*The goal of Data Service Infrastructure for the Social Sciences and Humanities (DASISH) is to determine areas of cross- fertilization and synergy in the infrastructure development and to work on concrete joint activities related to data, such as data access, data sharing, data quality, data archiving and legal and ethical aspects. As stated in DASISH task 5.6 description*, *researchers need to be able to store the results of collaborative intellectual work either as an annotation of a single fragment or in the form of typed relations between a number of fragments.*



**DWAN: DASISH Web ANnotator**

# Instrumenting collaboration with web-annotating tools

The availability of digital archives and other research data via the Internet creates new chances for collaboration. Indeed, equipped with special software, researchers from different institution, countries and fields can work together via the network. Such collaboration can take the form of annotating the data and sharing these annotations using an annotation infrastructure. By an annotation we mean a remark over a fragment(s) of a document(s).

# DWAN framework

DWAN is a software framework for possibly multiple annotation clients working together with a single backend. The backend consists of a database and a Representational State Transfer (REST) web service implemented in Java.

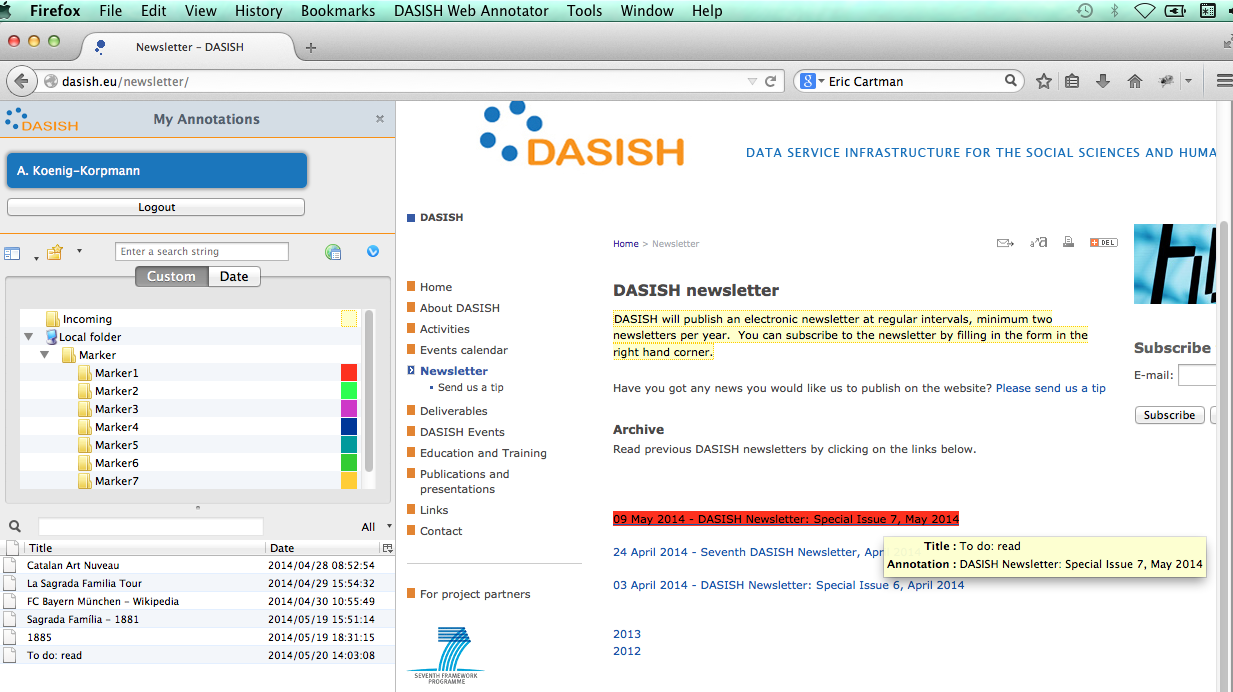
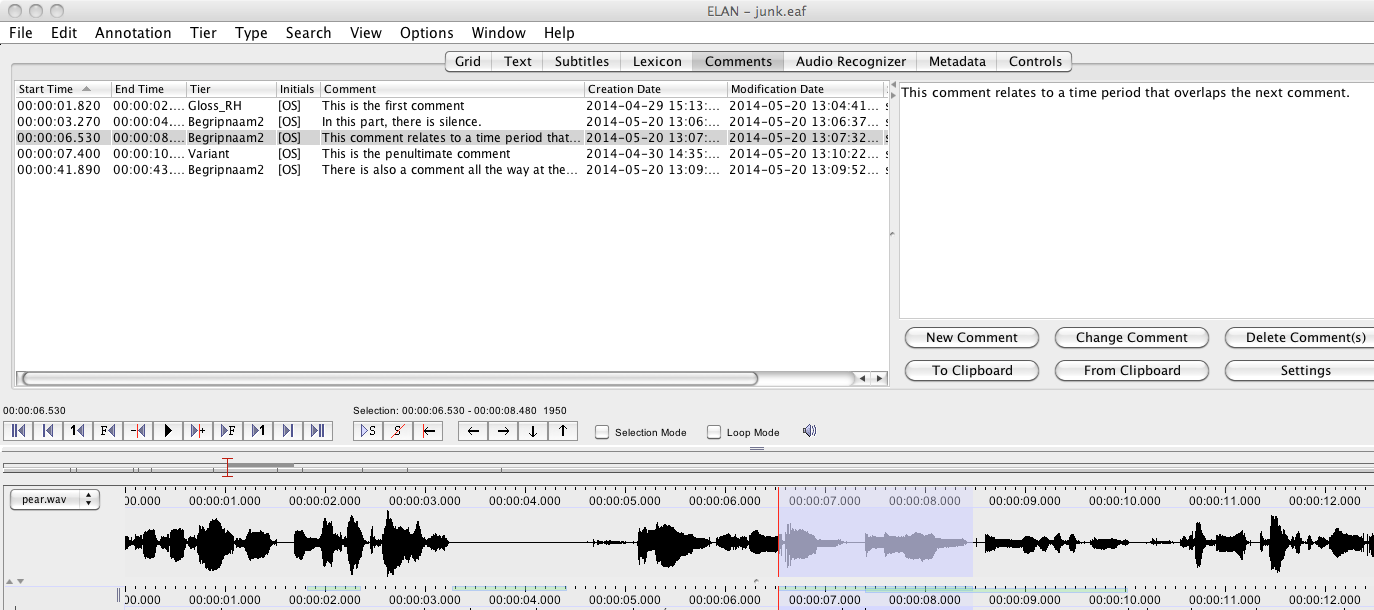
A client is developed specifically for a particular sort(s) of web-documents, whereas the backend is not specific and treats requests of all clients in the same way.

A client and the server exchange data via sending REST requests and responses. A client accepts and sends XML files that obey DWAN XML schema, which is a part of the server-side software. It mirrors a data model that has been designed to represent the main data classes, which are involved in constructing annotations, and relations between these classes.

# DWAN data model

relations *Annotation - Target*, *Target - Source*, *Target - Cached Representation* closely follow the *Open Annotation* (OA) standard. An annotation, i.e. an inhabitant of the class *Annotation*, is a structure that contains necessary information about user's annotation. In particular it contains the annotation's identifier, the reference to the owner and the time of creation. An owner is either the principal who has created the annotation or a user to whom the ownership has been assigned.

Yet another client



Besides the owner, an annotation has *readers* and *writers*. As one can expect, a reader is a user that can read the annotation, and a writer can also add changes to it. Thus, a registered principal can be related to an annotation by means of one of three access modes: *reader*, *writer, none*.

An annotation can have one or more *targets*. A target (i.e. an inhabitant of the "Target" class) contains the reference to the web-document (a *source*) and the precise description of the document's fragment, which is actually annotated. Moreover, a target may refer to one or more cached representations of (the relevant parts of) the target document with the precise descriptions of the annotated fragments for each representation.

The semantics of an annotation is given in its body. In the implementation a body is an arbitrary text or an XML text.

# DWAN backend

The core of the backend is the database where annotations and information about annotated targets are stored together with cached representations of the targets. A cached representation is a copy, e.g. a screenshot, of a target document. Storing cached representations allows to retrieve the copy of an annotated document when the actual web-document under the target’s URI has been updated so that locating the annotation in it becomes difficult or even impossible. It may happen when the corresponding fragment has been significantly changed or disappeared.

# DWAN frontends

**Wired-Marker-based DWAN client** has been developed on the base of the *Wired-Marker* freeware. It is a highlighter that allows marking fragments of a web-document by different colors. An annotated fragment can be a text fragment or the whole image. aAn annotated fragment is preserved not only in the local client’s database within the extension but also sent by the DWAN client as an XML file to the backend database where it is stored.

When a user creates an annotation, the client sends it to the server together with a cached representation of the annotated page (on the moment of annotation). A cached representation can be requested by the user later, for instance if the client cannot deliver the annotation because the page has been changed and the fragment cannot be resolved .

**The ELAN frontend** for the DWAN backend is being prepared in the context of the COLTIME project. ELAN is an annotation program for media files. Comments in ELAN-speak are called Annotations in DWAN-speak There was no specific support to comment on the ELAN annotations themselves. The DWAN backend is an ideal vehicle to store these comments: it is based on comments which refer to some URL, or even more specifically to some fragment of the URL by means of a fragment identifier.

**The frontend for ANNEX**. ANNEX is an open source online visualizer for time-aligned annotation files, primarily targeted at the EAF (ELAN Annotation Format) format. Primarily being a visualization tool for archived materials, ANNEX currently does not offer any creation or modification functionalities over time-aligned annotation files. This leverages the need for extra functionality allowing users to create comments on the existing annotations.

For this purpose the DWAN backend stands out as an ideal server-side engine to store, search and retrieve such comments.

# Outlook

Due to universality of DWAN backend, the whole framework has a good potential to be reusable by other research groups and institutions. An interested ogranisation can either reuse DWAN’s backend code, or share the database with another organization, and create its standalone frontend, or adjust already existing one. Amongst potential candidates to be used as DWAN frontend we would propose the following software: *ANNIS* (specific for multilingual linguistic corpora), *BIBLIOPEDIA* (designed to crawl scholarly resources). *LitBlitz* or NB (for faculties and students)

*MapHub* (specific on annotating historic maps)

Pliny (developed by KCL ), *Pundit*, *Uvic Image Markup Tool* (specific for images), *Virtual Lighbox* (for Museums and Archives), *Zotero* (probably one the most common and used annotation tools among humanists), *WebLicht* .

DASISH - Data Service Infrastructure for the Social Sciences and Humanities



DASISH - Data Service Infrastructure for the Social Sciences and Humanities www.dasish.eu

**DASISH**