Is the Mashup Technology Mature for its Application in an Institutional Website?

Serena Pastore
INAF – Astronomical Observatory of Padova
Vicolo dell'Osservatorio 5 – 35122
Padova, ITALY
e-mail: serena.pastore@oapd.inaf.it

Abstract—**Web design for general-purpose websites today requires framework software infrastructure built from different components. Among web server software as the front--end needed to process HTTP/HTTPS requests, the top-down layers consist of different software middleware of which CMS (Content Management System) technology is a primary component. However, with the proliferation of websites and the advent of Web 2.0 philosophy, much distributed information has become structured in interoperability file formats (i.e., XML, RSS, JSON, etc.), and thus, new technologies such as mashups have been developed to collect this information in a dynamic way. The document approaches mashup technologies in order to evaluate their application not only in an end-user environment to create a personalized start page that embeds different sources of information relevant to a specific user, but also in generic areas such as the development of an institutional website in terms of creating a business mashup.**

Keywords - Mashup applications; web widgets; website platform; XML formats; CMS technology; Web 2.0 tools.

I. Introduction

Most websites today are built upon a software infrastructure made of different components of which Content Management System (CMS) technology [1] is the primary one. Among the variety of commercial and open source solutions, the common aspect is the presence of a database management system (DBMS) where content is stored and an engine that queries such a system transforms content in an HTML page linked to stile sheets following some template. So, these days, communicating over the web means, in most cases, the setup of a website using a web CMS software and its integration in several ways with one or more of the so-called Web 2.0 tools [2], i.e., web systems able to allow a single user to interact by presenting videos stored on YouTube [3] or creating a link with Facebook [4], Twitter [5] and so on. The applicability in a broader context is called Enterprise 2.0 [6], a term that describes an organizational model for companies and the application of technologies aimed at distributing information and services through a collaborative lens.

Most distributed content is available in an interoperability format such as XML technologies [7] or related structures specific for some functionalities (i.e., content-specific markup languages such as RSS [8], Atom [9], JSON [10], KML [11], etc.) that share a markup language used to describe the content.

From a technical perspective, this has become a simple operation: an organization could implement a website by using web CMS software (i.e., Wordpress [12], Joomla [13], Drupal [14], Plone [15], etc.) which, in a few steps, allows a website to install a database management system as a backend. The graphical aspect is provided through numerous templates, while content editors take advantage of a simplified method for uploading files and sending commands through a web interface. This revolution in website development has undoubtedly enhanced a webmaster's work, but it has contributed to a shift in the attention from other aspects of a website (graphical presentation, the presence of feeds, videos and changing images on the homepage) away from content. Web design has, in some cases, been reduced in its importance as a roadmap made from specific steps whose final objective is to communicate information.

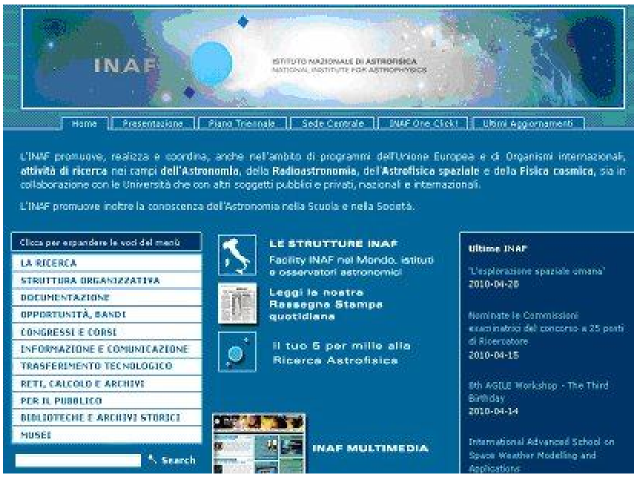
In our opinion, most of these websites result in appealing entry pages, with poor content since, if dynamically updated, it is reduced to a list of news. Most modern websites appear as web journals collecting the same chunks of information. Each organization has its own website, even if a great many have poorly designed content or show a duplication of information, instead relying on the force of an appealing interface and connection with Web 2.0 social tools. Among these, web feeds are a common method of distributing news [16]. Since, in a web infrastructure, information is distributed and no longer centralized, is a general-purpose website still useful for communicating over the web? Is it perhaps more useful to take advantage of web technologies (i.e. XML, REST [17], Ajax [18], etc.) to create a mashup application that uses web widgets in order to create a homepage that could aggregate and present information coming from different sources? After these considerations, a web project manager could approach a different solution in a restyling of a website and try to aggregate and present content that is already available on the web, taking whatever attention necessary to follow all the steps required for good web design.

Such technologies are normally used in an end-user environment (i.e., the iGoogle approach [19], the user's Google startup page that allows a user's customization of merging different feeds in a single page). This paper describes research we conducted for a project to restyle our website in order to evaluate if an institutional website could be reduced to a homepage made up of mashups and web widgets as interactive graphical elements and that adheres to web standards in terms of accessibility laws.

In our opinion, such as solution could contribute to reducing the duplication of information and making web content relevant again. This paper starts by describing our need for re-designing the actual website, and presents an overview of technologies related to mashups and the ability to collect data and services that should appear in an interoperability manner. Then, the paper discusses a prototype solution that could be adapted to the case study, outlining the issues related to collect data from different sources and publishing them in a user-friendly manner for different categories of users.

II. THE ISTITUTIONAL CONTEXT AND THE STATE OF ART

When, in 2005, we began to develop the new website for the INAF [20] institute (see Fig. 1), we focused our attention on the type of content that we should publish and the necessity of defining a publication workflow that starts from a web editor inserting content and then passes to a redactor and then a reviser.



L'INAF promuove, realizza e coordina, anche nell'ambito di programmi dell'Unione Europea e di Organismi internazionali, **attività di ricerca** nei campi **dell'Astronomia**, della **Radioastronomia**, dell'**Astrofisica spaziale** e della **Fisica cosmica**, sia in collaborazione con le Università che con altri soggetti pubblici e privati, nazionali e internazionali.

L'INAF promuove inoltre la conoscenza dell'Astronomia nella Scuola e nella Società.

Clicca per espandere le voci del menù
LA RICERCA
STRUTTURA ORGANIZZATIVA
DOCUMENTAZIONE
OPPORTUNITÀ, BANDI
CONGRESSI E CORSI
INFORMAZIONE E COMUNICAZIONE
TRASFERIMENTO TECNOLOGICO
RETI, CALCOLO E ARCHIVI
PER IL PUBBLICO
BIBLIOTECHE E ARCHIVI STORICI
MUSEI
Search
LE STRUTTURE INAF
Facility INAF nel Mondo, istituti e osservatori astronomici
Leggi la nostra Rassegna Stampa quotidiana
Il tuo 5 per mille alla Ricerca Astrofisica
INAF MULTIMEDIA
Ultime INAF

'L'esplorazione spaziale umana'
2010-04-20

Nominate le Commissioni esaminatrici del concorso a 25 posti di Ricercatore
2010-04-15

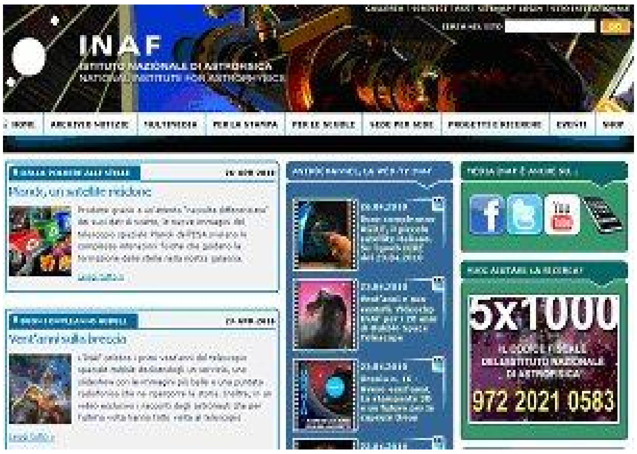
8th AGILE Workshop - The Third Birthday
2010-04-14

International Advanced School on Space Weather Modelling and Applications

Figure 1. How actually appears the homepage of the INAF institute

The choice fell [21] on Plone software as a CMS technology, which allows to define, in detail, the types of roles that different users can have and the actions the users can perform. Even though the project objective was satisfied, this website suffers from the fact that it does not truly represent the whole institute. This institute is, in fact, made up of 19 astronomical observatories and other institutes, together with two facilities hosting telescopes. Each of these entities, geographically distributed in Italy, Spain, and the U.S.A (in regard to facilities) has its own website, (i.e., the INAF, Astronomical Observatory of Padua [22]) implemented in a different manner from the others and with its own graphical aspects, etc.

For these reasons, most information is duplicated (present on both the national site and on the local site), making it difficult to recognize the single entities as part of the same institute. Moreover, in the last period, another website called "Media inaf" [23] dedicated to the general public as the Fig. 2 shows, has been developed to distribute information.



INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS
Vent'anni sulla breccia
5X1000
IL CODICE FISCALE DELL'ISTITUTO NAZIONALE DI ASTROFISICA
972 2021 0583

Figure 2. The start page of the INAF website dedicated to the public

Because the content is spread among different websites, related only by links or by entirely duplicating data, information updates are not frequent (especially for some sections of the site) and there are duplicate web servers, bandwidth costs and webmasters needed to manage them.

Moreover, this structure does not promote efficiency, since each server relies on its own software and backup scheduling. Considering the amount of information, a single server developed as a web server farm could replace all those sites in order to provide reliability and load balancing.

From this aspect, the required restyling of the website, necessarily implies some consideration of the possibility of using a general-purpose server. Considering that the information for publishing already exists in a distributed way among the local websites, from a technical point of view, modern technologies could help to aggregate and dynamically present the needed information. Mashups are an emerging technology in the Web 2.0: in the words of Wikipedia, a mashup is a web site or web application "that seamlessly combines content from more than one source into an integrated experience". Content comes from different sites and is aggregated into a start page, which could be an innovative starting point for the institute.

The idea is to transform the website into a single homepage that collects and presents information as a result of careful analysis. The visual design project, developed in a second phase, could provide an appealing aspect while still adhering to web standards and web accessibility guidelines with respect to laws and to maintaining the original idea of the web, that of being devoted to all users, especially those with disabilities. The paper approaches the feasibility study in order to verify if it is possible to use existing tools to develop and create such a page as chunks of information visualized in sections that are dynamically updated? Mashups are used in a single homepage and in an end-user environment. Why not introduce them into the official homepage of an institute or an organization? What is the difference between these two technologies in different contexts? Both a generic websites developed with CMS technology and those developed with web mashups focus on web content. Content is essentially structured in a database format (essentially, relational or object database) or in an XML-like technology to be imported or transformed onto an HTML/XHTML page

III. MASHUP TECHNOLOGIES AND EXPORTED DATA FORMATS

A Mashup [24] is a hybrid web application that integrates data and functionality using web technologies to create new services and can in general run either on the server or on the client side.

The term, as Fig. 3 shows, implies fast integrations, APIs (Application Programming Interfaces), and data sources. Categorization divides a consumer mashup aimed at the general public to provide personalization of data/viewing according to users' needs, data mashups that combine similar types of media from multiple sources and information into a single representation, and business mashups that focus on a single presentation.

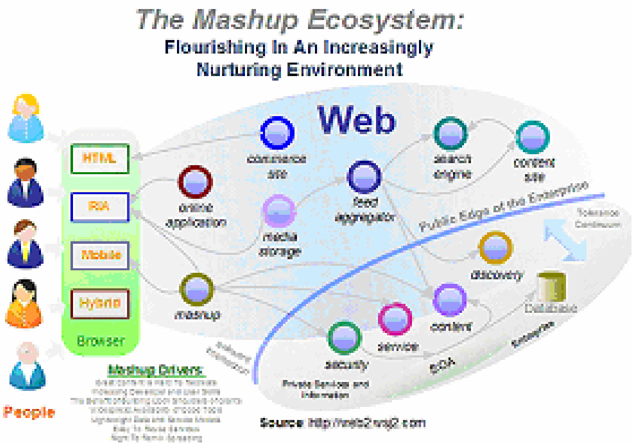


Figure 3. The Mashup ecosystem: involved components and technologies

Architecturally, there are two styles of mashups: web-based mashups that use a web browser to combine and reformat the data and server-based mashups that analyze and reformat the data on a remote server and transmit the data to the user's browser into its final form. For server side development, some technologies enable retrieving and processing data dynamically (i.e., PHP [25], Ruby [26] programming languages, etc.) and technologies implemented or supported in web browsers like Ajax, RSS, Adobe Flash [27], Microsoft Silverlight [28], etc.).

Both mashups and web applications generally use a browser as a client-side solution component that provides a user interface (UI) to a given application. They are centrally managed from a server in the enterprise, and both are deployed to the user's browser when the URL is used to access a web resource on the server. Business logic can be executed on systems, and data can be retrieved from databases, public data sources, or external services. Code that defines and makes up the UI and information extracted from internal and external systems is rendered within the browser, providing users with the means to view and act on the data. The content sent to the browser can include Javascript and scripting libraries that execute in the browser's runtime to allow custom logic to be run locally in the browser. A mashup can appear to be a variation of a Façade pattern, that is a design pattern that provides a simplified interface to a larger aggregate of different feeds with different APIs. A mashup can be used with software provided as a service (SaaS) [29]. Publishing and syndication formats require special attention since they are the glue that links mashed sites together [30].

Mashup selection can be performed dynamically identifying the situational set of widgets available at runtime. A widget is a software component used a by mashup assembler that encapsulates disparate data, user interface markups, and behavior into a single component that generates HTML code fragments. They are implemented as chunks of code that can be easily embedded in pages. Widgets are applications developed through RSS or Rich Internet Applications (RIA) [31] as means of combining web technologies (HTML/XHTML [32], CSS [32], ECMAscript [33], etc.).

Widgets are available on webtops (with drag and drop buttons) and by inserting source code or HTML pages. Mashups and widgets are available to end-users because most web 2.0 tools release example coding (usually as script in JavaScript programming) that can be integrated into a website. The technology takes advantages from structuring content in an XML-like format. Re-using and combining information and services is possible with a specific format (i.e., RSS or Atom format) using APIs, or web services developed in REST-style technology that have are based on HTML protocols. Even if the final goal is information publishing and thus HTML/XHTML and CSS language is used to visualize on a web browser such content, if information is exported in XML, transformation into whatever format is possible thanks to all XML technologies. In fact, most mashups include news feed available in RSS [8] or ATOM [9] format (which uses a markup language like XML) or Google maps, which make use of KML[11] format to structure information. The most common data formats that are used for exporting data are RSS and ATOM, even if XML languages or Google related formats are also used.

A. *Interoperability file formats*

JSON [10] is a lightweight data-interchange format based on a subset of standard ECMA-262, originally specified in RFC 4627 with an official Internet media type (application/json) and used for serializing and transmitting structured data over a network connection. JSON is primary used to transmit data between a server and a web application serving as an alternative to XML. JSON is supported by most browsers (i.e., Mozilla Firefox, Microsoft Internet Explorer, Opera and Webkit-based browsers such as Google Chrome, Apple Safari, i.e., browsers based on the webkit layout engine designed to allows web browsers to render web pages), and it is part of most popular Javascript libraries [34] like Prototype, jQuery, Dojo Toolkit, Mootols, Yahoo! UI library. XML is family of markup languages used to describe structured data and serialize objects with the primary goal for data interchange. When data are encoded in XML, the result is typically larger in size An examples used in such a context is KML (Keyhole Markup language) [11], an XML-based languages used to work with geographic data such as those used by Google Maps. KML uses a tag-based structure with nested elements and attributes (and an extension .kml o .kmz as zipped format). Finally, RSS and ATOM are the most frequently used as web feed formats to publish content as blog entries, news headlines, audio, and video. ATOM is specifically an XML language used for web feeds and it has a related protocol (the Atom Publishing protocol AtomPub or APP [35]) that is a simple HTTP-based protocol for creating and updating web resources. This format was published as an IETF proposed standard in RFC 4287, while the protocol in RFC 5023.

B. User web mashups

The primary application of mashups is in the area of end-users, since these activities involve little programming if a user chooses an already-publicized service. A webtop such as iGoogle is an example of a mashup that helps user to individually configure a start page with specific information. Users can propose favorite preferences and compose various widgets in a mashboard. This result in a new sort of interactive web applications whose effect gives new values to the end user. Web mashups use either public APIs from existing sites and/or web scraping to collect data and combine them on a single web page. Examples are mashup web applications using Twitter web services and Google maps services to display locations mentioned. Several examples exist in literature by using different programming languages [36] for data extraction on the server side, even if an alternative approach is to use visual tools [37].

Any public web applications that allows users to enter account information for other web sites needs to take suitable precautions safeguarding this information, and this is difficult dealing with thirdy-party web portals since all information should be stored in plain-text (i.e, Google and Twitter credentials). In this case, it is necessary to adopt standards for open authentication APIs like OAuth [38] a protocol that allows single sign-on solutions for authentication on web sites.

The literature and websites report many examples of how to create a mashup, but all of them concern the inclusion of so-called web 2.0 tools (i.e., Facebook profiles, Twitter connections, and so on) by using different visual mashup editors. Unfortunately, most initially available have been dismissed (i.e., Microsoft Popfly or Google Mashup editor that is migrated to Google App Engine [30]). Yahoo pipes [40] remains one of the tools useful for visually merging content from different sources and in some way “free” even if it requires a Yahoo authentication. IBM QEDWiki has graduated to IBM Lotus Mashup [41] as a commercial tool.

IV. MASHUPS FOR AN INSTITUTIONAL WEBSITE

The use of mashup to address institutional website requires available data sources. The great obstacle to this solution, whose resolution could free up many activities in the actual web management and could contribute to reducing the duplicating of information, is that each website needs to publish information in a format that can be easily collected. Actually, APIs available to include photos on Flickr [42] or sales item in eBay [43] refer to commercial sites and need a sign-up and thus specific credentials. In fact, usually mashup-maker tools use some programming languages for data extraction from an external data source on the server side and combine data together. Most of data that will be combined should be available in an XML-like format (i.e., RSS/Atom feed publishing and syndication formats that are used for exporting data) to be retrieved and processed dynamically. The problems are both how to export information to single sites in this way and how to collect them in order to design the homepage of a national institute. Until now, all mashup technologies have been devoted to end-users, since the aim is to create a user startup page suited to each user in a manner similar to portals. However, its use for an institutional website should be the collection of specific content, available as widgets, in order to form a single page

This represents a great constraint: most sites export data in the RSS feed format, but only for limited information, such as news content. Creating a dynamic start page for an institute involves different types of content being combined with static information (i.e., information about the main features of the institute and its organization, which are not constantly updated). However, using such software requires an availability of exported data and is difficult when integrating static information. The final start page should be a composite aggregation of dynamic and static information devoted to all kinds of people that use such a site. And content should be differentiated if the user is an internal employee, a professional astronomer, or simply a person who likes astronomy.

The mashup for this goal is necessarily server-based that combines sources from different websites. Simple techniques use cutting and pasting snippets of Javascript, using RSS feeds and XML to connect the various parts together and even on-line js inclusion that can pull in and integrate a powerful external component.

A first idea on the schematic layout of the homepage is presented in Fig. 4. This first prototype includes some external APIs from major thirdy-party sources (i.e., Google Maps, Flickr, Facebook, etc.) and several data from different RSS/ATOM feeds. This solution could be easily implemented since all single local sites export al least some information as RSS feed, especially those regarding all news (internal and external) about local activities and projects.

If such a technology could be easily used to merge content available on different sites, there should be savings in hardware, software and human resources.

Moreover, mashup technology, even if born around the single user and his or her ability to create his or her own content, functionalities, and services could be adapted to reach a group of users (i.e., a different category of users to whom a website is devoted).

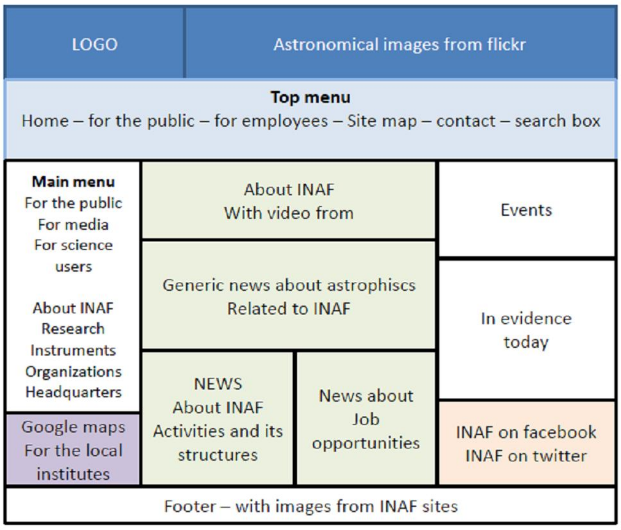


Figure 4. A draft of a webpage layout

A mashup could be a way to extend the concept of using Web 2.0 tools in an enterprise context [44] being a tool to create a innovative website.

A. Combining user partecipation in website and the need of controlling the published information

In our opinion, Web 2.0 is a way to allow end-users to insert their own content and functionalities in a website that could became a work tool that uses of a set of web technologies related to the fact the content should in some way imported and published in a customized manner. Close to the concerns of exporting data and services through the network, there is the problem of allowing different users to choose data and services to be published and at the same time avoiding transferring full control of the information to the end-users. We have outlined several categories of users who will be involved with our website. It is logical that internal users require information different from external users (science users or the media for example). Details of users is shown in Fig. 5.

The concern is to export data as feeds or in other XM-like format for all relevant information (i.e., customized different news) and to create some web services able to include other types of content in the page such the multimedia content that is non-necessary stored on a Web 2.0 websites such as YouTube or Flickr.

This is realized with a registration by e-mail in the iGoogle approach or by including a special key to be used in the exported API. However, this approach seems to be too user-customizable in an institutional website, and a tradeoff should be found to mix these requirements. The state of art of mashup technologies and literatures do not seem to face with this problem probably because the main aim of the instrument devoted to end-user.

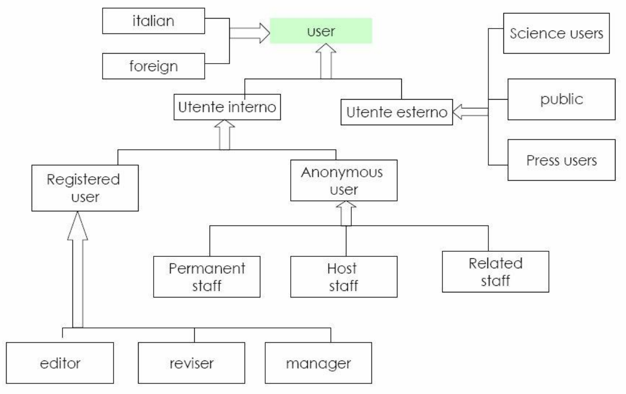


Figure 5. Users' categorization according an analysis of thw website's visitors

But, we think that this solution will provide many savings from technical and content viewpoint, and thus this project will be the aim of next months. Probably the main concern will be to have all main content structured in an XML-like format in order to apply XML technologies [45] (i.e., XSLT) helping in processing files and trasforming them in a format useful to be published on a web page.

V. CONCLUSIONS

This work is a preliminary discussion of the possibility of proposing a mashup as a substitution for a generic website. The need to avoid information duplication and for a website that is not used by its employees has lead to an analysis of the possibility of applying a different approach in the design of a website. Mashups and widgets could possibly be applied in such a solution, but it is necessary to make a further study in analyzing how to actually implement such a technology and its tools. One such solution is the use of mashup-makers for composing personal homepages, but could this tool also be applicable for a general homepage, such as that of a research institute? And how it is possible to define from which part of the homepage to collect information? For example, an important step is to understand how content should be exported in order to be collected (in an XML format, such as an RSS or Atom feed) and if a local website could contribute by implementing a web service interface through an API. All these requirements will be taken into consideration by implementing specific tools able to create a composite startpage. However, the first prototype that includes some features peculiarly to a web mashup has been developed. It collects the web feeds available from different website in order to compose a list of information about news in the astrophysical area and content about the institute's activities. Moreover some web services exported by famous commercial sites (i.e., Google, Facebook, and so on) could easily be embedded in the page. A Google map could help locate several institutes, and Google Docs could be useful for managing administrative documents before and after the publication on the website. A problem here is that information is strictly related to this site (i.e., there is the need for a specific Google key in order to import such a service) and related to business activities.

REFERENCES

[1] D. Addey, J. Ellis, P.Suh, and D. Thiemecke, Content Management Systems (Tool of the Trade), Apress, 2003.

[2] T. O'Reilly, What Is Web 2.0: Design Patterns and Business Models for th next generation of software, O'Reilly Press, 2005.

[3] YouTube site, http://www.youtube.com, 10.07.2010

[4] Facebook site, http://www.facebook.com, 10.07.2010

[5] Twitter site, http://twitter.com, 10.07.2010

[6] D. Tapscott, Enterprise 2.0, how social software will change the future of work, Gower Publishing, 2008.

[7] XML (eXtensible Markup Language) Activity Statement, http://www.w3.org/XML/Activity, 10.07.2010.

[8] RSS (Really Simple Syndication) board, http://www.rssboard.org, 09.07.2010.

[9] The Atom Syndication Format,.http://tools.ietf.org/html/rfc4287. 09.07.2010.

[10] JavaScript Object Notation (JSON), http://www.json.org, 09.07.2010.

[11] KML (keyhole markup language Tutorial http://code.google.com/apis/kml/documentation/kml_tut.html, 09.07.2010.

[12] Wordpress site, http://wordpress.org, 10.07.2010.

[13] Joomla!, http://www.joomla.org, 10.07.2010.

[14] Drupal, http://drupal.org, 10.07.2010.

[15] Plone CMS, http://plone.org, 10.07.2010.

[16] R. Jee, Pro Web 2.0 Mashups, remixing data and web services, Apress, 2008.

[17] R.L. Costello, Building web services the REST Way, http://www.xfront.com/REST-Web-Services.html, 10.07.2010

[18] Ajax and other "rich" interface technologies, http://www.owasp.org/index.php/Ajax_and_Other_%22Rich%22_Interface_Technologies, 10.07.2010.

[19] N. Conner, Google Apps: The missing manual, Pougue Press, 2008.

[20] Te National Institute of Astrophysics (INAF) website, http://www.inaf.it, 09.07.2010

[21] C. Boccato and S. Pastore, "The Web Information System of the National Institute for Astrophysics: different actors contributing to disseminate information". In Current Research in Information Sciences and Technologies, Multidisciplinary approaches to global information system, Proc. Of Int. Conf. on Multidisciplinary Information Sciences and Technologies (INSCIT2006), Merida, Spain, October, 2006, Vol. I, pp. 507-511.

[22] INAF OaPD, http://www.oapd.inaf.it, 09.07.2010.

[23] Media INAF website, http://media.inaf.it, 09.07.2010.

[24] D. Merrill, Mashups: the new breed of web app. An Introdution to mashups, IBM developerworks, 2009.

[25] PHP: Hypertext Preprocessor, http://php.net, 09.07.2010

[26] Ruby progamming language, http://ruby-lang.org, 09.07.2010

[27] Abobe Flash Platform Technologies, http:// http://labs.adobe.com/technologies/flash/, 10.07.2010.

[28] Microsoft Silverlight, http://www.silverlight.net, 10.07.2010

[29] Tim O'Reilly, Open Source Paradigm Shift, O'Reilly Media, http://tim.oreilly.com/articles/paradigmshift_0504.html, 13.07.2010.

[30] M. Watson, Scripting Intelligence: Web 3.0 Information Gathering and Processing, Apress, 2009, pp.269

[31] Putting the power of Web 2.0 into practice. How rich Internet Applications (RIA) can deliver tangible business benefits, IBM white paper, 2008.

[32] HTML &CSS, http://www.w3.org/standards/webdesign/htmlcss, 09.07.2010.

[33] ECMAScript programming language, http://www.ecmascript.org/, 09.07.2010.

[34] Javascript librieries, http://javascriptlibraries.com 09.10.2010.

[35] The Atom Publishing Protocol, http://www.ietf.org/rfc/rfc5023.txt, 09.07.2010.

[36] O. Beletski, "End User Mashup Programming Environments", Helsinky University of Technology, Telecommunications Software and Multimedia Laboratory, Seminar on Multimedia, 2008.

[37] A. Spoerri, "Visual Mashup of Text and Media Search results, in Proceedings of the 11th International Conference Information Visualization," IEEE Computer Society, 2007, pp. 125-134.

[38] OAuth Community site, http://oauth.net, 09.07.2010.

[39] Google App Engine, http://code.google.com/appengine, 10.07.2010.

[40] Yahoo pipes, http://pipes.yahoo.com/pipes, 09.07.2010

[41] IBM Mashup Center, http://www-01.ibm.com/software/info/mashup-center/, 10.07.2010.

[42] Flickr Photo Sharing, http://www.flickr.com, 10.07.2010.

[43] eBay, http://shop.ebaby.com, 10.07.2010.

[44] A. McAfee, Enterprise 2.0: new collaborative tools for your organization's toughest challenges, Harvard Business School Press, 2009.

[45] XML Technology, http://www.w3.org/standards/xml/, 13.07.2010

.

ISBN: 978-1-61208-100-7