
	<p><b>OPAALS PROJECT</b></p> <p>Contract n° FP6-034824</p>
---	--

## **WP9: Communication and Dissemination**

### **D9.3 – Stable OPAALS Web Office**

	<p>Project funded by the European Community under the “Information Society Technology” Programme</p>
---	--

**Contract Number** : FP6-034824

**Project Acronym:** OPAALS

**Deliverable N°:** D9.3

**Due date:** Month 12, June 2007

**Delivery Date:** Month 12, June 2007

**Short Description:** A web accessible point for visitors interested in Opaals, generated thanks to the OKS data.

**Author:** Thibaud Desodt, TechIDEAS

**Partners contributed:** LSE

**Made available to:** OPAALS Consortium and European Community

#### Versioning

Version	Date	Name, organization
1	10/05/2007	Thibaud Desodt - Techideas

#### Quality check

**Internal Reviewers:** Neil Rathbone, LSE

**Dependences:**

<b>Work Packages</b>	Indicate which WP could benefit from this Deliverables and if possible with planned tasks, including a clear justification
<b>Partners</b>	LSE
<b>Domains</b>	web publication, data aggregation
<b>Targets</b>	specify which is the target of this deliverable (Domain researchers, public administrations, SMEs, Scientific communities on specific domains, etc.)



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License. To view a copy of this license, visit : <http://creativecommons.org/licenses/by-nc-sa/3.0/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

This article is available on the OPAALS Wiki (<http://wiki.opaals.org/WP9-D9.3>)

## WP9-D9.3 : Stable OPAALS Web-Office

*Note: it [has be proposed](#) that the "Web-Office" should be renamed to "[WebSite](#)". Therefore, in the following article, any reference to Web-Office should be understood as [WebSite](#)*

---

Author : [Thibaud Desodt](#) ([TechIDEAS](#))

The Opaals [WebSite](#) is part the [9th Workpackage](#) ("[Communication and Dissemination](#)") of the Opaals project. It should be the central point for research collaboration.

### Contents

1. [Presentation](#)
  1. [What is the Opaals Web-Site?](#)
  2. [Functionalities](#)
    1. [Centralization](#)
    2. [Visibility](#)
    3. [Auto-feeding](#)
    4. [Privacy and security](#)
  3. [Qualities](#)
    1. [Based on Open-Source software](#)
    2. [Customisable/Extensible](#)
2. [An aggregation system for the OKS](#)
  1. [Basic idea](#)
    1. [Jabber](#)
    2. [Wiki](#)
    3. [Blogs](#)
    4. [Visualization](#)
    5. [More...](#)
  2. [Overview](#)
3. [Implementation](#)
  1. [Architecture](#)
    1. [Data Providers](#)
    2. [Data Aggregator](#)
    3. [Access provider](#)
    4. [Renderer](#)
  2. [Concrete implementation](#)
    1. [Data sources](#)
    2. [Aggregation](#)
    3. [Access provider](#)
    4. [Naming conventions](#)
    5. [Rendering](#)
4. [Conclusion](#)
  1. [Interest of the solution](#)
  2. [Current state](#)
  3. [What's next?](#)
5. [See also](#)

## Presentation

This section consists in an analysis of its role among the OKS which will allow to focus on the expectations for such a tool.

### What is the Opaals Web-Site?

The workpackage 9 focuses on **spreading the excellence of the Opaals project**. It aims at increasing the community's reputation among Europe and beyond, to make its research **available to worldwide researchers and interested people**. This dissemination can be done using the usual tools (paper materials such as brochures, posters (...), branding...), which is the goal of Task 9.1. It allows communication and "advertisement" in a traditional static way. Nevertheless, the Open Knowledge Space (OKS)'s very dynamic architecture makes room for a other kind of dissemination : **dynamic dissemination**, as all the constantly-evolving information it contains can be used to provide an up-to-date vision of the project.

The Open Knowledge Space is a very important part of the Opaals project. One of its main characteristics is its "distributedness", which is, so far, rather limited, but will be in the future improved to reach a totally decentralized network of knowledge in the next phases. More and more, the information flows between participants will transit though Peer-2-Peer Channel, rather than the centralized web-based OKS tools. We will then need a "meeting-point", to access the distributed knowledge in an easy way, thus hiding the complexity to the end user.

While every participant in the project does his own research and/or development, on its own, or in collaboration with other members of Opaals, it has become indispensable to provide **a global view of its advancement and activity**. We need a place where a member of Opaals, or any visitor interested in the project may be able to get informed about it, without needing to browse the different tools of the OKS (wiki, blogs...) one by one. This is an important element of the Dissemination and Communication phase of the OKS.

To achieve this **centralized web-presence**, the task 9.2 proposes to implement the OPAALS Web-Site (finally renamed to Website ?), a web-published place where up-to-date Opaals-related information should be easily browsed through by visitors.

### Functionalities

As the centre of the OPAALS community, we expect a certain set of functionalities from the Web-Site.

#### Centralization

The first goal of the web-site is to provide a **central point** to stay informed about Opaals. This means that the system should be able to retrieve data from the different components of the OKS, and make it publicly available. The data may then be processed to generate user-readable data (HTML web-pages), or be used by other computer-tools for specific treatment (statistics, visualization...). Therefore, an essential part of the task is the **definition of naming conventions** that should allow any human or computer to know where he/it can find the information he is looking for.

#### Visibility

One of the objectives of the implementation of the web-site is that it should help broadcasting the OPAALS' activity. Therefore, efforts should be done to **improve its visibility on the internet**. This is

achieved mainly thanks to the indexation by search engines such as Google, which partly depends on the number of links pointing to a website. Visibility may also be achieved by using all the branding material developed for Opaals to customize the look of the web-site.

### Auto-feeding

Another expectation for the Web-Site is that it should rely on all the existing infrastructure and data of the OKS for the generation of web-accessible pages. It can then **publish up-to-date information automatically**, showing the dynamic and evolving nature of the OKS.

### Privacy and security

One of the major issues in the dissemination of data is the one of privacy. Only the information which has been set as public should actually be visible. Particularly in a field such as research, **circulation of confidential information** is a big issue. Therefore, it is important that the web-site should not open any security hole, and limit the publication to data that has been validated.

---

## Qualities

The previously listed functionalities constitute the core features that must be achieved by the Opaals Web-Site. To be really interesting for the future, it should have certain qualities, that will make it more and more useful. This is the topic of the following paragraphs.

### Based on Open-Source software

For the implementation of the web-site, we chose to use open-source components. The open-source and free software offers many assets that we may require in our project.

The very philosophy of free software integrates perfectly with the idea of the OKS. It assumes that knowledge should be free (as in "free beer"), and free (as in "freedom"). It then promotes the **sharing of ideas**, and the fact that anybody should be able to access them and add his own ideas and knowledge to the public domain. This philosophy focuses on the **notion of community**, which is what Opaals is all about. It is therefore normal that Opaals project should promote open-source software, using it in its developments.

As they are maintained by a community of developers, every bug is always quickly corrected, so that any released stable version of open-source projects is always as **good, efficient and well-designed** as a "closed-source" equivalent.

For all these reasons, we decided that the web-site should be based on open-source software components to provide an open, customisable, stable and extensible platform.

### Customisable/Extensible

Another expectation for the Web-Site is that it should be totally customisable and configurable. We should then be able to easily change its look to adapt it to the Opaals identity. It should also be easily integrated among the various tools and services of the OKS. Finally, improving its behaviour should be easy.

To achieve this qualities, the use of open-source components already provides big help. By their very nature, they are modifiable by developers, so that the behaviour can be perfectly tuned to our needs, or adapted to integrate better with other components.

The integration of the web-site may be facilitated by the use of standard and open formats for the circulation of data. These standards have been designed so that they could easily be processed by any computer-based application. This also means that adding other elements in the future would be easy.

In order to easily customise the look of the web-site, we should rely on the data/presentation model, so that the way information is presented to visitors can be changed without modifying the data layer. That means it would make it easy to separate the work of developers and the work of designers.

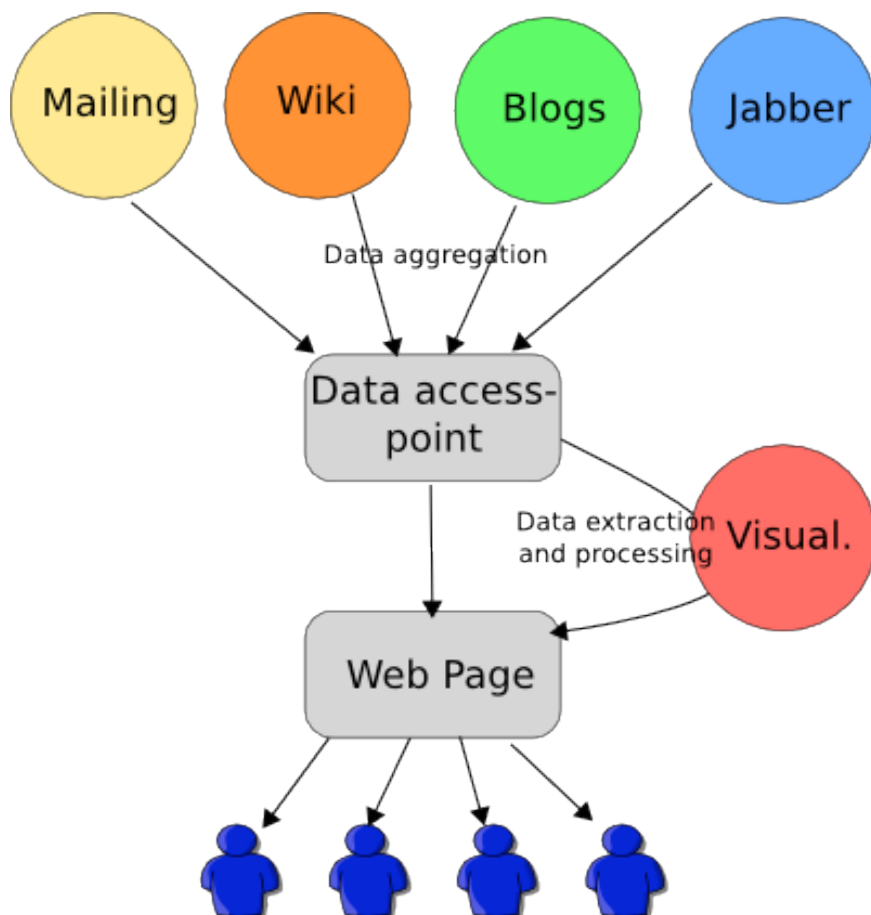
## An aggregation system for the OKS

Considering the requirements and expected functionalities, the Web-Site will be a kind of aggregator. This paragraph is partly based on Bob's presentation for the Web-Site in Kassel. It presents briefly the general idea behind the Web-Site.

### Basic idea

The basic idea is to provide content from the tools we are already using, such as Jabber, the Wiki, the Blogs in order to make it available to developers for processing (for instance visualization, statistics), and to generate the web-interface of the Web-Site.

The following drawing sums up this idea.



*Illustration 1: Overview of the data aggregation principle*

The different tools of Opaals provide very different kinds of data, which may then be brought together to provide an exhaustive Opaals information channel.

## **Jabber**

The instant messaging infrastructure **Jabber** mainly provides information about the members of the Opaals community such as their presence status and availability.

## **Wiki**

The **Opaals Wiki** is the main source of information regarding the general activity of the project. This is where one should search if he needs some details about a particular topic covered by Opaals.

It can provide detailed and up-to-date information about the project and its tasks as well as about the members of the Opaals community (institutions and individuals). Its user-friendliness and accessibility makes forebode that it will be the main data source. It can also generate statistics that may be used to monitor the whole project.

## **Blogs**

Every member of the project has his own **Opaals blog**, where he can post personal messages, ideas, comments or everything he may think about. For this reason, it is the ideal source for member-specific information, as well as the place to look for the latest news and events concerning Opaals. All the blogs may be used to generate a "community blog" with messages of everybody.

## **Visualization**

Some **visualisation tools** may also be used in the web-site context. The idea is that the visualization would both get data (about users, news...) and generate other data (images, charts...) that would then be integrated in the public interface.

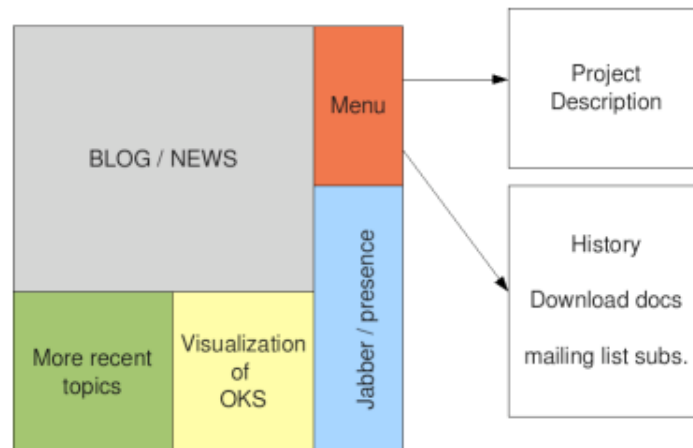
## **More...**

The aggregator shall be used with other sources, such as document repositories, or mailing list subscriptions... The more datasources we use, the more relevant the combination of data will be.

## **Overview**

Here is a very basic suggestion of what we could find in the public web-interface of the Opaals web-site. The look should of course be designed in collaboration with the team responsible for the Opaals branding, so that Opaals may have its identity on the web.





*Illustration 2: Layout prototype for data presentation*

*The Website aggregator makes it possible to have access to the big quantity of already existing data, with a minimum of effort. Its architecture has been designed to offer many points of extension, so that the system may be reused in other situations.*

## Implementation

In this section, we will speak a bit more about technical details such as system architectures, technological choices, and implementations.

The important features of this implementation are:

- the definition of naming conventions to access data
- the definition of a data format to retrieve data
- the adaptation of the different tools of the OKS in order to provide coherent data
- the choices that were done at implementation-time

## Architecture

The aggregation system counts with different parts:

- One or several **data sources** : providing requested data, in an exploitable format.
- A **data collector** (=aggregator) : it is responsible for retrieving data from the data sources and storing it in a centralized sink, and may also be responsible for filtering the obtained data (elimination of redundancy among the sources).
- An "access provider" : its role is to provide an access point that allows to retrieve data from all the sources, in a normalized way, and in a given format
- A **renderer** (facultative) : its role is to display the data in a "nice" way for the user.

## Data Providers

The idea is that each source should be able to provide data in a **standard format**, so that it can be automatically processed by the aggregator. For the best interoperability of all the components, it has

been decided that the data would be transmitted either in plain-text (for text data to present to users) or in XML (for data that should be processed by computers).

Data providers may propose data such as:

- **text-data** : for instance, content of articles
- **HTML** : for content of articles, keeping some layout informations
- **XML** : for structured data transmission. A particular kind of XML data is the **news feed XML format** (also known as RSS or Atom) which we study more in detail in the Opaals [NewsLetter](#) Deliverable.

These formats are official standards. Therefore, many tools exist to generate, read or process it. As they are text-based, they are also human-readable.

### Data Aggregator

This is mainly a process whose job is to "download" and update the data from the different sources, to **make it available** in a centralized point. It stores the data that the access-provider grants access to.

### Access provider

This is the public access point to the data of the OKS. Its main goal is to provide a web access to a the OKS' data, in the defined data formats.

It provides a **web-access to the data** stored by the aggregator, according to naming conventions that allows to retrieve a given piece of information from a given source.

### Renderer

This facultative part is the part responsible for converting the Opaals data into a **user-readable web page**. It may be integrated as a part of the Access Provider for simple processing, or as an external tool for more advanced transformation.

## Concrete implementation

Most of the current implementation is made using the object-oriented scripting language Python, which is also the language used by most of the tools of Opaals. It will make it easier for the developers of the community to modify it.

### Data sources

As stated before, the preferred data format we will extract from sources is text or XML. Some of the OKS tools have been modified in order to provide data in these formats. For informative purpose, here are some of our implementations or the existing functionalities we use. They should allow to generate an index, as well as retrieving news.

In **Frog**, the Python blogging system, we can:

- extract a message posted by a user, in HTML directly from it's web-page ([example](#))
- extract a news feed (RSS 2.0) with the last posts of a given user ([example](#))
- extract a news feed (RSS 2.0) with the last posts of a everybody in the community ([example](#))
- news feed filtering by category ([example](#))

- extract an XML data structure containing information about the users, and the categories they have created ([example](#))

In [MoinMoin](#), the wiki provider:

- extract the content of an article in HTML ([example](#))
- extract a news feed (RSS 2.0) of the last changes in the wiki ([example](#))
- extract a news feed (Atom) of the last changes in the wiki (with filtering possibilities) ([example](#) -

[example with filter](#))

- extract an XML index of all the pages in the wiki (with filtering possibilities) ([example](#) - [example with filter](#))
- extract an XML index of all the categories defined in the wiki ([example](#))

## Aggregation

The data aggregator is basically a Python script run by the server. It is able to **download a file** from a given URL, and **stores its data and its data type in the cache**. It's currently rather basic, but efficient enough.

## Access provider

It is actually a simple Python [HttpServer](#). Depending on the HTTP request (see Naming conventions), it can **retrieve data from the cache**. It can then apply some basic transformation, according to the data format (see rendering).

The result of the transformation can then be sent to the client

## Naming conventions

The **proposed URL scheme** uses this kind of syntax:

```
http://[server_name]/[data_source]?page=[page_name]{&nb=number_of_
entries}{&user=[user_name]}{&category=[category_name]}{&norender=t
rue}
```

where the mandatory elements are:

- `server_name` : the host of the aggregator (for instance `opaals.org:4711`)
- `data_source` : the source where data should be retrieved from. (at the moment, only `wiki` and `blogs` are valid entries)
- `page_name` : the name of a page to retrieve. It can be either:
  - the relative path of *the page to retrieve* on the source server. For example, for the wiki, it would be the name of the article, for the blogs, it could be `user/[user_name]/` to access the user's blog
  - a *special page*, starting with the character `_`, which allow access to special features. Current implementations allows : `_last` to retrieve last changes of blogs or wiki, `_list` for a list of articles of the wiki, or a list of blogs' users

and the optional elements are:

- `user_name`: only used for the blogs so far. It allows to filter the retrieved data by user.

Allows access to user's feeds, and user's posted articles

- `number_of_entries` : allow to limit the number of entries which will appear in the generated feed (default value is 10)
- `category`: only used for the blogs to far. Allows to filter the user's feeds using a given category name, or with the id which corresponds to the id supplied by the Frog blogs (in this case, use `cat=myid`)
- `norender`: by default, **data is obtained as HTML**, but setting a value to `norender` indicates that the data will be sent in the format used by the data source.

Note: parameters have to be URL escaped

## Rendering

A basic renderer is already part of the access provider. It is responsible for providing data at least in HTML format. Other renderers may be implemented, internally or externally for more advanced processing.

The obtention of HTML is made either by:

- extracting a given portion of an html page
- converting XML data thanks to XSL Transformation stylesheets

The result of the transformation can then be sent to the client

## Conclusion

### Interest of the solution

With the Opaals Web-Site, we have implemented a centralized and extensible solution to centralize the distributed data generated by all the members of the Opaals community. It should become the centre of information for all Opaals-relative information, thus contributing to a better visibility and dissemination of its excellence.

### Current state

The OKSAggregator is currently a Html page, calling other pages through "iframes". It is currently hosted at <http://oks.opaals.org/website/oksaggregator.html> but it may change. Here is what it looks like:



Illustration 3: Screenshot of the beta version of the Opaals Website

The Website has now been updated with nicer graphics for public visitors. Please note that it is still a beta-version. Its home page looks like this:



Illustration 4: Screenshot of the OPAALS public website

For testing purposes, here are some examples of usable URLs:

- Wiki:
  - home page: <http://opaals.org:4711/wiki?page=>
  - an article: <http://opaals.org:4711/wiki?page=WebSite>
  - list of pages: [http://opaals.org:4711/wiki?page=\\_list](http://opaals.org:4711/wiki?page=_list)
  - XML-list of pages: [http://opaals.org:4711/wiki?page=\\_list&norender=true](http://opaals.org:4711/wiki?page=_list&norender=true)
  - last changes: [http://opaals.org:4711/wiki?page=\\_last](http://opaals.org:4711/wiki?page=_last)
  - RSS - last changes : [http://opaals.org:4711/wiki?page=\\_last&norender=true](http://opaals.org:4711/wiki?page=_last&norender=true)
- Blogs:
  - home page: <http://opaals.org:4711/blogs?page=>
  - list of users (and categories): [http://opaals.org:4711/blogs?page=\\_list](http://opaals.org:4711/blogs?page=_list)
  - XML-list of users : [http://opaals.org:4711/blogs?page=\\_list&norender=true](http://opaals.org:4711/blogs?page=_list&norender=true)
  - last posts of everybody: [http://opaals.org:4711/blogs?page=\\_last&user=all](http://opaals.org:4711/blogs?page=_last&user=all)
  - RSS - last posts of everybody:  
[http://opaals.org:4711/blogs?page=\\_last&user=all&norender=true](http://opaals.org:4711/blogs?page=_last&user=all&norender=true)
  - home page of Javier Noguera:  
<http://opaals.org:4711/blogs?page=&user=Javier.Noguera>
  - last posts of Javier Noguera:  
[http://opaals.org:4711/blogs?page=\\_last&user=Javier.Noguera](http://opaals.org:4711/blogs?page=_last&user=Javier.Noguera)
  - RSS - last posts of Javier Noguera:  
[http://opaals.org:4711/blogs?page=\\_last&user=Javier.Noguera&norender=true](http://opaals.org:4711/blogs?page=_last&user=Javier.Noguera&norender=true)
  - last posts of Mehita Iqani in category "Opinions" (cat\_id=6):  
[http://opaals.org:4711/blogs?page=\\_last&user=mehita&category=opinions](http://opaals.org:4711/blogs?page=_last&user=mehita&category=opinions)
  - RSS - last posts of Mehita Iqani in category "Opinions" (cat\_id=6):  
[http://opaals.org:4711/blogs?page=\\_last&user=mehita&category=opinions&norender=true](http://opaals.org:4711/blogs?page=_last&user=mehita&category=opinions&norender=true)

## What's next?

We have now delivered a first stable version of the Opaals Web-Site. The next steps in its finalization are:

- Integration with the [Newsletter](#) : propose suscription directly from the web-site
- Integration of other data sources from the OKS, to keep on increasing the amount of available data
- Design of a public web-interface, to propose the data in a user-friendly way, according to the branding policy of the Opaals.
- Integration with vizualisation
- and more !

## See also

- [OPAALS Newsletter](#)

- [about XML on Wikipedia](#)
- [about Open-source software on Wikipedia](#)