OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

DISCLAIMER

FAIR USE POLICY

CONTACT

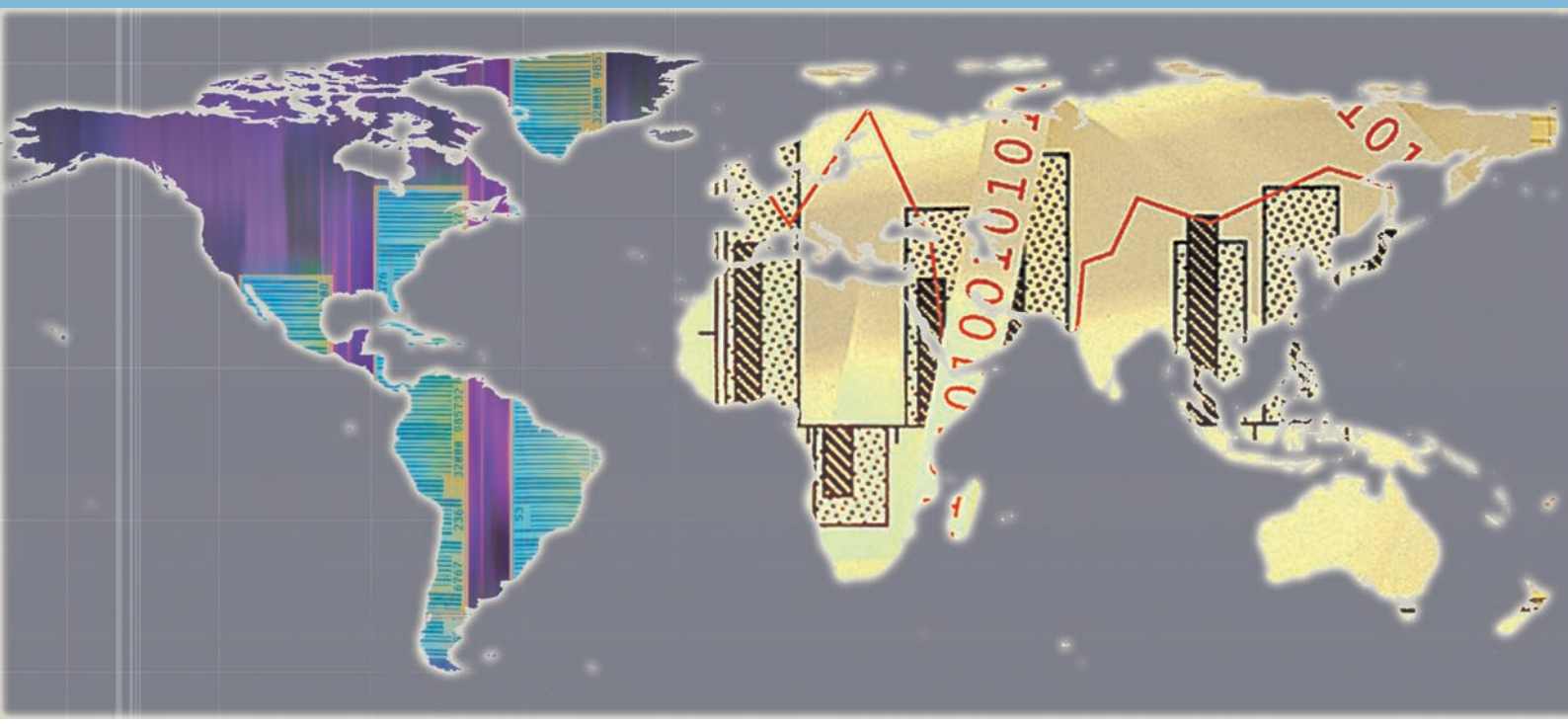
Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

Research and Statistics Branch

Working Paper 10/2009



Virtual Teams:
Practical Guide to Wikis and other Collaboration Tools

UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

RESEARCH AND STATISTICS BRANCH
WORKING PAPER 10/2009

Virtual Teams: Practical Guide to Wikis and other Collaboration Tools

Valentin Todorov
Research and Statistics Branch
Programme Coordination and Field Operations Division
UNIDO

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna, 2010

Acknowledgements
We are thankful to Peter Holzleitner from UNIDO, Huzaifa Zoomkawala from "HM Consultants" and the participants of the MSIS Task Force on sharing software, particularly Alistair Hamilton, Marton Vuscan and Carlo Vaccari for provided information, comments and suggestions and to Diman Todorov from Cardiff University for coauthoring an earlier version of the paper.

Table of Contents

List of Tables

List of Figures

1 Introduction

The term Web 2.0 refers to collaboration on the World Wide Web. Prominent examples for Web 2.0 applications are *flickr*, *facebook* or *del.icio.us*. The 2.0 version identifier plays at the fact that, prior to Web 2.0, the web communicated with its users not in a dialogue but, similarly to print media, in a monologue. Content was published by a single author and was not modifiable by consumers. Interaction with the web site was rudimentary at best. Now the World Wide Web has become a platform where the users benefit from a greater interaction with the applications and advanced and at the same time easier communication between the content producers, i.e. the users themselves.

The purpose of the paper at hand is to provide an overview over the strengths and weaknesses of wikis as a tool for collaboration. First, we introduce the concept of wikis. Next we provide a brief overview over technical characteristics of wikis. We compare wikis to other collaboration tools like mailing lists, forums and blogs, paying particular attention to points characteristic in wikis. We further provide the historical and social context of the emergence of wikis. Having drawn this picture of wikis and their context we present scenarios in which the application of wikis is particularly appropriate or not. In conclusion it can be said that while wikis are a powerful concept they leave many questions open. Wikis are well suited in social environments with a flat hierarchical structure in which there is no single person or committee responsible for maintaining content.

The already mentioned above *del.icio.us* is one of the most popular social bookmarking services which is a method for allowing the Internet users to share, organize, search, and manage bookmarks of web resources. Closely related to the socual bookmarking is the collaborative reference management and these two topics will be briefly discussed in Section 7.1.

1.1 What is a Wiki

Wikis appeared on the internet landscape in the mid 1990's. They were developed by Ward Cunningham (Leuf and Cunningham, 2001) and named after the Hawaiian word for "quick". Wikis can be used in different contexts mainly for developing written text, not only WWW but also locally or Intranet. A wiki enables documents to be written collaboratively, in a simple markup language using only a Web browser. They do not require users to install additional software. Meaningful topic associations between different pages are created by making page links. This is almost intuitively easy and shows whether an intended target page exists or not.

Louridas (2006) extends the definition of a Wiki also to the software that makes it possible for anyone to edit web sites and the philosophy surrounding how users edit these web pages. The philosophy and the approach to collaborative editing the web sites is usually more important than the technology that supports it.

Wikis are considered a Web 2.0 technology. In order to appreciate the implications of this

categorization let us briefly review the characteristics of Web 2.0. The Web 2.0 world is distinguished by a set of tools and technologies as well as a particular vision of the business strategies that makes explicit reference to new social and cultural trends oriented to foster relationships between actors in the network. These are tools that allow the users to create directly content on the web (*user generated content*), social networking, collaborative authoring (*wiki*), tools for publishing information and opinions (*blog*) and tools for "labeling" the content of web sites (*social bookmarking* and *folksonomy*). Our personal opinion is that the name "Web 2.0" is misleading as it hints at a further development of the World Wide Web. Originally the web was intended to be an interactive collaboration platform (Berners-Lee and Fischetti, 1999) and the reason why it was not collaborative from the very beginning is that at the time the technological means were not sufficient for such a demanding task.

There is a wide discussion over the set of characteristics defining an application as being "Web 2.0". One commonly accepted set of characteristics is:

- participatory
- decentralized
- linked
- emergent

Wikis fulfill all of these criteria. They are participatory, because every consumer can participate in the maintenance of the content. They are decentralized because they work best when they are managed not by a single person or committee but when the responsibility over the content is shared among the users of a wiki—this claim will be elaborated later in section 3.3. They are linked because single pages in a wiki are often only meaningful in the larger context of the entire content of a wiki. They are emergent because wikis are a relatively new technology, the flexibility and limitations of which are still being tested and at the same time their content rather than relying on fully predefined structures is allowed to emerge over time, their success comes from cooperation, not control.

Judging by the purpose and functionality of wikis they can be categorized as a kind of content management system (CMS). Unlike traditional CMSs, however wikis are unstructured. There is no imposed structure on content. While this means that a wiki navigation path can be shaped in any desired way, it also means that the content graph can become unwieldy and impractical if its growth is not being monitored. Content is not the only part of wikis that is unstructured—wikis also lack user hierarchies. Although there are methods to restrict editing rights on wikis, the intent is to grant write access to anyone who has access to the wiki. This has several social implications which will be discussed in more detail in section 3.2.

Wikis were designed upon the premise that content is never complete or error free. To allow for faster convergence towards optimal content wikis allow as many editors as possible and give them as much freedom as possible. As radical as this approach may appear at first, it is actually the way the World Wide Web was envisioned in the early 1990s.

Figure 1: An example of wiki syntax (from Wikipedia:Cheatsheet). The leftmost column describes what you want to do, in the middle column is shown the necessary command in wiki syntax and the rightmost column presents the achieved result. The most basic commands are for formatting (**bold**, *italics*, ***bold and italics***), sections and headings, bulleted and numbered lists, internal and external links and images.

Wikipedia Cheatsheet

Shortcut: WP:CHEAT

For more advanced details, see Help:Wiki markup

Description	You type	You get
Italic text	`''italic''`	*italic*
Bold text	`'''bold'''`	**bold**
Bold and italic	`'''''bold & italic'''''`	***bold & italic***
Internal link *(within Wikipedia)*	`[[Name of page]]` `[[Name of page|Text to display]]`	Name of page Text to display

External link *(to other websites)*	`[http://www.example.org Text to display]`
	`[http://www.example.org]`
	`http://www.example.org`
		Text to display
		[1]
		http://www.example.org

Redirect to another page	`#REDIRECT [[Target page]]`	↳ Target page

1.2 Technical Aspects

In this section we will cover some technical aspects characteristic of wiki software. Content is entered into wikis using a special markup language that provides syntax for simple formatting of text as well as inserting links to other documents and external content. Figure 1 shows an example of wiki syntax. Some of the basic commands are for formatting (**bold**, *italics*, ***bold and italics***), sections and headings, bulleted and numbered lists, internal and external links and images. For more details see for example `http://en.wikipedia.org/wiki/Wikipedia:Cheatsheet`.

Although the markup language is kept simple, it poses a significant entry barrier. Its advantage is that it can be used on any web browser and on any device with web browsing support. Practically all current wiki software support WYSIWYG (What You See Is What You Get) content entry. The WYSIWYG editor makes all the features of the markup language accessible from a toolbar. Text entry using the toolbar may be slower because users will need to alternate between the keyboard and the mouse often. Although the WYSIWYG editor is often only supported on the most common web browsers, its main advantage, a low entry barrier, makes its existence legitimate. Examples of wiki toolbars are shown in Figure 2. The upper panel presents the *Mediawiki* toolbar and the lower - the *Wikispaces* one. In both shown toolbars the most basic commands for formatting (**bold**, *italics*, ***bold and italics***), sections and headings, bulleted and numbered lists, internal and external links and images) are accessible by clicking the corresponding button. The code shown below represents the internals of the wiki page in Figure 3.

Figure 2: Examples of wiki toolbars which have buttons for easier access to commonly used wiki syntax commands. The upper panel shows the *Mediawiki* edit toolbar and the lower - the *Wikispaces* one.

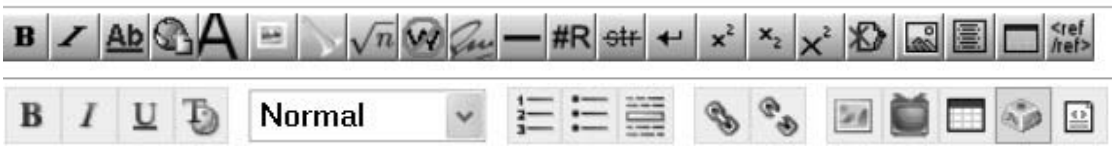
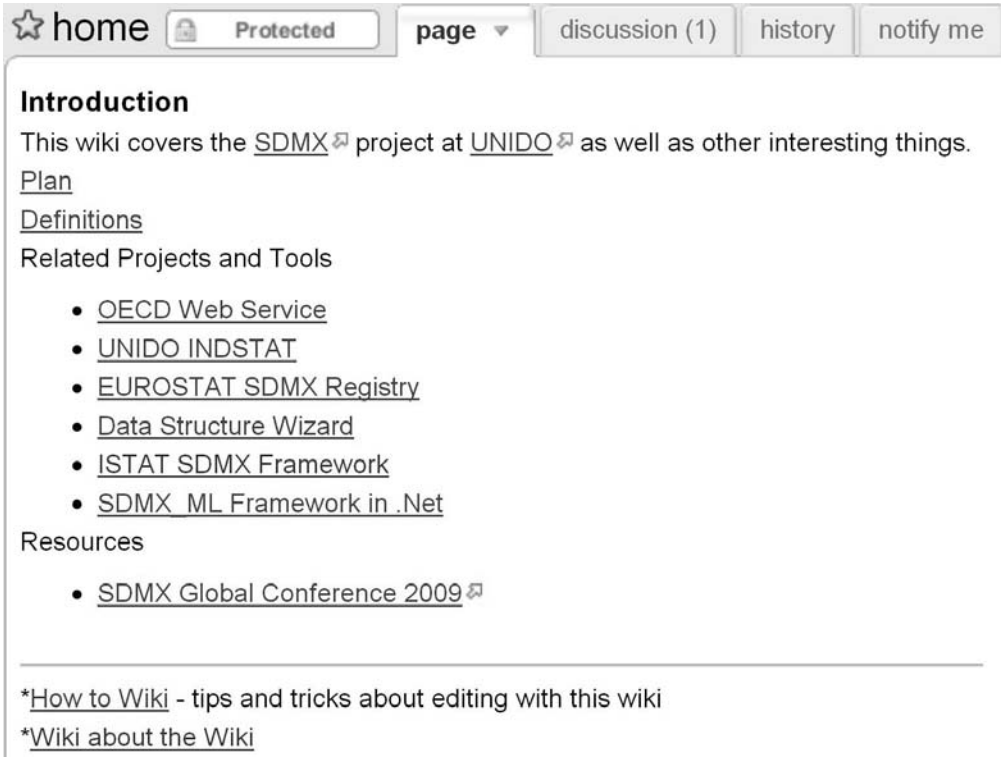


Figure 3: Example of a wiki page (UNIDO SDMX page, see Section 4.2 on *Wikispaces*).



```
====**Introduction**====
This wiki covers the [[http://sdmx.org|SDMX]] project at
[[http://unido.org|UNIDO]] as well as other interesting things.
[[Plan]]
[[Definitions]]
Related Projects and Tools
* [[OECD Web Service]]
* [[UNIDO INDSTAT]]
* [[Eurostat|EUROSTAT SDMX Registry]]
* [[Data Structure Wizard]]
* [[ISTATSDMXFramework|ISTAT SDMX Framework]]
* [[SDMX_ML Framework in .Net]]
Resources
* [[http://www.oecd.org/dataoecd/6/54/41959211.pdf?
    contentId=41080934|SDMX Global Conference 2009]]
*[[How to Wiki]] - tips and tricks about editing with this wiki
*[[Wiki about the Wiki]]
```

As outlined in the previous section, one of the main characteristics of a wiki is that its content is dynamic. Wikis implement several mechanisms to ensure a high standard of the quality of content in spite of its fluidity. A complete revision history of every document is stored so that after an erroneous or malicious change the content can be easily reverted to a previous state. Users can opt to receive emails when the content of a document is changed. This may result in an Inbox full of notifications of typo corrections. To prevent this, editors can mark changes as minor. Minor changes are not sent to all users for review. Users can also subscribe to an RSS feed (*Really Simple Syndication*, which itself is one of the main "components" of Web 2.0), publishing major changes instead of requesting emails. The main advantage is that RSS infrastructure is better suited for this news flash style solicitation of information.

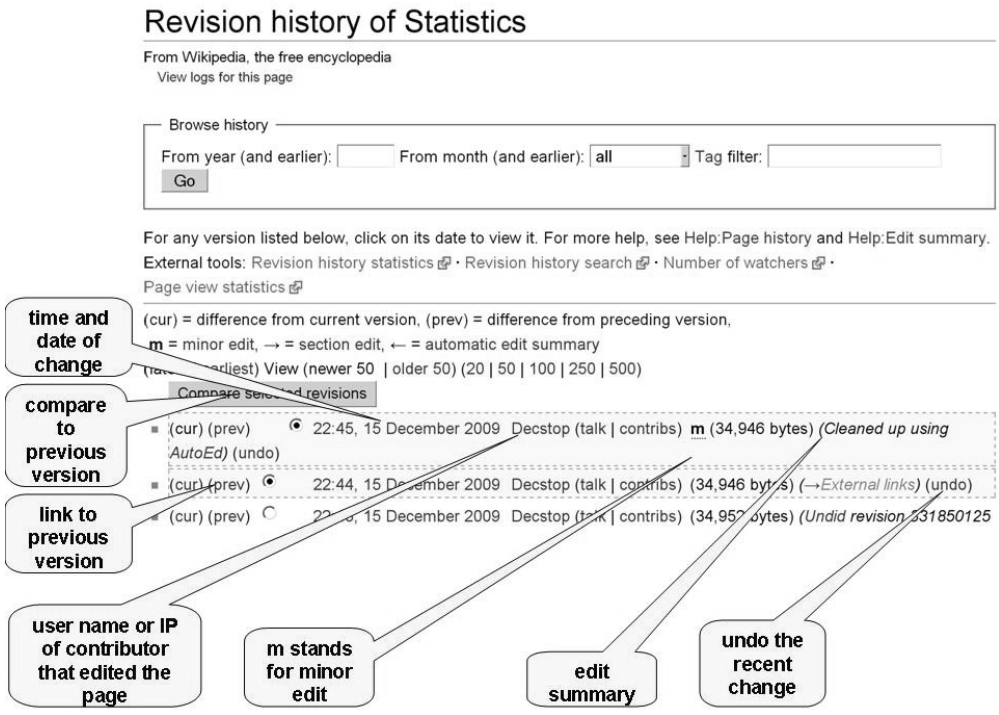
Wikis are available through a wide variety of services and open-source software tools. Generally there are the following two categories of wikis:

- Wiki services or Wiki farms: the wiki pages are hosted at the service provider server or array of servers, require no local software installation and can be either free or fee-based. A comparison of many Wiki farms can be found at `http://en.wikipedia.org/wiki/Wiki_farm`. Also very helpful, especially if one is looking for a very simple solution, is the following blog entry and the subsequent discussion: `http://pascal.vanhecke.info/2005/10/30/`.
- Self-hosted Wikis: there exists a variety of open source wiki software like MediaWiki and TWiki. The advantage of self-hosting the wiki allows for maximum control over the access and security as well for fulfillment of specific requirements. A disadvantage is the necessity of an own server and some technical and network experience which results in longer start-up time.

Choosing the right wiki can be based on answering some question regarding the key features of the wiki, the available resources and technical support:

- How many users?
- Are separate groups necessary?
- Is interaction between the groups necessary?
- How secure should the pages be?
- Public or private pages?
- How skilled are the participants (simple markup or WYSIWYG)?
- How important is the layout?

Figure 4: An example of wiki revision history (as shown by Wikipedia). All past changes to the selected page are listed in reverse-chronological order. One can view a specific version (click a date) or compare versions(click *cur* or *prev*).



2 Comparison with Similar Technologies

To present a profile of wikis we will compare them with other collaborative technology. We will compare wikis with mailing lists, forums and newsgroups, blogs, traditional content management systems and powerful proprietary collaboration tools like Microsoft SharePoint and Lotus Notes/Lotus Domino.

Mailing lists. A simple and widely spread collaboration method are mailing lists. In a mailing list emails are sent to a designated email address. The receiver of that address is not a person but

a software, which broadcasts received emails to all subscribers. Although it is possible to send emails to all receivers in the first place, the automated method makes it simpler to administrate subscribers of the mailing list. The main difference between mailing lists and wikis is that mailing lists are private. The receivers are usually restricted and although there is some support for publishing mailing lists on the web the information circulated through mailing lists is usually not accessible to users who are not subscribed. It is interesting to note that although it may not be obvious at first sight, archived mailing lists provide functionality to achieve tasks similar to the ones accomplished with a wiki.

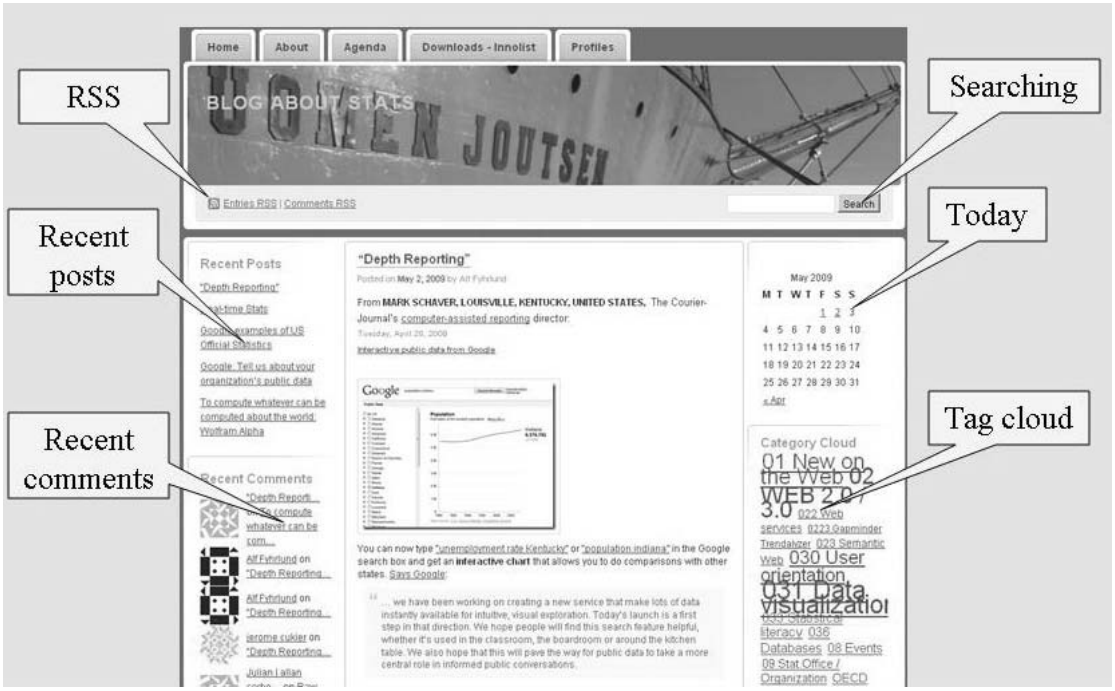
Web forums. Another comparable technology are newsgroups and their more modern counterpart, web forums. Although the difference between wikis and forums seems obvious, it is tricky to define precisely. Unlike mailing lists, in forums users can change their own content after it has been posted. The archival of communication threads is intrinsic to the medium. The difference to wikis seems to be mainly in spirit: while wikis are more document centric newsgroups forums are more communication centric. A forum with sufficient article persistency can be used in the same way as a wiki. However, wikis provide better support for authoring, retrieval and interrelation of documents. For example URLs to documents in wikis are often self explanatory, whereas in forums URLs are identified by a unique numeric ID, which does not give any clues about the content at the location. The qualities of both systems are difficult to compare because their performance depends heavily on the culture and discipline of the community they are used in.

Blogs. Although blogs are not a collaboration tool in the strictest meaning, they are still covered here since they are a significant step up from the monologues Web 1.0. Blogs, short for web logs, are comparable to periodic columns in print media. The main difference is that they usually have an interactive comments section right beneath each article. The main articles are usually written by one or few authors. The comments however can be written anonymously. This policy may be a good tradeoff between the strict access control in traditional content management systems and the liberal approach in wikis. The content in blogs is organized chronologically, hence the name web logs. The content published by the *blogger* can be a specific subject, personal information or a combination of these. Thus several cases of blogging can be distinguished (see Grossenbacher, 2009):

1. personal blogs (personal content would not be mixed up with companies content)
2. personal blogs overlapping with topics of the company or institution
3. official blogs of a company or institution, general or restricted to a special theme

The above cited *Blog About Stats* `http://blogstats.wordpress.com/` is a good example of the second type - it is published by Armin Grossenbacher from the Swiss Federal Statistical Office (FSO) and is not an official blog. As its name suggests the subject of the blog is *All about dissemination of Official Statistics* and is intended to be a multi-author blog. It provides the latest news and information about developments in communicating statistics on the Web. In Figure 5 is shown an example page of this blog illustrating the main elements - an RSS feed link, panels for recent posts and recent comments, calendar and a tag cloud.

Figure 5: Example of a blog main page: *Blog About Stats* - the blog of Armin Grossenbacher with a subject *All about dissemination of Official Statistics*.



When considering the activities of statistical offices, the "blogging" itself is not the only form of presence in the blogosphere - another angle is the monitoring of the blog posts or comments written by others about statistical offices or products (Gardner, 2009) (which can be seen as part of a complete media monitoring strategy). For example, Statistics Denmark has reported at a recent UNECE meeting that they systematically check the blogosphere for relevant posts, record them in a database and respond where and when necessary (Gardner, 2009). Table 1 summarizes the main features of these three (blogs, web forums, wiki) technologies (see Wagner and Bolloju, 2005).

Content management systems. As mentioned above, wikis are a subgroup of content management systems (CMS), but they have some peculiarities which make them stand out from what is commonly understood under the concept of a CMS. While the features and functionalities are the same, a different emphasis is placed on the same features in a CMS and in a wiki. Generally a wiki is more open for authorship. The support for authorization and authentication is less sophisticated than in a CMS. Because of the open authoring policy wikis implement elaborate inter personal conflict resolution mechanisms which are usually not found in a CMS. Another difference, which is more a guideline than a rule, is that wikis can be implemented efficiently relying solely on free open source products while traditional CMS are often expensive, proprietary solutions (with some exceptions like Joomla, see also `http://www.cmsmatrix.org`). The last major difference is the organization of content: in a CMS content is organized in hierarchies while the organization structure of wikis is flat.

Table 1: Comparison of different collaboration tools

	Web forum	**Blog**	**Wiki**
Speed of publication	Yes	Yes	Yes
Ease of publication	Yes	Yes	Yes
Knowledge representation	Chronological organization	Chronological organization	Topical organization as well as chronology of changes
Team support	Open or closed set of members; moderators	Individual publishing but some tools offer team support	Inherently open to public but can be restricted to closed groups
Security	Yes	Yes	Yes
Version management	Not provided	Not provided	Versions and history changes are provided; rollback possible

Microsoft SharePoint. A completely different platform for collaboration on the web is provided by Microsoft SharePoint products and technologies. SharePoint web sites and pages are commonly used to build Intranet and Extranet portals and team sites, as well as public Internet sites. It shows maturity in terms of user interface, database design, workflow and communication features. SharePoint includes two platforms: Windows SharePoint Services (WSS) and Microsoft Office SharePoint Server (MOSS). WSS is more basic and is used to create web sites for team collaboration on a common project. It can serve small companies and individual departments. It comes as a free extension to the Microsoft Windows Server 2003 and higher. MOSS builds on WSS and provides capabilities for portal publishing, enterprise search, enterprise content management (ECM), and many more. MOSS targets the management and control of a company's diverse knowledge assets. It is available in Standard and Enterprise Editions. Both WSS and MOSS are .NET 2.0 applications with XML web service interaction layers and ASP.NET presentation layers. SharePoint is highly integrated with MS Office and thus very strong when type specific documents (Word, PowerPoint, Excel) are concerned. It integrates also nicely with MS SQL Server. But SharePoint ultimately turns into just a file share. Another disadvantage is the required maintenance of the user access rights. SharePoint 2007 includes a wiki but it is very low profile when it comes to be benchmarked with other wiki engines, not to consider its price. In summary, it is not necessary to compare SharePoint to wikis because they have completely different purposes (complexity vs. simplicity, expensive vs. free) but rather SharePoint has to be compared to other Content Management Systems like *Joomla*, *Alfresco*, *drupal* and others.

Lotus Notes/Lotus Domino 7.0 Lotus is one of the original players in the collaboration and messaging market and has made a name with business users and IT specialists. Basically there are two products - a Lotus Domino Server which provides the back-end services and Lotus Notes serving as a client application. The tool Lotus Notes Designer helps for developing collaborative applications but it requires specialized skills. Lotus Notes has good interoperability with DB2, JSP and XML but poor third party support. Similarly as for SharePoint, there is some integration possible between wikis and Notes but different communities tend to draw the dividing line between Notes content and wiki content differently, depending on what suits the preferences and "culture" of the particular community. However, Lotus Notes tends to be preferred for hosting large reference documents, or more formal documents, within the Intranet.

3 Purpose, Strengths and Limits

Wikis are a collaboration tool. In the context of this paper we treat wikis as a means for collaborative knowledge management. To understand better the tradeoffs involved in wiki collaboration we present a historical background of wikis, a treatment of the social mechanisms at work in communities using wikis and finally we present some best practices for the successful implementation of wikis.

3.1 Historical Background

To understand the social context and original intent of wikis it is necessary to look at the history of their development. Wikis were made famous by the open source community. They were used as a simple documentation tool for projects. Since documentation is often written by the people who wrote the software, the documentation tool needed to cater to the organization structure of open source projects. The structure of open source projects is covered in great detail by Raymond (2001). This kind of projects have several characteristics usually not found in commercial projects. The hierarchical structure of the contributors is flat. The release cycle is very short, Raymond goes as far as phrasing the mantra "publish soon, publish often". Every user is a potential contributor. The schedule of the project needs to be flexible—in many large open source projects there is non stop development because there is always a time zone with contributors currently working on the project. It is easy to see how wikis fulfill all these requirements.

Eric Raymond explains the motivation driving open source software as the need to scratch an itch. In this metaphor contribution to open source software is treated as the solution to a minor annoyance. Although there is no monetary return, the effort required to solve a minor annoyance is low. Writing documentation however does not add to the solution of the annoyance, it just makes the solution viable for other people than the contributor at an added effort. To make open source contributors document their software at all they must be able to do this in a simple and quick manner. What makes the documentation complete is the sum of small contributions of all programmers. To make the process of documentation even simpler, the need for categorization of information is removed all together.

3.2 Social Mechanisms

Wikis are a tool for the externalization of knowledge. As such knowledge is transferred and synthesized along several vectors in a wiki community. To externalize knowledge an individual writes or changes a wiki article. However, the process of externalization can lead to individual learning processes in the contributor because in order to write something down a clear understanding of the matter is required. Knowledge in wikis also flows in the opposite direction. When users read articles, they internalize the knowledge in the wiki corpus, this leads to expansion of the individual's knowledge on a topic. The crucial flow of information in wikis however is the synthesis of knowledge which was previously neither present in the wiki nor in the individual. When newly acquired knowledge interacts with previous knowledge the so called emergent knowledge is created. Emergent knowledge is a direct result of collaboration, it is more than the mere sharing of knowledge. Within this theoretical structure an attempt can be made to explain what motivates the contribution of knowledge to a wiki. Cress and Kimmerle (2008) propose the theory that what drives people to edit wikis is an incongruence between the knowledge in the wiki and the internal knowledge of the contributor. One such incongruence could be information which is missing in the wiki but present in the internal knowledge. Another incongruence is conflicting information—when the wiki contains information which is contradictory to internal knowledge a reader is prompted to edit the article. Cress and Kimmerle (2008) model the motivation for contribution as a function of the size of incongruity and the importance (valence) of the topic to the contributor. Curiously dealing with the resolution of contradiction conflicts between different users and the corpus in the wiki is also the largest challenge in running a wiki. Failure to deal with such conflicts leads to the inevitable failure of a wiki.

As mentioned above, wikis emerged as a documentation tool for software projects. When writing documentation usually all programmers document their own modules. The subtle implication of this setting is that there are hardly ever inter-personal conflicts about the content. Such conflicts are possibly the most obvious reason not to use a wiki. Although most wiki software implements elaborate conflict resolution mechanisms, all of them fail without a fair and clear policy as well as the willingness of the community to adhere to such a policy. Some of the conflict resolution mechanisms are discussion forums, moderation or user authentication. Authentication is targeted specifically at conveying a sense of responsibility to users and thus preventing vandalism. A possible policy to deal with controversy is to assure that all aspects of a controversial topic are covered in an article—this is the modus operandi in the online encyclopedia Wikipedia.

3.3 Best Practices

The historical roots of wikis have strong implications for their applicability in environments outside of the open source community. For wikis to be useful a large user base is needed. Wikis work best when the organization structure of the users is flat. In commercial environments with a strict hierarchy wikis often do not work well. In a hierarchy subordinates often do not carry the initiative required for the productive implementation of a wiki. For a wiki to work its users

need the confidence that they take on responsibility for parts of the content without intervention of superiors. In hierarchical contexts often information is power. When this is the case, participators are usually unwilling to give away information, for example by entering it in a wiki, and thus forfeit power. A related problem frequently found in strict organizations is the fear that due credit will not be attributed for contribution to a wiki. The users worry that they will not receive credit for their good ideas and hard work and thus refuse to participate. This problem can be greatly alleviated by using the change monitoring mechanisms described in section 1.2 in combination with an administrator who makes sure that deserved credit is given for good work.

In open source projects there is most often a flat hierarchy of contributors led by a so called benevolent dictator who makes sure that there is a common vision and who steers the project towards it. This is the setting in which wikis work best. The content of successful wikis is monitored by the community, not by a single person. In the case of irresolvable problems the issue is escalated to a higher authority. This approach works well for two reasons. On the one hand, the information stored in wikis is often too overwhelming for a single person or a small team to monitor in sufficient detail so that they can identify problems. On the other hand, bearing responsibility and custodianship for the content of an article is an additional authoring incentive.

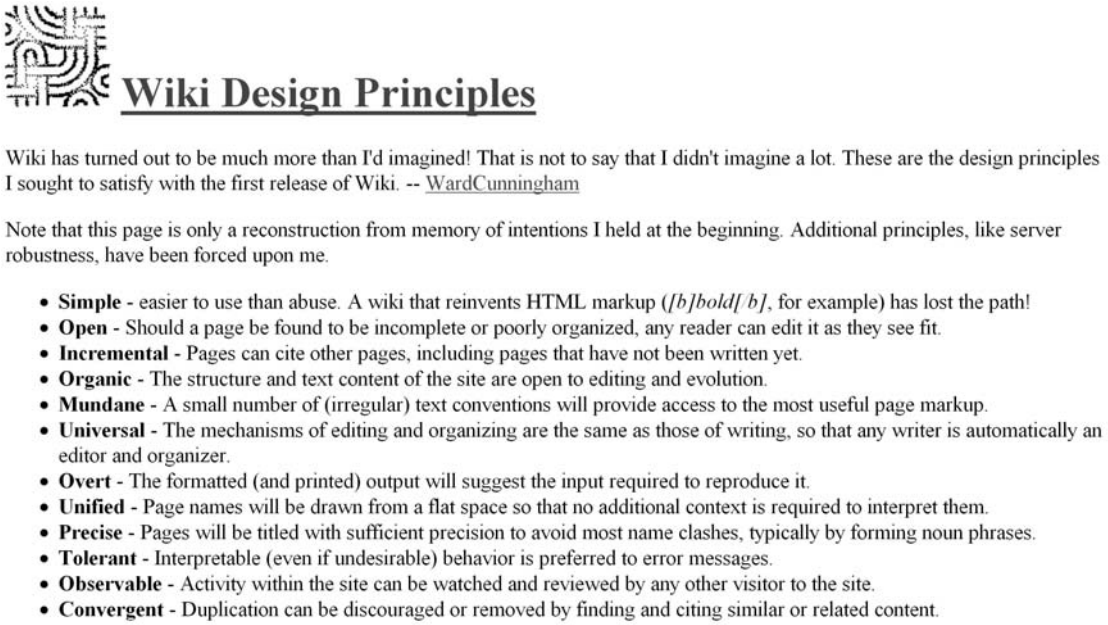
The smooth working of such a culture is best supplemented with conduct guidelines. These guidelines should state what kind of information should be entered in the wiki, it should specify conflict resolution procedures, it should settle a guideline for copyright issues and so on. A good example for successful guidelines are the five pillars of Wikipedia (see Wikipedia, 2009a). Often the presence of a clear policy in combination with technical conflict resolution support will prevent conflicts from appearing in the first place. Anecdotal studies (see Wikipedia, 2009b, and the references therein) have shown that when relying on community monitoring vandalism has practically no negative impact on the integrity of wikis.

When introducing wikis into social contexts without prior wiki experience there are some common misunderstandings. Often expectations are too high, users need to be made aware that not everything will work from the beginning. Wikis are not a solution for every collaboration difficulty. Although wikis are simple, they need to be introduced. A tutorial or workshop in which technical as well as social aspects of wikis are explained ensures that they are used as intended. Usually it needs to be made explicitly clear that wikis are never finished and that they require constant commitment. Another way to reduce the entry barrier for new users is to have already some content present at the time the wiki is introduced.

4 Example Scenarios

To draft a picture of possible applications of wikis we present several fictional (and real) scenarios.

Figure 6: Wiki design principles



Wiki Design Principles

Wiki has turned out to be much more than I'd imagined! That is not to say that I didn't imagine a lot. These are the design principles I sought to satisfy with the first release of Wiki. -- WardCunningham

Note that this page is only a reconstruction from memory of intentions I held at the beginning. Additional principles, like server robustness, have been forced upon me.

- **Simple** - easier to use than abuse. A wiki that reinvents HTML markup (*[b]bold[/b]*, for example) has lost the path!
- **Open** - Should a page be found to be incomplete or poorly organized, any reader can edit it as they see fit.
- **Incremental** - Pages can cite other pages, including pages that have not been written yet.
- **Organic** - The structure and text content of the site are open to editing and evolution.
- **Mundane** - A small number of (irregular) text conventions will provide access to the most useful page markup.
- **Universal** - The mechanisms of editing and organizing are the same as those of writing, so that any writer is automatically an editor and organizer.
- **Overt** - The formatted (and printed) output will suggest the input required to reproduce it.
- **Unified** - Page names will be drawn from a flat space so that no additional context is required to interpret them.
- **Precise** - Pages will be titled with sufficient precision to avoid most name clashes, typically by forming noun phrases.
- **Tolerant** - Interpretable (even if undesirable) behavior is preferred to error messages.
- **Observable** - Activity within the site can be watched and reviewed by any other visitor to the site.
- **Convergent** - Duplication can be discouraged or removed by finding and citing similar or related content.

4.1 Building collaboratively a knowledge base (in a given domain)

This is maybe the most common usage of a wiki with Wikipedia being the most famous example. The content is created and used by the community and the size of this community can vary from a small team to potentially the whole world in Wikipedia. Several examples of particular knowledge base wikis are in order.

4.1.1 METIS Wiki

URL: `http://www1.unece.org/stat/platform/display/metis/About+this+wiki`.
The METIS wiki is part of the effort to help experts in statistical agencies develop metadata management systems and processes. The meetings of metadata experts from statistical offices are held every 1-2 years, with the purpose of sharing experiences and developing a Common Metadata Framework (CMF). Published on-line, the framework is available at `http://www.unece.org/stats/cmf`. Each part of the CMF concentrates on different practical and theoretical aspects of statistical metadata systems. Part D focuses on the experiences of national statistical offices that have recently implemented or re-engineered their statistical metainformation systems. Statistical organizations are describing their approach to metadata management in a series of case studies which are published on the METIS-wiki, a platform where contributors can keep their own material up-to-date as their metadata projects progress. The wiki engine behind is the enterprise wiki *Confluence* - see `http://www.unece.org/stats/cmf` hosted by UNECE. The content is structured around "case study" entries at several levels and only authorized users (i.e. the contributors of the respective case studies) can add or edit content. The case studies follow a predefined template (available also as a MS Word template) containing general information about the organization, six topics (which could slightly vary from organization to

organization) and the complete case study as a PDF document. If the case study was created as a MS Word document it can be easily uploaded to the wiki using the import capability provided by the engine. The front page of the wiki contains information about the common metadata framework, about the wiki and the available editing tools as well as further contact information.

4.1.2 R Wiki

URL: `http://wiki.r-project.org/rwiki/doku.php`
As described by the R-core development team on its web page, R (R Development Core Team, 2009) is "a system for statistical computation and graphics. It provides, among other things, a programming language, high-level graphics, interfaces to other languages and debugging facilities". R is a free software and is developed and distributed under the GNU license and enjoys a very large community of users and developers. It provides a wide variety of statistical and graphical techniques - linear and non-linear modeling, classical statistical tests, time-series analysis, classification, clustering, robust methods, etc. Hundreds of specialized statistical procedures for a variety of applications are available from the *Comprehensive R Archive Network (CRAN)* in the form of contributed R packages, which can be downloaded in source form or installed directly from the R console. A few examples relevant for national and international statistical organizations are: survey analysis (**survey**, **pps**, **sampling**, **sampfling**), handling of missing data (**VIM, mice, mi, mvnmle, mitools, EMV, mix, pan**), time series analysis, robust statistics (**robustbase, rrcov, robust**). More information on R can be found at the CRAN web site (see also Todorov, 2008).

R comes with several official manuals and a general collection of useful information for users on all platforms (Linux, Mac, Unix, Windows) can be found in R FAQ. Additionally there are two platform-specific FAQs for Windows and MacOS. The main location for general questions about R is the R-Help mailing list. It is useful (if you dare ask a question) but very busy and quite unfriendly - it happens that the same topics are discussed repeatedly and the long-term members always complain that the newbies never read the archives and the documentation. May be this was the reason to launch an R Wiki as an easier platform to search, find and contribute information all around R. There existed other R Wikis before, but they did not gain much success due mainly to the lack of publicity and the use of a too simplistic wiki engine which did not provide R-specific features like code-highlighting, direct links to documentation for R functions and packages, etc. The new R Wiki launched in 2006 (see Grosjean, 2006), uses the DokuWiki engine, targeted to software documentation and extended with R-specific plugins to make it more suitable for creating R documentation. The main principle for structuring the content is to distinguish between large guides or books and short tips. It seems that the first category is not quite relevant since its content did not go further than importing a small part of the book *Statistics with R* by Vincent Zoonekynd and several short tutorials.

Currently there are about 2800 pages in the database which includes also the wiki version of the R documentation. Probably the number of *h2*-titles is more representative of real wiki pages—there are 460 pages that are legitimate content pages. There are 650 registered users.

4.1.3 Ubuntu Wiki

URL: `http://wiki.ubuntuusers.de/Startseite`
Ubuntu is a community-developed computer operating system based on the *Debian* Linux distribution and is available as free and open source software under the GNU General Public License (GNU GPL) and the GNU Lesser General Public License (GNU LGPL). The development is sponsored by the UK based company Canonical Ltd., owned by the South African Mark Shuttleworth. *Ubuntu* provides an up-to-date, stable operating system for the average user, with a strong focus on usability and ease of installation.

Ubuntuusers is a portal (in German) providing everything necessary about *Ubuntu* and its derivatives. It is the single entry point to a web forum where one can ask questions, a wiki where one can read (or write) guides and explanations and a blog which publishes news from *Ububtuusers*. The wiki has a clean structure built on a small number of top level categories like download, installation, drivers, security, programming and so on, which can be further expanded.

4.2 Collaborative software development

This scenario entails a small team of software developers working on a mid-range project. The team consists of five people, one of which is the senior developer and another is responsible for providing the team infrastructure. The team uses a wiki in their development process. The content of the wiki will not be a part of the product developed by the team, it is rather part of the supporting infrastructure, similarly to a version control system. It is used to document peculiarities of the development cycle, to share know-how about third party software, to collect information about how problems are tackled by competing products or perhaps to store meeting minutes and to-do lists. The wiki is probably an inappropriate medium for a technical documentation of the source code. For such a task a wiki is far too informal. One of the advantages of using a wiki in this scenario is that common tasks, which are not performed frequently enough for the team members to remember the exact steps, can be documented. In these cases the developers will have a guideline for solving a specific problem and will not have to go through the same documentation over and over. An example for such a task is the build process. Suppose that the developers are employing unit test techniques. Occasionally it is necessary to build the project project despite some of the tests failing. The wiki could store the command line options required to ignore failed tests and produce a project build. The wiki corpus would also be advantageous to developers newly introduced into the project. On the one hand, new developers can be referred to the wiki on trivial issues. On the other hand, these new developers will probably discover several minor mistakes in the wiki and will thus contribute to a better corpus.

Many additional challenges poses the development of a pilot project in an area where neither clear requirements are articulated nor sufficient knowledge is available. Investigating, learning, capacity building, experimenting, programming and documenting should go hand in hand and the collected information should be reliably stored in the fastest and easiest possible way. A new dimension of complexity can be added to the organization and coordination of the project if the development team is geographically spread around the world. As an illustrative example

of this type the pilot *SDMX project* which is currently carried out by UNIDO will be described below.

The SDMX initiative is an international project carried out by several international organizations, including the United Nations, specialized agencies, OECD and Eurostat. SDMX has been endorsed by the UN Statistics Commission as the preferred method to be used by the international statistical system. SDMX aims at defining standard formats, information technology architecture and content-oriented guidelines for the national and international exchange of statistical data and metadata. The SDMX Technical Standards Version 2.0 developed and reviewed with the goal to replace, within the context of the International Organization for Standardization (ISO) the previous version (ISO/TS 17369:2005 SDMX) provides technical specifications for the exchange of data and metadata based on a common information model. Its scope is to define formats for the exchange of aggregated statistical data and the metadata needed to understand how the data are structured. The main focus is on data presented as time series, although cross-sectional XML formats are also supported. Version 2.0 Technical Standards are compatible with the earlier Version 1.0 efforts, which focused on XML- and EDIFACT-syntax data formats. The latest work broadens the technical framework to support wider coverage of metadata exchange as well as a more fully articulated architecture for data and metadata exchange.

As it takes years for large number of national and international agencies to adopt common standards, SDMX started with several agencies that have the required technical facilities in place and are familiar with sharing data. The practical utilization of SDMX standards is still in its infancy, not only in UNIDO but also in most international organizations. Some prominent pilot projects (not a complete list), from which lessons can be learned include:

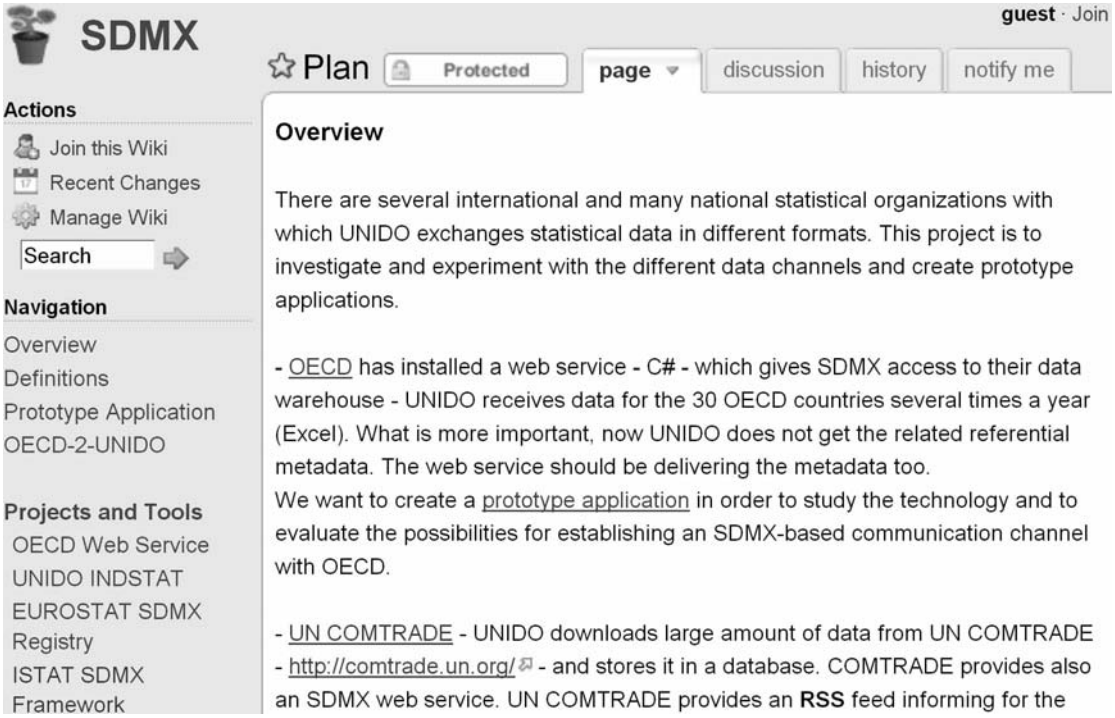
1. *SDMX Open Data Interchange (SODI)* which is a data-sharing and exchange project within the European Statistical System. The project started with a pilot exercise involving National Statistical Institutes of France, Germany, the Netherlands, Sweden and the United Kingdom. The statistical institutes of Denmark, Italy, Norway and Slovenia joined the pilot exercise in 2006, while Finland and Ireland joined in 2007;

2. *FAO CountrySTAT*, which is based on the application of data and metadata standards of FAOSTAT and SDMX, is a web-based system being developed since May 2004 using PX-Web at FAO Headquarters. It was successfully tested in the statistical offices of Kenya, Kyrgyzstan and Ghana during 2005. Many other developing and developed countries have shown an interest in and are adopting it (see `http://www.fao.org/es/ess/countrystat/`);

3. Data exchange between OECD and IMF: Exchange Rates data from IFS.

There are many ways to use SDMX to exchange data. For example, a primary distinction can be made on whether the data are being sent by one counter party to another (called a "push" scenario) or whether the data are posted in an accessible location, and then obtained when needed (called a "pull" scenario). In the push mode, which is the traditional data-sharing mode, different means, such as e-mails and file transfers, are used to exchange data. This is how UNIDO

Figure 7: One of the main pages of the UNIDO SDMX wiki.



and many international agencies collect data from NSOs and international organizations. To use SDMX for data-reporting or data collection, which are actually two aspects of the task of data exchange, at least two counter parties are required, one or more providing data to another. Any of these counter parties must adopt the same technical standards, have a common data structure and use common vocabulary. The lack of necessary technical facilities could be a serious stumbling block for developing countries involved in the process.

As a first step towards SDMX utilization, UNIDO is currently developing a data and metadata exchange procedure based on the web service provided at *OECD.Stat*. This is the central repository where validated statistical data and metadata are stored, and is intended in due course to become the sole coherent source of statistical data and related metadata for the OECD statistical publications. Utilizing the OECD web service will allow to automatically retrieve and process data for all OECD member countries, which currently is done by transferring Excel files (see Upadhyaya and Todorov, 2008). Being a pilot project with the goal of demonstrating the applicability of SDMX in the UNIDO statistical production process, it was organized on a volunteer basis with almost no budget. The team consists of several members at distant geographical locations - one in Karachi, Pakistan, two in Vienna, several in Paris and possibly later some in Washington. The profile of the team members was different and the first task was to build an SDMX knowledge to be used throughout the project and at the same time to specify the basic requirements and the desired output. To start with, a wiki was set up as a central point for storing the project materials and knowledge. The fastest, easiest and cheapest option was to use one of the available free web farms and the choice fell quite soon on *Wikispaces*:

www.wikispaces.com/. Figure 7 shows one of the main pages of the UNIDO SDMX wiki. The structure of the wiki is extremely simple and at the highest level consists of three main parts:

- OECD web service: specification, pointers and other information about the OECD web service
- UNIDO INDSTAT: information about the required data and the structure of the UNIDO INDSTAT database
- General SDMX information, pointers to tools, useful documents and links, etc.

Figure 8: Example of a wiki page including code (in this case XML).

```
<?xml version="1.0" encoding="UTF-8"?>
<message:QueryMessage xmlns="http://www.SDMX.org/resources/SDMXML/schema
   http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message http://www.
   <message:Header>
      <message:ID>none</message:ID>
      <message:Test>false</message:Test>
      <message:Prepared>2007-04-25T00:00:00-01:00</message:Prepared>
      <message:Sender id="OECD"/>
      <message:Receiver id="OECD"/>
   </message:Header>
   <message:Query>
      <KeyFamilyWhere>
      <Or>
         <KeyFamily>SSIS_BSC</KeyFamily>
      </Or>
      </KeyFamilyWhere>
   </message:Query>
</message:QueryMessage>
```

Since one of the main goals of this pilot project is to investigate the functionality of the OECD Web services and to create modular and well tested building blocks for establishing a production SDMX-based communication channel to the UNIDO statistical database the most interesting page is the one with pointers to the functions to be developed, like authentication, retrieval of Data Structure Definition (DSD), retrieval of generic data, etc. The wiki engine allows to store code examples (XML) in a readable form as shown in Figure 8.

4.3 Wiki for education, teaching and training

Modern education theory states that meaningful learning cannot be accomplished solely by performing passive activities like reading or listening. Rather students should be engaged in the learning process. Following this logic a teacher is not one who provides content but one who provides context for learning. If passive learning were effective, students would read a book on

a topic and become experts. Since this is not the case, the job of a teacher or instructor is to provide a context in which effective learning can take place.

Wikis can serve as a tool which inherently requires students to actively contribute to the learning process. When applying wikis in an educational context, similar problems as well as best practices apply like in other scenarios—see Section 3.3, one major difference being that in educational settings the contributors to a wiki meet frequently in person. Resolving issues in a classroom setting is much easier than when the contributors to a wiki are scattered around the world.

Several applications have been identified for wikis in education (West and West, 2009). They can be used for knowledge construction. Students will compile an information corpus on a subject, field or discipline in order to gain a deeper understanding of the matter by being forced to write down their thoughts. Another application of wikis is to hone critical thinking. Although critical thinking is a skill mainly promoted in academic settings, its goals, namely the analysis and evaluation of a subject, are also applicable outside universities. A third application of wikis in a classroom is to provide contextual application. Contextual application is the practical application of newly acquired knowledge with the purpose of reinforcing the grasp on the material. In this case wikis can be used to coordinate and organize the implementation of a practical project.

4.4 Collaborative Authoring

In this scenario we consider a small team, say of up to 10 members (possibly distributed geographically and organizationally) with the task of creating one or more documents (known as collaborative writing or collaborative authoring). Editing of the document can be done either in real-time or asynchronously. There exist software tools and technologies that facilitate the editing and reviewing of a text document by multiple individuals which can vary a great deal and can range from the simplicity of a wiki system to more advanced systems. The collaborative writing tools are characterized by several key features: supported file formats, text chat or conferencing, tracking changes and support of revisions, support for RSS feeds, email updates, support of private and public sessions, real time coediting, possibility to add comments, spell checker. A useful guide to online collaborative writing tools based on these criteria can be found at `http://www.kolabora.com/news/2007/03/01/collaborative_writing_tools_and_technology.htm`.

We illustrate this scenario with a recent, very successful example of application of wiki by a cross organizational working group - the MSIS (Management of Statistical Information Systems)[1] Task Force on Software Sharing. The task force was launched during the 2008 MSIS

[1]The UNECE, in partnership with Eurostat and the OECD, organizes annual meetings on the management of statistical information systems (MSIS). The aim is to: (i) provide a forum for exchange of experiences; (ii) collect, discuss and make available examples of good practice and (iii) facilitate implementation of relevant standards and recommendations across the UNECE region. MSIS meetings consider issues related to information technology governance and management, system architecture, accessibility and usability. These meetings are prepared by the MSIS Steering Group, which: (i) ensures continuity over time, and follow-up of actions between meetings; (ii) includes representatives of national and international statistical organizations and (iii) reports to the Conference of

meeting and was asked to prepare a report on possible future work on sharing statistical software and components. The working group had 10 members (representing five national statistical offices and five international organizations) and no face-to-face meeting was envisaged. The work started by conventional mail exchange, than for a short time a web forum was introduced and finally the group settled on a wiki generously hosted by ISTAT. Some initial structure was proposed and it turned out to be sufficient. The editing was performed both asynchronously (adding and editing the bulk of the text and writing comments) and in real time, during a teleconference (hosted by UNECE). The asynchronous mode demonstrates how wikis allow both discussion of content, i.e. comments, that do not directly change the base content and evolution of the content itself. The real-time mode is even more impressive when compared to alternative methods like email discussion or web forum. Of course this could be achieved through "heavy duty" on line collaboration tools, but wikis are free and provide the simplest means - no need to download specialized software, no problems with the organizational "borders" (no organization could host the group easily in its *SharePoint* or *LotusNotes* environment). The work of the task force lead to the proposal of a Sharing Advisory Board (SAB) with formal terms of reference which was endorsed by the May 2009 MSIS meeting. The mission of the SAB is to provide strategic direction to international work on the convergence of statistical business architectures and to promote favourable conditions for the sharing and joint development of software tools and components amongst national and international statistical organizations. The wiki of the task force was transformed into a new one hosted by UNECE and running on *Confluence*. Now the work of the Sharing Advisory Board continues successfully with regular teleconferences and a face-to-face meeting once a year at the annual MSIS meeting. A section of the wiki is opened to the general public where the SAB outputs are presented, a regular news letter is published and tools for gathering information from the readers as well as feedback are provide.d

4.5 Intranet Wikis

Most Intranets follow the model of *browsing already authored texts* and the content is created by a small number of employees assigned to this task and supported by the IT department. It is considered that the publication process is too complicated and has to be done by specialists. Recently several large companies like Google, Nokia, Motorola presented their internal web organization based around wiki (Leuf and Cunningham, 2001; Buffa, 2006). Within an Intranet, wikis are a good means for the quick and uncomplicated collecting of information. This creates a knowledge base as well as a platform for communication that is always available to the participating employees. At the TWiki web page `http://twiki.org/cgi-bin/view/Blog/2008-03-17-wiki-intranet` one can read the top ten reasons why there should be a wiki in every Intranet.

Here we will consider the example of the UNIDO Intranet. UNIDO launched a wiki as an Intranet platform in 2006 as a successor of the previous conventional HTML web site. The adopted engine is Mediawiki with many optional modules. The Intranet wiki is accessible in the

European Statisticians. For more information see the MSIS wiki hosted by UNECE at `http://www1.unece.org/stat/platform/display/msis/Home+Page`.

Headquarters in Vienna as well as in all field offices throughout the world (directly or indirectly).

Currently there are about 20000 pages in the database which includes "talk" pages, pages about UNIDO Intranet, minimal "stub" pages, redirects, and others that probably do not qualify as content pages. Excluding those, there are 8000 pages that are probably legitimate content pages. There are 650 registered users and about 200 of them actively contribute content. The rest are simply using the provided services (may be without even knowing that they are working with wiki software) like calendar of events, room reservations or library access. Recently a Semantic MediaWiki (SMW) extension was installed that helps to search, organize, tag, browse, evaluate, and share the wiki's content (see section 6.3).

4.6 A lobbying group

The next scenario describes lobbying group with a small core team and a number of freelances. In this setting a wiki can be used to sketch a common argumentation strategy of the group. It would be used to structure excerpts of publications of directly or indirectly involved parties in such a way that an optimal strategy can be devised. For example, the wiki of a group lobbying for cyclists in urban environments would contain references to road traffic regulations, excerpts of health advisories supportive of bicycling or studies demonstrating how bicycles are a lot more environmentally friendly than fossil fueled vehicles. While the wiki is probably too unstructured for a contact list of government representatives responsible for road traffic, it can be used to document an escalation plan along government instances in case of a specific issue. If the lobby publishes a newsletter, the wiki can be used to store and collaboratively proof read articles.

Recently (April 2006), a letter signed by twenty-three British academics, expressing their concerns over the current progress and direction of NHS Connecting for Health's National Programme for Information Technology (NPfIT) was prepared and sent to the Health Select Committee. To work together and keep a dossier of information relating to this topic they set up an own wiki in which media reports were tracked and was made available to a general readership (`http://editthis.info/nhs_it_info/Main_Page`). For this purpose they used the free service *EditThis.info* (`http://editthis.info/`). The maintenance of the wiki continues up to now (last update as of 15 December 2009) and contains about 1400 pages of which 118 are content pages and five uploaded PDF files.

5 Wiki engines

Without going into details—comparing and choosing a wiki or a wiki engine is a topic in itself—we will briefly present several of the most well-known wiki engines. The majority of engines are open source/free software, often available under the GNU General Public License (GPL). It is hard to determine which wiki engines are the most popular, although a list of lead candidates include TWiki, MoinMoin, PmWiki, XWiki, DokuWiki and MediaWiki. Some engines include many non-wiki features (news articles, blogs, etc.) like those in TikiWiki CMS/Groupware

and can be considered a wiki-Content Management System hybrid. An excellent resource for choosing a wiki is the site Wikimatrix: `http://www.wikimatrix.org/`.

5.1 Mediawiki

MediaWiki is the most popular wiki software powering Wikipedia and other projects of the non-profit WikiMedia Foundation, as well as many other wikis. MediaWiki supports many languages, web site user styles, multimedia and extension features, index of content items, edit tracking, talk pages and a lot more. MediaWiki is suitable for personal and education use. MediaWiki is a free software package licensed under the GNU General Public License (GPL).

5.2 TikiWiki

TikiWiki—`http://tikiwiki.org/` is a powerful open-source "groupware" tool as well as a Content Management System. TikiWiki can be used to create web sites on the internet and intranet. It offers great resources as a collaboration tool. One can use TikiWiki for forums, chat rooms, poll taking, blog, file and image gallery, FAQ, calendar, and even more.

5.3 DokuWiki

DokuWiki—`http://www.dokuwiki.org/dokuwiki` is an easy to use and standards-compliant wiki system. DokuWiki is the best choice to write a small or medium size documentation. It eases the creation of structured content, has a powerful syntax, and data files can be read outside the wiki. DokuWiki helps teams and workgroups interact much easier while working on a project. All data are stored in plain text files and no database is required. DokuWiki as well as MediaWiki and TikiWiki are written in PHP.

6 Information Retrieval in Wikis

One common critique of wikis is that it is hard to find information when you are unfamiliar with the specific wiki. Since the content structure is loose, there is often no other feasible way than a full text search to find a specific article in a wiki. There are several strategies to deal with the unstructured information graph. Three of the approaches are discussed in this section: (i) structuring the wiki itself, (ii) using a "structured wiki" engine and (iii) using a semantic wiki.

6.1 Structuring the Wiki

The wiki is by definition an informal database with free-form entries and no specific structure imposed. Nevertheless often useful structures are created (evolving over time) by the users. The structure can be also viewed from an administrative point of view—a particular structure can be created, suggested and in some cases enforced. There are different ways to structure the content of a wiki and all are supported by the underlying hyperlink mechanism. In order to support both contributors and users of a wiki, most wiki implementations, like MediaWiki, TWiki, or PmWiki, offer structuring elements such as namespaces, subpages, categories, different types of links (internal and external), keyword or full-text search, templates, or skins.

To impose or not a structure on a wiki depends very much on its purpose and content. Wikis used for education and training need much more structure than general discussion wikis. A focused team consisting of people used to working with wiki usually needs only a very rudimentary structure. One way to "improve" the structure of a wiki is to periodically set up or renew the main topic content (Leuf and Cunningham, 2001) by selecting "core pages" including

- A page explaining the topic and its scope which will be the top page
- Pages setting the major entry points (first-level pages)
- A page explaining the "Find" capabilities
- Pages explaining how to use the editing capabilities

Much more insight in the structuring issues of a wiki can be gained by investigating particular case studies (see Leuf and Cunningham, 2001, page 363).

6.2 Structured Wikis

Structured wikis combine the benefits of plain wikis and database systems, although these two seem to belong to contradicting worlds. The result of such a combination is a collaborative database environment where knowledge can be shared freely, and where structure can be added as needed. In a structured wiki, users can create applications that are very specific to their needs, such as *call center status boards*, *to-do lists*, *inventory systems*, *bug trackers*, *calendar of events* and more. A wiki can become a structured wiki thanks to a combination of a number of features such as templating system, formatted search. There are several wiki engines which support structured wikis—*TWiki*, *TikiWiki*, *PmWiki* and others.

6.3 Semantic Wikis

Semantic wikis follow the idea of the semantic web as envisioned by Tim Berners-Lee (Berners-Lee *et al.*, 2001). The main idea behind the semantic web is to enrich information on the World Wide Web with machine processable information. When information is stored in a machine processable format, the retrieval, evaluation and combination of search results can be automated. A common example is one of a person looking for a doctor's appointment. In today's web one would search for doctors in the vicinity, look through their opening times and match these against their existing appointments—all by hand. Suppose that the address of the doctors' offices and their opening times were not only presented in a human readable format on their web sites but are also contained in some metadata which is hidden from human users. An agent software could then sift through the metadata and automatically combine the opening times with the calendar of the human user. Effectively a semantic web would not be restricted to full text searches, as is the case even with today's most sophisticated search engines, but would be able to produce meaningful results based on the meaning of information.

Adding semantic metadata to wikis relies on the fact that hyperlinks are the main method for structuring information in a wiki. When analyzing the relationship between articles and links one can only tell how much links are leading into or out of an article. Such an analysis would give a measure of the importance of an article. In fact, this measure is called the page rank and is used by Google.com to index the web. The importance of an article, however only provides a useful way to sort matches to full text queries. The deficiency here is that although it is possible

to enumerate links and articles and produce a graph of the information in a wiki, there is no feasible method to tell a computer the meaning of a link or what kind of articles it is connecting. To alleviate this problem such information is simply added to the wiki content. Each hyperlink is treated as a subject, predicate, object triple. Let us look at a concrete example. Suppose one is looking for all movies starring Great Actor in a wiki containing movie information. In such a wiki, articles on movies would belong to the class *movies*, actors would belong to the class *actors*. Links would serve as predicates so there would be a predicate called *playsIn*. If articles and links are enriched with these respective data, it is easy to ask a machine for all movies which have an incoming link of the type *playsIn* from a specific actor.

For further information on semantic wikis the interested reader is referred to the seminal paper on semantic wikis Krötzsch *et al.* (2005). Semantic wiki implementations are currently mainly a research topic and there are so far none known to the authors which work well with large amounts of information. An example implementation is the MediaWiki extension SMW (Semantic MediaWiki) which allows users to add "semantic annotations" to the wiki: `http://semantic-mediawiki.org/wiki/Semantic_MediaWiki`.

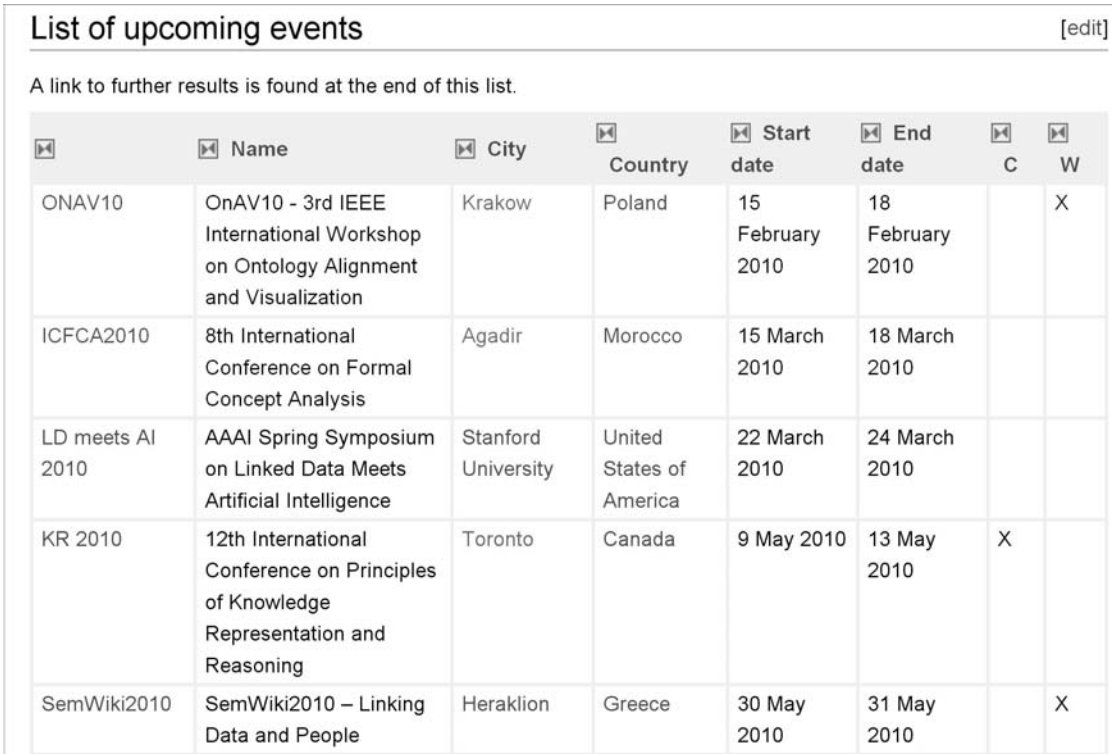
An example of a semantic wiki, powered by the SMW extension can be experienced at `http://semanticweb.org/wiki/Events` where a list of upcoming events, mostly conferences and workshops related to semantic technologies is presented. The events are described by pages, which should make use of an *event template* and can be entered by anybody who is logged in. The list of events (containing title, city, country, start and end date), sorted by end date is shown in figure 9 and the query used to create this list looks as follows:

```
== List of upcoming events ==
A link to further results is found at the end of this list.
{{#ask:[[Category:Event]]
[[end date::>{{CURRENTYEAR}}-
             {{CURRENTMONTH}}-
             {{CURRENTDAY}}]] |
?title = Name|
?has location city = City|
?has location country = Country|
?Start date|
?End date|
?Category:Conference = C|
?Category:Workshop = W|
format=table|limit=50|sort=end date
}}
```

7 Collaborative Reference Management and Social Bookmarking

In this section only a brief introduction to social bookmarking and collaborative reference management will be presented and the main material will be deferred to another document. Storing links to web pages and other internet resources (favorites or bookmarks) in the web browser means that they are only accessible on the local desktop PC and cannot be easily shared. Social bookmarking on the other hand offers many advantages as a personal and collaborative infor-

Figure 9: An example of a semantic wiki: List of upcoming events (conferences and workshops related to Semantic Web technology)



List of upcoming events [edit]

A link to further results is found at the end of this list.

	Name	City	Country	Start date	End date	C	W
ONAV10	OnAV10 - 3rd IEEE International Workshop on Ontology Alignment and Visualization	Krakow	Poland	15 February 2010	18 February 2010		X
ICFCA2010	8th International Conference on Formal Concept Analysis	Agadir	Morocco	15 March 2010	18 March 2010		
LD meets AI 2010	AAAI Spring Symposium on Linked Data Meets Artificial Intelligence	Stanford University	United States of America	22 March 2010	24 March 2010		
KR 2010	12th International Conference on Principles of Knowledge Representation and Reasoning	Toronto	Canada	9 May 2010	13 May 2010	X	
SemWiki2010	SemWiki2010 – Linking Data and People	Heraklion	Greece	30 May 2010	31 May 2010		X

mation management tool. The concept of social bookmarking is extremely simple, but possibly one of the most effective social software tools available. The online social bookmarking services allow one to create an online account and store bookmarks on the server. Logging into the account from any networked computer is enough to access the stored bookmarks. The interface provided by most of the services is easy and comfortable, providing quick buttons that can be installed in the browser and allow to bookmark a page while browsing the web. Bookmarking web sites of interest is closely related with collecting and maintaining references while researching on a particular topic. This topic, known as reference management will be considered in the next Section 7.1.

7.1 Reference Management

A crucial part of any research is the literature investigation and the citation. The maintenance of the bibliographical references, their collecting and formatting has always been the sole responsibility of the authors. The number of resources which could be searched electronically has increased tremendously in the recent decade and thus have posed a new dimension of complexity to the routine research work. Many reference management software tools have been developed to facilitate the task of collecting, organizing, maintaining and using the references. These tools can interact with the text processing systems (Microsoft Word, LaTeX, etc.) to easily insert these references into the text and produce list of references relevant to each particular publication. The key features of such a tool are:

- Maintain a database in which full bibliographic references can be entered
- Integrate with word processors so that citation can be included in the text while writing
- Automatically produce a reference list in the appropriate format which will reduce the risk that a cited source is not included in the reference list. Ideally the style of the reference list and the citations can be easily modified according to the requirements of the particular publication or journal
- Import references from other sources - from simple text files to proprietary formats of other reference management tools
- Enable the user to search references from online libraries, e.g. using the `Z39.50` public protocol

Collaborative research and collaborative writing of documents require strong synchronization among authors. This is valid especially for the references. Reference management tools can be divided roughly into two categories - local or personal manager programs which are to be installed on the local computer (*Reference Manager, EndNote, Refworks, Visual Composer, synapsen, Librixx, Bibliographix, JabRef*) or social bookmarking services which are available on the Internet (*CiteULike, Connotea, Bibsonomy, Del.icio.us*). While the former are mainly commercial products for which license fees have to be paid, the latter are freely available. A special case is *Zotero* which is an add-on for the *Firefox* browser (locally installed but free, open source). A very detailed comparison of almost all available reference management software tools can be found at `http://en.wikipedia.org/wiki/Comparison_of_reference_management_software`. It considers different aspects of the tools like operating system support, import and export file formats, supported citations styles, reference list file formats, word processor integration, database connectivity, networking, security and password protection. The information is presented in tabular form and no comments or recommendations are provided. Another valuable resource reference management software and social bookmarks is the master thesis Kerschis (2007) which describes in detail and compares most of the important tools but unfortunately is already somewhat outdated. Munushree (2008) considers the reference management from a computer science point of view and evaluates 87 (most of them open source) tools. The main criteria of the evaluation are similar to those listed at the beginning of the present section, but also other aspects, like organizing of ideas and references (mind mapping), conversion between different formats, duplicate discovery, are reflected. Special attention is devoted to the four tools considered to be the best in most of the evaluation criteria: *Aigaion*, *Bibsonomy*, *Zotero* and *Jabref*.

The most prominent commercial tools are *Endnote* and *Reference Manager* both by *Thomson Reuters*, the latter being also the oldest tool of this type (originally developed by Ernest and Earl Beutler, in 1982 for the CP/M operating system, ported to DOS and then Microsoft Windows and later the Apple Macintosh). *Reference Manager* is preferred for its multiuser and collaboration capabilities maintaining a shared central database of references. Although the current version of *EndNote* has networking capabilities and its files can reside on a central server it does

not, however, support collaborative editing of a single bibliographic file.

In the following sections several complementary tools will be briefly presented, which are considered to be applicable in the heterogeneities software environment of the UNIDO Research and Statistics Branch. This set of tools provides both desktop (*JabRef*) and web-based (*BibSonomy*) means for reference management, has a common internal format (BibTeX) and can integrate in LaTeX, *MS Word* and *Open Office* text processing systems. It requires no extra costs - neither for obtaining the tools which are open source nor for training of the users since the interface is easy-to-use and intuitive.

7.1.1 BibTeX and LaTeX

Bibliographies which are generated by LaTeXand BibTeX using a BibTeX file can be easily formatted to suit any journal reference list specifications through the use of different BibTeX style files. The development team of *JabRef*, Section refsection:jabref, supports the initiative to build a searchable database of BibTeX style files, organized by journal names - see `http://bst.maururu.net/`.

7.2 JabRef

JabRef Alver and Batada (2003) is an open source reference management desktop program that uses BibTeX as its native storage format. It is freely distributable under the terms of the GNU General Public License, version 2. *JabRef* provides an easy-to-use interface for editing BibTeX files, for importing data from online scientific databases, and for managing and searching BibTeX files. The application runs on the Java VM and is available for Windows, Linux and Mac OS X. In Figure 10 the main window of the program is shown. It displays the list of bibliographic entries with a default set of fields. The BibTeX fields are categorized into required, optional, general, abstract, review fields based on the entry type. The user may view a selected bibliographic entry in a formatted bibliography style (shown in Figure 11), which may be further modified.

A very useful feature of *JabRef* is the capability for duplicate discovery. Two entries are duplicates if they are exactly the same or if both have the same required fields. The two entries will be displayed and the user is given the option to keep either the first or the second entry or both of them. The entries that are not selected by the user are discarded automatically. Exact duplicate entries can optionally be deleted without further user interaction. *JabRef* supports plug-in interface and the most valuable (known to us) plug-in is the one from *BibSonomy* which allows to easily exchange entries between *JabRef* and *BibSonomy* (see Section 7.3). This approach combines the advantages of maintaining a local BibTeX-file with the comfort and usefulness of a collaborative reference management platform.

Figure 10: The main window of *JabRef*.

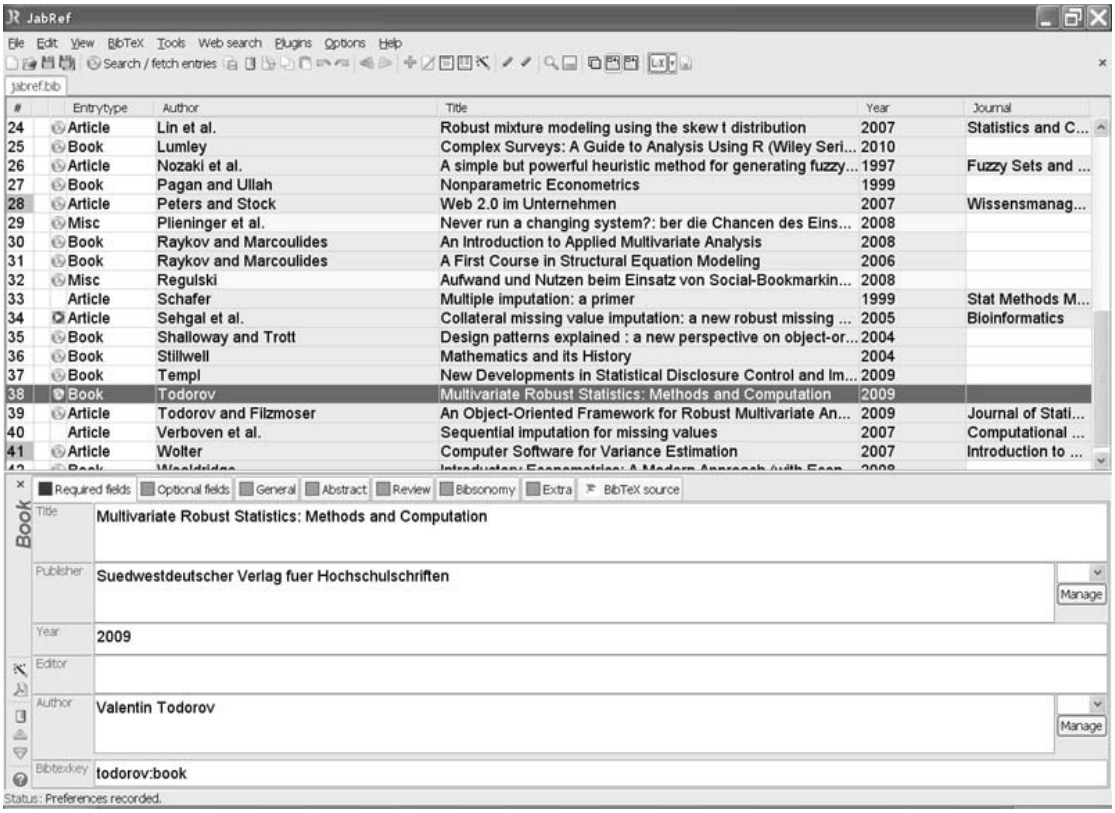


Figure 11: Bibliographic entry presented in a formatted style in *JabRef*.

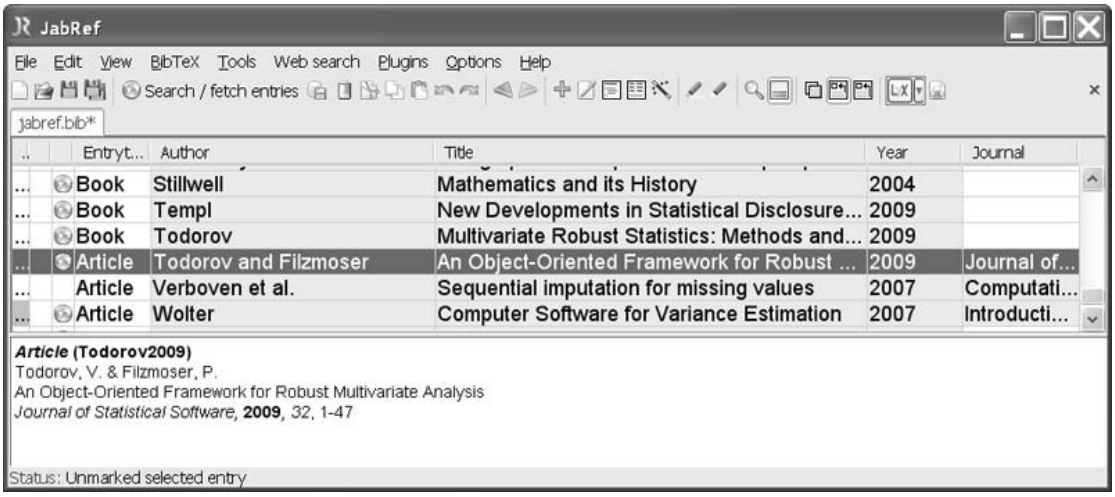
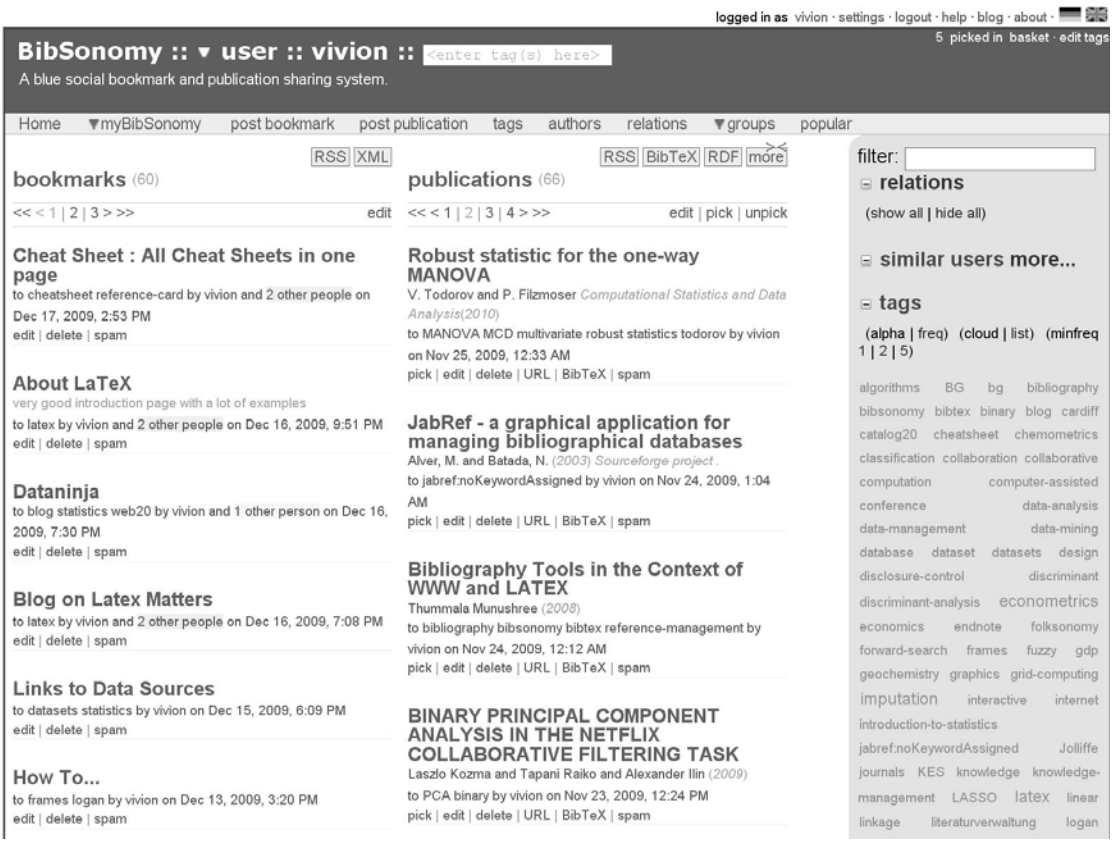


Figure 12: A typical page of BibSonomy displaying simultaneously bookmarks (left panel) and bibliographic references (right panel).



7.3 BibSonomy

BibSonomy (Hotho *et al.*, 2006) is a system for sharing bookmarks and lists of literature. When discovering a bookmark or a publication on the web, one can store it on the BibSonomy server. The different entries can be tagged (by adding keywords - tags) which helps to find again and organize a collection of bookmarks and publications. The idea is very similar to the bookmarks/favorites that are stored within a browser but the essential advantage of BibSonomy is that one can access her data from anywhere given an access to the Internet. It is also possible to discover bookmarks and publications stored by colleagues, friends members of special interest groups and other BibSonomy users. The data model of the publication part is based on BibTeX (Patashnik, 1988) which is a popular reference management system typically used with LaTeX (Lamport, 1994).

Similar to *BibSonomy* are the services *CiteULike* and *Connotea*. All these are online services which do not require installation of software on the own computer, so that the stored references and bookmarks are available from any place. This is not the case of desktop applications like, for instance, the semantic web based *Bibster* (Haase *et al.*, 2004) - `http://bibster.semanticweb.org/` and *Citavi* - `http://www.citavi.com/`.

8 Conclusions

In today's era of globalization the establishment of virtual teams, which are geographically separated and work across boundaries of organizations, space and time becomes more and more important. This would be impossible without appropriate computer driven communication technologies and here is where web 2.0 enters the scene. Web 2.0 is an evolutionary enhancement of the original web (or web 1.0) as an interactive collaboration platform. This is not a correction of previous shortcomings but rather leveraging the nowadays already available technological power to create revolutionary way of managing online information and knowledge. The operation, utilization and success of the web 2.0 tools worth studding from different perspectives.

In this paper the concept of wikis as collaboration tools is introduced and their strengths and weaknesses as compared to other collaboration tools are considered in a number of scenarios. Wikis have the advantage of being simple and inexpensive (if not completely free) which promotes widespread application with moderate resources. They have the ability to disseminate information across various domains unlimited in time, distance and organizations. A small group of people working intensively on related material is the ideal scenario for applying the wiki technology. The area of application could be documenting a product or a conceptual framework, a software development project, knowledge capturing, creation and sharing, education, teaching or training, "help desk", collaborative authoring and many others. A moderately sized company or organization could successfully utilize wikis for building Intranet but we do not know how well this architecture would scale.

A common critique of wikis is that the content structure is loose and it is hard to find information if one is unfamiliar with the specific wiki but there are strategies and technologies that could mitigate this problem such as "structured wiki" engines and semantic wiki. A potential hindrance for the adoption of wiki as a knowledge management platform is the possible mismatch with the organization culture.

Social bookmarking is a powerful Web 2.0 application and has tremendous potential to aid resource discovery, maintenance and sharing between team members. The closely related collaborative reference management is an essential tool in any research activity. An experimental platform for collaborative research based on existing free tools such as *JabRef*, *BibSonomy* and LaTeX is proposed but further study and evaluation of these applications is necessary.

References

Alver M. and Batada N. (2003) Jabref - a graphical application for managing bibliographical databases, sourceforge project.

Berners-Lee T. and Fischetti M. (1999) *Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*, Harper, San Francisco.

Berners-Lee T., Hendler J. and Lassila O. (2001) The semantic web, *Scientific American*.

Buffa M. (2006) Intranet wikis, in: *Proc of Intraweb workshop, WWW Conference 2006*.

Cress U. and Kimmerle J. (2008) A systemic and cognitive view on collaborative knowledge building with wikis, *International Journal of Computer-Supported Collaborative Learning*, 3, 2, 105–122.

Frongia D. and Vaccari C. (2008) Introduzione al web 2.0 per la statistica, *Contributi ISTAT*, 4, 2, 88–91.

Gardner J. (2009) Blogs, wikis and official statistics: New perspectives on the use of web 2.0 by statistical offices, UNECE Annual Report Economic Essays 11, UNECE.

Grosjean P. (2006) Collaborative writing of R documentation using a wiki, in: *Abstracts UseR'2006, Vienna, 2006*.

Grossenbacher A. (2009) Blog about stats, http://blogstats.wordpress.com/.

Haase P., Schnizler B., Broekstra J., Ehrig M., van Harmelen F., Menken M., Mika P., Plechawski M., Pyszlak P., Siebes R., Staab S. and Tempich C. (2004) Bibster–a semantics-based bibliographic peer-to-peer system, *Web Semantics: Science, Services and Agents on the World Wide Web*, 2, 1, 99–103.

Hotho A., Jschke R., Schmitz C. and Stumme G. (2006) Bibsonomy: A social bookmark and publication sharing system, in: *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, Kang M., ed., Aalborg University Press, 87–102.

Huettner B., Brown M.K. and James-Tanny C. (2007) *Managing Virtual Teams: Getting the Most from Wikis, Blogs, and Other Collaborative Tools*, Wordware Publishing.

Kerschis A. (2007) Literaturverwaltung und Wissensorganisation im Vergleich. Das Angebot von Literaturverwaltungsprogrammen und Social Bookmarking in Bezug auf die Benutzbarkeit in Bibliotheken, Master Thesis.

Krötzsch M., Vrandecic D. and Völkel M. (2005) Wikipedia and the semantic web - the missing links, in: *Proceedings of Wikimania 2005 - The First International Wikimedia Conference*, Wikimedia Foundation.

Lamport L. (1994) *LaTeX: A document preparation system: User's guide and reference manual*, Addison-Wesley Pub. Co., Reading, Mass.

Leuf B. and Cunningham W. (2001) *The Wiki Way*, Addison-Wesley, Boston.

Louridas P. (2006) Using wikis in software development, *Software, IEEE*, 23, 2, 88–91.

Munushree T. (2008) Bibliography tools in the context of WWW and LaTeX, Master Thesis.

Musser J. and O'Reilly T. (2007) *Web 2. 0 Principles and Best Practices*, O'Reilly Media, Inc.

O'Reilly T. (2005) What is web 2.0: Design patterns and business models for the next generation of software, http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

Patashnik O. (1988) *BibTeXing*.

Raymond E.S. (2001) *The cathedral and the Bazaar*, O'Reilly Media, Inc.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Todorov V. (2008) R: An open source statistical environment, meeting on the Management of Statistical Information Systems (MSIS 2008) (Luxembourg, 7-9 April 2008).

Upadhyaya S. and Todorov V. (2008) UNIDO data quality: A quality assurance framework for UNIDO statistical activities.

Wagner C. (2004) Wiki: A technology for conversational knowledge management and group collaboration, *Communications of the AIS*, 13, 256–289.

Wagner C. and Bolloju N. (2005) Supporting knowledge management in organizations with conversational technologies: Discussion forums, weblogs, and wikis, *Journal of Database Management*, 16, 2, 1–8.

West J.A. and West M.L. (2009) *Using Wikis for Online Collaboration*, John Wiley & Sons, Inc., San Francisco.

Wikipedia (2009a) Wikipedia:Five pillars — Wikipedia, The Free Encyclopedia, `http://en.wikipedia.org/w/index.php?title=Wikipedia:Five_pillars&oldid=332864270`, [Online; accessed 23-December-2009].

Wikipedia (2009b) Wikipedia:WikiProject Vandalism studies — Wikipedia, The Free Encyclopedia, `http://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Vandalism_studies&oldid=327745710`, [Online; accessed 22-December-2009].

Index

Printed in Austria
V.09-89454—January 2010—200

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Telephone: (+43-1) 26026-0, Fax: (+43-1) 26926-69
E-mail: unido@unido.org, Internet: www.unido.org