

Helene : A collaborative server to create and securely deliver documents for the blind

Guillaume du Bourguet, Benoît Guillon, Dominique Burger

Association BrailleNet, Université Pierre et Marie Curie, INSERM U483

Keywords: DAISY - Braille - Digital Library

The access to written information is essential for the inclusion of individuals in modern societies. At school, at work it is an important success factor. At home it is source of pleasure and cultural development.

In this paper we describe a service that has been developed to improve the cooperation between the different actors involved in producing and distributing books in alternate formats for visually impaired persons.

The Helene Server 2.0 provides a complete framework to manage books information, publishers relationships (authorization and retribution) and to track editing and delivering workflow. All exchanges from Helene are highly secured to protect publishers from unfair use. We also provide a set of tools to bring plain text files to a more structured DAISY 3.0 format ready for Braille printing.

<http://www.serveur-helene.org>

1 Adapted publishing in France

Making a book accessible to visually impaired people means its digitization, several format conversions and a human expertise on the content (captions on graphics and pictures, science content transcription, ...). In France, this work is done in a great part by the adaptation centers. They can be small sized non-profit associations or specific departments of schools and public libraries. Books are transcribed and printed in Braille upon request of teachers, pupils, parents or any visually impaired reader.

1.1 Adaptation techniques

Adaptation centers work very rarely on source documents provided by publishers. Most of the time, they need to digitalize or even type the text. The resulting file is an unstructured text which needs to be enriched for Braille printing or computer browsing.

For Braille printing, it means to add specific flags or markup in the text (centered text, headings, tables, footnotes, contracted Braille ...) then to process the text for a specific make of printer. Different software are available to edit source files for Braille, convert them in proprietary formats and print them in Braille devices. Most of the adaptation centers use their own software and format (e.g.: BrailleStar, Abrotec, Gibon, ...). Adapted files cannot be easily reused and shared between centers. Most of the adaptation centers are aware of upgrading their archives to a more general format, from which adapted files could be generated upon demand.

1.2 Issue with copyright in France ¹

Unlike numerous countries, the French legislation does not include a copyright exception for adapted publishing. Adapting a book needs the publisher's agreement. A National Committee for Adapted Publishing report shows that 2/3 adapted materials are published without any agreement of the intellectual property rights owner. This situation is the consequence of the difficulties encountered by adaptation centers to discuss adaptation rights with publishers. Actually, most of the publishers are reluctant to give their authorization to these numerous small sized centers. They need feedbacks on how books are processed and distributed and more warranties on the respect of the copyright. Even when positive, publishers answers spend too much time to satisfy needs.

1.3 The need to create a cooperative framework

In spite of efforts made, these facts show some great limitations in adapted publishing in France. The Internet and electronic publishing provide an opportunity to create a cooperative framework where:

- resources would be shared in an intermediate accessible format and delivered in multiple formats,
- cooperation between adaptation centers would be made easier,
- requests and answers of publishers authorization would be centralized and better monitored,
- the intellectual property rights would be better guaranteed.

The Helene Server has been designed and developed by the BrailleNet association, the Pierre et Marie Curie University, INSERM U 483 and the INRIA. Its aim is to fill the needs of adapted publishing. A preliminary version has been developed and tested in real situations during 2001 and 2002. Several functionalities have been added and improved in the version 2 released in the first months of 2003.

2 The Helene Server, an accessible digital library

2.1 The catalogue

The main functionality of the Helene Server is an Internet bibliographical catalog. Each book is characterized by its title, author, publisher, ISBN, public or private domain, ... Books can be easily retrieved using multi-criterion searches, filters or alphabetical index.

2.2 Management of publishers' authorizations.

Each private book recorded in the catalog is attached to the authorization of its publisher. It contains a status (authorized / not authorized / in progress), a start date, an end date, a renewal date, and the price to pay to the publisher for each copy done. Unauthorized books are not displayed in the catalog until the publisher gives its agreement.

Administrators of the catalog can edit authorizations or create different lists of books regarding to their authorization (authorizations to ask, not authorized books, authorization in progress, ...). A contact manager of recorded publishers is also at the disposal of administrators.

The requests to publishers for literacy books are handled by the BrailleNet association using the Helene Server. More than 170 French publishers have been contacted, some of them have established flowing relations with BrailleNet, where whole collections of books received authorizations.

2.3 Website accessibility to visually impaired people

A particular effort is made on the compliance with the w3c guidelines for accessibility of the Internet ². For instance, lists of books (catalog, search results, index ...) are presented in an accessible HTML table with a summary and references to headers in each cell. These tables are easily understood with a screen reader or a speech synthesis. Invisible internal links are available on each page for blind users, they provide quick access to contextual menus and contents. Each page also contains a quick search form. The whole HTML code of the public web interface is XHTML 1.0 compliant.

3 Users level access

3.1 Registered partners

The catalog of authorized books is public and can be browsed by anybody on the Internet, but the main functionalities of the Helene Server and access to file under copyright are reserved to registered adaptation centers.

Each registered center signs a contract of intellectual property rights respect and gives a list of transcribers allowed to use the service. An account is created for each transcriber (a login and a password is given) so that they can authenticate themselves and access to private files and advanced functionalities.

3.2 Access levels

Administrators of the Helene Server can fully edit users account and choose their access level between :

- transcribers, who can access to any file, modify them and add new files and books. They are also allowed to use statistics on formats to retrieve files to adapt and can browse the list of books added or downloaded by their center.
- printers, who can access to already adapted files only.
- authorization administrators, who can edit the rights of books they have inserted in the Helene Server. This access level is used for school books whose authorizations are not managed by the BrailleNet association.

The different functionalities offered to authenticated users appear in contextual menus on concerned web pages (e.g.: add a file to this book, renew this authorization, ...) and in a specific tool bar in the top of each page (add a new books, display authorizations the user has to administrate, ...).

Authenticated users can access to their home page (history of user) and edit information about their organization and their account.

4 Files management and format conversions

Files storage

To provide a reasonable security level, files of books are gathered in an isolated secured server. This server can only be accessed by the Helene Server's web site via an encrypted protocol (HTTPS). The file server is protected from external requests by a firewall which filters incoming and outgoing network traffic.

The file server also saves deleted or modified files. Previous versions of files can be restored and a history of modifications is kept for each file.

Formats analysis and harmonization

The great variety of formats gathered in the first two years by the Helene Server led us to start a fundamental work on digital formats.

In outline, input files can be:

- **plain text or unstructured formats.** They are the result of scanning, typing (performed by adaptation centres) or conversions from mainstream publishing formats. The most encountered formats are: RTF, Microsoft Word, unstructured html, plain text.

- **adapted formats for Braille printing.** These formats are helpful to print in Braille but they cannot be processed to create accessible digital documents. Most of the files we received were encoded in BrailleStar (text or Microsoft Word). A lot of other specific formats were found.
- **semantically rich formats.** With the expansion of computing technologies, most of the publisher have streamlined and automated their production. The formats they use are in major part based on XML, using particular document type definitions (DTD). Some publishers supply their XML files to the Helene Server. These files can be transformed almost automatically in accessible digital documents.

To provide an attractive publishing scheme for adaptation centres, the Helene Server should be able to output files in:

- **adapted formats for Braille printing.** These formats are still essential to Braille printing, and an effort is being made to unify software and formats used in France (Duxbury DBT).
- **accessible digital formats.** Braille printed documents are cumbersome and not practical. At school, with “Digital School Bags”, at work or at home, visually impaired people are led to use computers with helping technologies. When provided with a rich structure and alternate contents for blind people, XHTML 1.0 ³ is a good example of accessible digital format. It can be read using speech synthesis or a Braille device. It can contain internal links to create table of contents and browsing functionalities. Its modularity allows to include mathML data for school books.

Harmonizing a such variety of files requires a central intermediate format from which different type of documents could be automatically generated. We chose the DAISY 3 ⁴ format (ANSI/NISO Z39.86-2002). Based on XML, it offers a general matter structure and can be extended with specialised modules. It allows text and sound synchronization; but at this time the Helene Server handles text-only DAISY 3 files.

Conversion tools

- **Upgrading files to DAISY 3.** Converting BrailleStar files or unstructured files to DAISY 3 needs human expertise for each file processed. A Microsoft Word Template has been developed to make source file edition easier. It helps to create a RTF document, parsing a BrailleStar file. It allows to edit styles applied to the content - especially Braille specific styles provided in the template.

To convert these RTF files to DAISY 3, the Helene Server provides a conversion service based on upCast, a java tool created by Infinity Loop. Braille specific styles are also imported in the resulting DAISY 3 files.

- **Generating output formats from DAISY 3.** To generalise file conversions, the Helene Server implements an unified interface for conversion services. A set of XSLT style-sheets provides an XHTML output with table of contents and internal links. These XHTML files can be shrunken in multiple linked files and so used as an accessible digital book.

A Duxbury converter is developed to provide a Braille output to DAISY 3 files with Braille styles.

These conversion services can be used by authenticated users. The Helene Server manages files synchronization: when a source file or a converter is modified, the corresponding source files become deprecated.

5 Files delivery and usage monitoring

5.1 Security issues on file delivery

To guaranty the intellectual property rights of private files, the Helene Server outputs them in S/MIME⁵ format. This standard widely used for secured e-mails allows to encrypt a file using asymmetric encryption methods.

Each registered user has a X509 digital certificate. It contains its private key and information on its identity. To decrypt files sent by the Helene Server, the targeted user needs its own certificate, generated by the Helene Server.

5.2 Usage statistics and reports to publishers

Every action of an authenticated user (file download, books added, files modified, ...) is stored in a database. The use of the service can be easily monitored by the administrators (e.g.: download list and evolution, added files, use of centers, history of files,).

Adaptation center can edit the number of copies they printed for each file downloaded. A general report can be generated for publishers every year. It sums up downloaded files, copies done and retribution for each publisher.

Conclusion

The Helene Server provides an efficient framework for adapted publishing. It helps to centralize efforts made for books adaptation and improves collaboration between centers and formats harmonization around the DAISY 3 standard. To ensure intellectual property rights respect, it offers a way to monitor publishers authorizations and usage of files.

¹ Bridging the Gap Between Regular and Special Publishing
Catherine Desbuquois, chief librarian, Ministry of Culture and Communication
<http://www.snv.jussieu.fr/inova/villette2002/act11.htm>

² Web Accessibility Initiative (WAI)
<http://www.w3c.org/WAI/>

³ XHTML 1.0
<http://www.w3.org/TR/xhtml1/>

⁴ DAISY 3
<http://www.daisy.org>

⁵ Secure/Multipurpose Internet Mail Extensions (S/MIME)
<http://www.ietf.org/rfc/rfc2311.txt>