the ultimate Web 2.0 improvements and innovations.

[1] http://flickr.com/
[2] http://www.youtube.com/
[3] http://del.icio.us/
[4] http://www.tribe.net/
[5] http://www.43things.com/

The protocol used for accessing web platforms is HTTP, so enhancements are made over it. One of them is **REST** architecture, the light version of Web Services.

The most representative browser technology from Web 2.0 is Asynchronous JavaScript and XML (**AJAX**). It is "a Web development technique for creating interactive web applications. With this technology, web pages evolve and start to look like desktop applications. The line dividing both worlds, local and remote, vanishes.

Bookmarklets improve browser functionality. Small JavaScript snippets, they are stored within a bookmark. They can modify of the way the web page is displayed, extract and mangle data from a web page, or send this data to search engines or other web pages. Bookmarklets point out the way towards users new roles. Instead of mere information consumers, users manage, subscribe to, search, mangle, and publish a wide amount of information. Users are able to transfer information between websites in a quicker and more efficient way. The next step in this way is *Flock*[6]. It is an enhancement on free browser *Mozilla Firefox*[7] integrating social software services, e.g. bookmarks sharing or blog publishing. Flock represents the paradigm of user tool for the new web collaboration scenario.

B. Web Services

Web sites are less and less information silos. They make their information available for interaction. These practices facilitate collaboration. The technologies used for this are Web services, "a software system designed to support interoperable machine-to-machine interaction over a network" [21]. There are two types of Web Services:

- **SOA**. Simple Object Access Protocol, standard protocol developed by Microsoft, IBM and others, at present under the W3C support. It provides a basic messaging framework based on XML data interchange [21]. SOAP framework is robust, complex and pretty static.
- **REST** Representational State Transfer, is an architectural model oriented to distributed hypermedia systems [21]. Its web services implementation uses XML and HTTP standards for data interchange. It doesn't have SOAP extra abstractions what facilitates services implementation and deployment. This is contributing to a fast spread and adoption by many web sites. REST depend on *microformats* for service related information management.
  Microformats provide information structure for REST transfers. RSS and Atom are the most representative microformats, XML standards primary used for web subscription and publishing. Other microformats types are FOAF (for machine-readable modelling of homepage-like content and social networks), iCalendar (calendar data interchange) and vCard (personal data interchange).

With these technologies, web sites become service providers. Every organization or individual is able to built or host its own service and offer it to the world. One step further on server side collaboration is building web server tools using web services from external sites. This is called a *Mashup*, and it opens another collaboration way. It brings the opportunity to create a complex mesh of web services offering, use and dependency.

C. Tags

Another technical feature emerging from Web 2.0 are tags. Content management web sites allow their users attaching key works to the media they store. These words are called tags, and provide semantics to content. They compose "folksonomies", a categorization system built by principles. Opposite to ontologies, this type of categorization is user oriented, which follows Web 2.0 features. Tag use facilitates content search. Search web sites like *Technorati*[8] base their search on blog posts in this tags. It also allows subscription to certain tags.

D. Collaborative Web Sites

Web sites as a whole also reach new stages in this collaborative framework. One example are collaborative content sites, which are focused on collecting information about certain topic. The most representative of them is the *Wikipedia*[9], a wiki web site focused on collaborative writting a full free encyclopedia in many languages. The web site is almost fully editable by users. Beyond many predictions, collaboration works in a surprising way. Wikipedia even comes close to Britannica in terms of the accuracy of its science entries [10].

Other examples of collaborative web sites are news ones. *Slashdot*[10], *Kuro5hin*[11] or *Digg*[12] content is contributed by users, and it is published to the front page in several different ways, that range from editors control to users moderation.

IV. Web Collaboration Architecture

All these pieces build a new collaboration framework. As opposed to old CSCW solutions, this framework is *integrated in users habits*. They don't have to download just another program, or learn another interface. Internet Supported Collaborative Work (ISCW) become part of people lifes. Users become part of the global web collaborative work. The paradigm of this new collaboration framework is Free / Open Source Software [18], because software builds the framework itself. But other knowledge fields are to be affected.

The new collaboration framework aims for solving at least some of old CSCW drawbacks. It is user oriented, encourages authorship. It provides new ways of searching for content. Folksonomies emerge as user-build vocabulary. It builds bidirectional communication channels.

Users collaborate on the web using a Personal Collaborative Framework. It resides in the user client (PC, mobile, etc..). It has browser capabilities, it not only collects and renders web

[6] http://flock.com/
[7] http://www.mozilla.com/firefox/
[8] http://www.technorati.com/
[9] http://wikipedia.org/
[10] http://slashdot.org/
[11] http://www.kuro5hin.org/
[12] http://digg.com/

pages, but it also supports described communication actions (publish, subscribe, search and eventually, site collaboration). It has an integrated editor for posting news and comments. an integrated photo tool for managing personal and other people images. Bookmarks are shared in a transparent way to multiple sites. An integrated web feed reader manages subscription to web sites feeds. Information is easily moved from one tool to another. Identity is also managed into the framework. Users log *into the PCF* rather than in web sites. In this way, synchronization is performed between the PCF and the web servers, so the user is able to access her post services, subscribed feeds, photo albums or bookmarks. The PCF is able to manage multiple identities at the same time.

Beyond traditional web communication (HTTP), the PCF uses web services to support communicative and collaborative actions. It uses a composed REST/SOAP component to communicate with different web servers APIs. Web servers services are accessed through their public APIs. These services include blog sites, photo sharing sites, bookmark sharing sites or personal information. Users may have their services spread among a handful of different services (personal, job, association servers...). One web server may provide multiple services. For example, an University provides their students with blog accounts, searching engines, content sharing services and a feed reader. Students just log in their Personal Collaborative Framework (not in the University site), providing an URI like *user@university.edu* or *user.university.edu* and a password. The PCF authenticates them against the University site and discovers the services offered by the institution, which are instantly available for the student. Students browse the web and are able to use University services at the same time.

Flock browser is a paradigmatic initiative of PCF. It currently supports blog, images and bookmarks publishing, news feed reader and search engines integration. On the other hand it lacks of a consistent identity system, every service has independent account management. Service discovery is only available for subscription feeds. It also lacks of feeds and personal information publication and storage at an external web site, in order to access feed reading from different PCFs.

As Web 2.0 communication actions become more popular, the PCF will move to the desktop integration on the PC and the software of other mobile artifacts.

V. Conclusions and Future Work

With the arrival of the next generation web environment, Computer Supported Collaborative Work has taken one further step, which we call “Internet Supported Collaborative Work”. Web 2.0 provides a incredible framework for collaboration.

We propose a Web Collaborative Architecture centered on individuals. It involves the Personal Collaborative Framework, a set of web tools that integrates the following actions: publish, subscribe, search and site collaboration. It communicates through web services (SOAP and REST) with server sites APIs. Web sites offer several services to users, setting web services APIs. Organizations and individuals set up their own web sites, promoting web presence and collaborative actions.

The identity problem stands out among the pending issues. The PCF must manage identities and services associated with them. A consistent identity model is needed. Other pending issue has to do with service discovery. Awareness of the services offered by web sites visited must be integrated in the PCF.

Acknowledgments

This work is partly funded by European Commission via the IST projects iCamp[13] and ProLEARN[14].

References

[1] Allen, C, *Tracing the Evolution of Social Software*, Life with Alacrity blog, October, 13, 2004, http://www.lifewithalacrity.com/2004/10/tracing_the_evo.html

[2] Boyd S., *Are you ready for Social Software?*, Darwin Magazine, May 2003, http://www.darwinmag.com/read/050103/social.html

[3] Cross, R., *A Practical Guide to Social Networks*, Harvard Business Review, March 2005.

[4] Downes, S., *mIDm - Self-Identification the World Wide Web*, May 4, 2005, http://www.downes.ca/idme.htm

[5] Economist, T., *The new organization*, The Economist Surveys, January, 19, 2006, http://www.economist.com/surveys/displaystory.cfm?story_id=5380483

[6] Eseryel D., Ganesan R., Edmonds G.S., *Review of Computer-Supported Collaborative Work Systems*, Educational Technology & Society 5 (2), 2002, ISSN 1436-4522, http://ifets.ieee.org/periodical/vol_2_2002/eseryel_ganesan.html

[7] Farmer J., *Centred Communication: Weblogs and aggregation in the organisation*, Blogtalk Downunder, May 19-22 2005, Sydney, http://incsub.org/blogtalk/?page_id=54

[8] Farrel, J., *Wikis, blogs and other community tools in the enterprise.*, Solve enterprise application problems with social collaboration technology, IBM developerWorks 03/23/2006, http://www-128.ibm.com/developerworks/web/library/wa-wikiapps.html

[9] Fitzpatrick B., *OpenID Specification*, (work in progress), http://openid.net/

[10] Giles, J., *Internet Encyclopaedias Go Head to Head*, Nature, Dec. 15, 2005, http://www.nature.com/nature/journal/v438/n7070/full/438900a.html

[11] Gotta, M., *Rise of the Social Enterprise*, Collaborative Thinking Blog, March, 17, 2006, http://mikeg.typepad.com/perceptions/2006/03/rise_of_the_soc.html

[12] Gradman E., *Distributed Social Software*, December 2003, http://gradman.com/projects/dss/

[13] Grudin J., *Computer-Supported Collaborative Work: History and Focus*, IEEE Computer, Vol. 27, No. 5, pp 19-26, May 1994

[14] Hardt, D., *Identity 2.0*, 2006, http://www.identity20.com/

[15] Himanen, P., *The Hacker Ethic*, Random House, New York, 2001

[16] Majumdar, S., *A conceptual model for sustaining competitive advantage in digital economy*, Proceedings of WEBIST 2006, 2nd International Conference on Web Information Systems and Technologies, Setúbal, Portugal, April, 11 - 13, 2006.

[17] McAfee A.P., *Enterprise 2.0: The Dawn of Emergent Collaboration*, MIT Sloan Management Review, Reprint 47306; Spring 2006, Vol. 47, No. 3, pp. 21-28.

[18] Shroff G., Mehta A., Puneet P., Sinha R., *Collaborative Development of Business Applications*, Proceedings of WEBIST 2006, 2nd International Conference on Web Information Systems and Technologies, Setúbal, Portugal,.April 11 - 13, 2006

[19] Stenmark, D., *Designing the new intranet*, Gothenburg Studies in Informatics, 2002, http://w3.informatik.gu.se/~dixi/thesis/thesis.pdf

[20] Valdes R., Smith D. M., *Web 2.0: Get Ready for the Next Old Thing*, Gartner Inc., 28 December 2005

[21] Wikipedia, http://en.wikipedia.org/

[13] http://www.icamp.eu/

[14] http://www.prolearn-project.org/