Organizational and Technical Aspects of an Electronic Publishing Framework for Historical Sources

RIK HOEKSTRA
Institute of Netherlands History (ING)
Prins Willem Alexanderhof 5, The Hague, The Netherlands
e-mail: rik.hoekstra@inghist.nl

This paper discusses the way in which we constructed a framework for the publication on the web of historical sources at the Institute of Netherlands History (ING) in The Hague (The Netherlands). The framework is built using the web application platform Zope and is flexible because of its modular construction. This makes it possible to publishing heterogeneous research data with their own requirements for accessibility and usability, while maintaining uniformity in presentation and access whenever possible. Because a number of different approaches were used in its design, it is not the end solution to a problem, but ready to grow and evolve as new features are required.

Keywords: web publications; web framework; Zope; historical sources

INTRODUCTION

This paper presents a discussion of the construction of a web publication framework at the Institute of Netherlands History in The Hague (Netherlands).[1] ING is the leading publisher of historical sources about all aspects of the history of the Netherlands since a hundred years. In 2000, ING decided that the web would be its preferred media for electronic publishing of research results. The main goal was to provide a platform that would allow each publication be accessed in its own right, while keeping as much unity in presentation as possible. This paper addresses the challenges we faced when we had to devise a scalable and future oriented way of publishing a wide range of historical sources. We believe that our publication framework has provided us with the technical and organizational solutions to these challenges, but, of course, we do by no means claim to have resolved all issues.

Below, I will first discuss the different aspects of the situation at the institute as we perceived it. ING is an institute with a long history in research and in the publication of sources. The analysis of the situation led to a number of starting points, guidelines and approaches we have taken to come to a solution that had to be sound, workable and as much future proof as possible. The design of a web publication framework was based on the preparatory work that made it possible to formulate a number of technical requirements. For the implementation we needed a technological toolkit, in our case that came to be Zope, the open source web application development framework. In implemented the framework in Zope we made use of its the object oriented nature, that allowed us to divide responsibilities between different layers. Some examples are presented to make the working of this construction more concrete. The final section of the paper is concerned with future developments in the framework.

THE SITUATION AT THE INSTITUTE OF NETHERLANDS HISTORY

The purpose of the Institute of Netherlands History is to publish historical sources. In elaborating and editing sources there is a century long experience. During that time, a lot of knowledge about the practice of making sources accessible, of annotating and indexing and of publishing them were accumulated. This has led to a number of explicit and implicit best practices and standards in source publication.

Of course, this applies to paper sources, as ING publications always were in the form of books. Since the beginning of the twentieth century in total some 350 of these State Historical Publications have been published, amounting to a wall of respectable green volumes. If anything, the institute is known for respectability and reliability in historical source publications. Our most important users have always been professional historians, the main language of introductions and annotations always was Dutch, the documents inside the volumes were in many different languages. For example, the letters from and to William of Orange were in Dutch, French, German, English, Latin and Spanish and some of them were encrypted. The use of (sometimes many) different languages, sometimes even within one document, and of old versions of Dutch poses many particular problems with regards to information retrieval. They go beyond the topic of this paper, though I will breifly return to them in the section on future developments.
In recent times, at ING there has been an increasing awareness that the future of source publications would for an important part be electronic. Historical source publication are hardly ever meant to be read from front to back, and in many times the table of contents in the beginning of a publication, the index at the back of it or cross references inside were used as guides and search tools. It was plain to everyone, after the usual initial resistance in organizations to the innovation of electronic means had been more or less overcome, that ‘computers’ would be much better suited to search and index source publications, especially after a number of databases had come into use in a number of research projects. At the time, databases in use were of the common desktop type: DataPerfect, Dbase III, Dbase IV and MS Access. For use in projects MS Access is still regularly used, and though we try to find alternatives, these are not so simple to find for practical reasons.

After some thinking and searching, it was decided that institute policy would be that in principle publications would be published on-line, by way of the Institute website. Other media (cd-rom/dvd) were never excluded, but the web had clear advantages over them. In academia it seems to have become common practice to publish research results on-line. The web and the browser have become a versatile and adaptable way of publishing research results, both documents and dynamic applications like databases. Distributing CDs is much more burdensome, CDs are platform dependent and user support and trouble shooting or upgrading applications are complicated and expensive. The only reason to publish on CD as I see it would be that it is easier to ask money for them – that is not to say that it is easier to earn money because of the high costs involved in developing CDs. All materials on the ING website are freely accessible, even if there is no guarantee that this will stay like this in the future. At this time, there have been not yet been CD-ROM publications. If in the future it would be decided to use CD or DVD as a media for a particular publication, it is almost certain that they would use formats and standards that were web compatible (XML, browser based interface etcetera).

Institute research is organized in projects – each project results in one or several publications. Research projects are entities in themselves. They treat a subject from the history of the Netherlands, but do not relate it to a larger story. Therefore, the publications cannot be read as one continuing story that begins at the time when the first people arrived in the marshes that are now part of the territory of the Kingdom of the Netherlands to the present day. Seen from this larger perspective, institute research are all snippets. Moreover, the orientation of the publications varies a lot, as they range from archival guides to source-text publications. The publications of the institute are all concerned with historical sources, but they are very diverse in orientation and form. In preparing the publications, researchers at the institute use the common variety of techniques: word processing, imaging and scanning (mainly of manuscripts), databases, XML and all possible combinations to prepare the publications. Each of the publications has its own requirements in terms of usability and accessibility.

To give an impression of the research material on our site and of its diversity, here is a limited list of examples, with a short characterization of content and technical format. The examples are as they now are accessible, many have already been converted from earlier technical formats:

- the reconstructed archives of the Centrale Inlichtingen Dienst (Central Intelligence Service): this is a web accessible database with the description of some 2000 secret police files from between the world wars. The files themselves are attached (ca. 20.000 images)
- Biografisch Woordenboek Nederland (BWN; Biographical Dictionary of the Netherlands): digital version of nearly 2000 biographies of well-known Dutchmen, that had previously been published in print. XML-files with digital indexes, suitable for making selections of the biographies
- Provincial finances in the Netherlands, 1600-1800: Financial accounts of the Dutch provinces. Now, data about three provinces out of seven have been published. Relational database with a web interface - under construction
- overview of the information processes of the 'Bataafs-Franse' government: overview of requests for information issued between 1795 and 1813 with the answers. Accessible through a relational database with a web interface
- guide to small political parties in the Netherlands 1900-1967: web accessible relational database with descriptions of the bewildering variety of political parties that have always characterized the Dutch political spectrum
- chronicle of Johannes de Beke: digital publication of a previously published book in pdf format

In addition the following publications are in preparation for the short term (next eighteen months):

- an edition of William of Orange's correspondence (relational database of 10.000 16th century letters and attached scans of the manuscript)
- a guide to the sources to the history of Social Security in the Netherlands in the 20th century (descriptions of archive collections in XML)
- Dutch Asiatic Shipping, a database of all voyages (ca. 8000 in total) of the East Indies Company (VOC), relational database on basis of retro-digitized existing ING volumes.
- an edition of the medieval chronicle of Melis Stoke (transcriptions in XML format of various manuscript versions with images)
- an edition (on an XML basis) of the resolutions of the seventeenth-century States General

The heterogeneity of the research of the institute are not likely to change in the future. More projects will not add up to a continuing story and attempts at standardization will not make publications more uniform in outlook, organization or structure, because their nature is different. Still, when uniformity is not the goal, there are aspects that do lend themselves for a more uniform treatment. This is why we chose to build a web framework for our digital publications instead of publishing each publication separately on our website and leave it entirely to the user to try and integrate data in them. The framework is meant to provide context and cohesion to the publications which they would not have when they would be stand-alone digital publications, but not to force them into a model or a technological straightjacket. The idea was that as the framework grows, so will the context for each publication.

**APPROACHES FOR A WEB PUBLISHING FRAMEWORK**

The users of the institute publications always were professional historians. Books have to be bought, or borrowed in a library and it would be vary rare that someone would end up in a State Historical Publication by unintentionally or, once they held a volume, not know by its appearance it was a serious work of learning. In contrast, by nature the web has a much lower threshold than books and, once on-line, anyone could end up in our research data more or less by accident. However, our publications traditionally were sound from a professional research point of view, but not easily accessible for anyone, because they contained documents in foreign languages, contained handwritten sources that were hard to decipher or are of a highly specialized nature. Therefore, we had to be explicit about who we saw as the users of our website and what they should be able to find on it. While everyone is of course welcome to visit the site and use the materials on it, the most important target group of the website still are the professional historians, with journalists and interested lay people as a secondary target group. It would be an illusion to think that our research material would be of interest to a general public. Making the site a something of a portal for Dutch history would require a persistent effort that is beyond our means and does not respond to the aims of ING. At the same time, we strive to make the material as easily accessible as possible, by giving proper guidelines for use and making searching and querying the research projects as easy as we can.

Anyone who comes to our site should be able to find quickly whether there is anything on it to his interest. We have strived to make searching and navigating as transparent and coherent as possible. This is difficult to accomplish, however, and it is a point of our ongoing attention. Finally, the site is almost exclusively in Dutch, because most of the underlying research material is in Dutch and users should be able to read Dutch anyway. Making introductory material available in English would only be giving users false expectations. Nonetheless, it is likely that in the future more material will be available in English.

The web publishing framework implementation of the Institute of Netherlands History consists of a framework part, giving a helicopter or introductory view of the Institute and our research and a structured way to access the ‘projects’ that organize the research and publications. Navigation to projects is on basis of a menu. In accordance with usual practices on the web all site-wide navigation takes place in one place, in our case on the left of the browser window. The projects all have a structured homepage, that provides basic information about the research and links to details about paper publications, but they also provide entrance points for exploring the project and the material in them (if any) further. Elements of the standardized description are the timespan of the project, institute staff related with the project, contact information, the publications originating from the project and access to electronic materials attached.

The projects themselves are together a collection of sub-websites of their own that are much more specific and tailored for the research and/or publications they give access to. These sub-websites of the various projects do not have a uniform structure. Even if layout and design are should match that of the main framework, we strive to give each of the current projects something of an identity of their own. The reason is to make them recognizable for both its target group and the institute staff that is working on it. There are over one hundred projects available now. Many of them just describe traditional publications and the number of projects with electronic publications attached is still limited. However, content is growing fast and projects will be added each year. The same is true for the framework as a whole, and features are added on a regular basis. It is designed to be flexible and the intent is that each feature contributes to make it stronger, without the necessity to make fundamental changes or making the construction less maintainable.

Many projects relate to each other, because, for example, the same people played a role in different events, or several events took place in the same town at different times. In this way, they various research projects provide
context for each other. Context may be synchronic because, for example, events in different places may be related or of importance to each other or diachronic, for example because various members of one family play a role for several centuries. Because the purpose of ING is to publish sources and not interpretations, it is impossible to provide all possible context for researchers. On the contrary, in electronic publishing we should provide not the context but make it possible for researchers to create their own context. The tools to do so depend, however, on a certain level of standardization. Of course, researchers are allergic to what they perceive as too much structure, because they tend to see them as limiting in their way of working or forcing them to cram reality into a model that comes from the drawing table. It is hard to prevent this sentiment altogether, but we have sought for a way to do so. To promote standardization we have adopted several of what we call data structures.

Data structures are small entities of coherent information about a specific entity, such as a person, a geographical place, a date, etcetera, and guidelines how to fill them in. In other contexts they would perhaps be called ‘business rules’. In examining available research and databases at ING, it was clear that in making information accessible, researchers tend to make use of a number of the same descriptive entities, of which people, places, keywords and dates were the most important. The data structures can be easily combined into, for example, a relational database schema. Many research projects also call for a more specific or more extensive description of certain of the data they contain. The data structures do not pretend to describe more then their own entity, they can easily be extended in a specific application. Because the data structures are logical, they are neutral with regards to technical implementation. In other words, they work for XML as well as for a relational database or for Zope objects.

In using a publishing framework for providing a platform for publication, uniformity is brought to the presentation of ING research project on the web. Because the framework was built as a container, it provides a general method of access to diverse materials. In the navigation we used project meta-data (time span, a limited set of keywords) to access the projects. The sub-websites of the projects can be almost completely tailor made for the project data. There, the framework interferes as little with presentation and logic as possible. The framework is the top down approach to the challenge of uniformity and diversity the availability of much diverse materials in one institute poses.

The data structures also provide structure, but approach the issue from the other side, that is, bottom up. Because they describe entities as atomic as possible (or useful), they have a minimum of impact on project structure as a whole. Ideally, they just are standardized building blocks that combine with other entities to a structured description ‘format’ that may at a later time be employed to retrieve data from different project at a time in a structured way. At the moment, the amount of available data in related publications is about to pass the threshold in which we may start to combine them, for example in making information on people that appear in more than one project accessible from all projects involved.

RESEARCH PROJECT DESIGN

As I described above, existing knowledge of ING researchers had to be employed as much as possible. This was not so easy, as much of it focused on paper publications. Much of the knowledge is also usable for digital publications, either unaltered, for example in selecting material or in editing sources, or with modifications, as with the decisions what data about or from sources should be indexed and in what way. However, information technology uses much more structured ways to record data than paper does and knowledge and even awareness about techniques or methodology in this respect is limited in researchers with a humanities background, with a few exceptions. Arguably, in the long run at least the awareness should be on a higher level, but for the time being it was not realistic to assume this would change quickly. The most important question in this respect was (and remains) therefore, how to make researchers think about the electronic form and accessibility of their publications, without making them think about technology directly.

We have now adopted a methodology that makes use of target groups and use cases to communicate the purpose, structure and construction of the technological component of a project with a researcher. Usually, there is an analysis of the content and extent of a research project. To start working, a researcher needs either a database or some other structure to describe his sources in. The methodology is really quite simple, because all a researcher has to do is to write down who he intends to be the different groups of users of his project and what they should be able to do with the sources he has made accessible. However, it has proven to be very useful, as it forces the researcher to focus on the application of what he is doing or going to do and how it may be accessed and not so much on all scraps of information available in the sources and the possible loss of particularities in the work of them. When thinking about use, everyone is able to see that too much detail will not make material more accessible, but just clutter your interface. Moreover, a description in terms of use cases is concrete enough to make it possible for developers to communicate with researchers, so that it will prevent misunderstandings in the development cycle, at least to a certain extent.[2]
FRAMEWORK IMPLEMENTATION USING ZOPE

The discussion of the features of the publication framework as they emerged from the discussion of the situation at ING led to a number of technical requirements. They ask for a versatile toolkit that is capable of many things: it should be open, flexibility and support technical standards (existing and emerging). We based the publication platform on the leading open-source Web application platform Zope.[3] It would go too far to discuss Zope’s technical features here in detail, so I will only outline those that were of most important for ING.

Zope has a modular architecture and encourages the separation of data access, retrieval and presentation. This makes the platform very flexible and allows for seamless integration of a variety of resources, of which relational databases of most common varieties and XML documents are the most important. Zope operates on basis of an object oriented database, in which everything is an object. Objects in Zope come basically in two varieties, container object hold other objects, content objects may hold ‘anything’ that may be considered content. These include templates for presentation (Zope has its own templates), (Python) scripts for logic or for querying either a relational database or an XML document, but also document in any format. Objects are versatile building blocks that can be used to build a large site from smaller parts and that make it possible to combine and reuse parts once they have been built.

In addition, Zope comes with so-called catalogs, a built-in machinery that makes it possible to index objects contained in the object database on basis of their properties. The properties include what can best be called meta data and may contain all sorts of values, from integers to strings to dates and keywords. Catalogs also have a text index that may be used for searching documents (or other objects) on basis of a query on words or phrases.

Zope offered features with which we could build a containing framework that set standards for presentation and for consistent behaviour of all the objects contained in it. It also enabled us to plug in new projects and to use their own data access, retrieval and logic. At the same time they share most of the publication logic and presentation facilities available in the platform itself. Zope scales well and performance has not been an issue till now. If it becomes an issue, it is easy to distribute applications over extra hardware. Zope itself comes with a toolkit to enable load balancing. In addition skills and knowledge were available in the organization as I myself had previous experience in programming Python and in using Zope. When I started to use these tools many of my requirements were comparable to those that we outlined at ING.

Zope has a development philosophy that makes it possible to do most developing and also content management through the web. It comes with a built-in extensive security mechanism, that is specific to Zope. It does not make it necessary to give users or developers physical access to the machine Zope is hosted on. Here, I will not go into these features any further, but for us it has proven a very useful feature as different developers can develop different applications, or fix some bugs, at the same time and remotely, real time and without having to take the website off line. Obviously, development work is done on a testing machine and all applications are extensively tested before they go live, but still this has proven very useful at times. At ING we do not, at the moment, make extensive use of the content management facilities of Zope, because we have rather strict editorial policies about in which only a few people are allowed to make changes. Probably this is going to change once our intranet is in full operation, but at the time of writing, it is still under development.

LAYERS AND RESPONSABILITIES

The way we employed Zope is best explained by an outline of the implementation of the publication framework. The framework consists of several layers, that all have their own responsibility. The framework layer has the following responsibilities:

- making ‘everything’ inside it as uniformly accessible as possible
- provides services that are shared throughout the framework. These should make development for individual project shorter, because much of the necessary machinery is already there. Zope comes with many services built-in and there are many Zope extensions that make it possible to plug in new functionality. The use and handling of these services is transparent, not only for the user of the web site but also for a large part for the developers, even if sometimes an initial installation effort is necessary
- provide a standardized and automatic mechanism for navigation and searching data inside the framework. Navigation is dynamic; the framework has to reflect objects that are added, deleted or changes in the navigation structure and its search results
- provide context to research data: ideally more research data should contribute to each others context. There are several levels where this may be done, the most obvious is that they appear in the same menus in the site hierarchy, but we have not yet used the possibilities here in full
- provide a connecting mechanism for research data and for documents and relate data whenever possible. This is still in early development.

All in all the responsibilities of the framework may be summarised as enabling for the applications and data it contains. It makes it possible to publish our research results (databases, XML content) and other content in such a way that we can concentrate on the features in it that are specific for the content in question and do not have to worry about their basic functionalities.

The content layer, of which the research project layer is the most intricate and most important has the following responsibilities:

- making research data accessible in the best way available. This is always a compromise between available means, available research material and the elaboration needed and desires of researchers and institute management. The design is done by a developer in co-operation with the researcher, by employing use cases, as explained above.

- where possible the project has to make use of the services of the framework. In fact, it is the responsibility of the developer working on a project to take this into account. The constraints the framework imposes are kept to a minimum. The most simply use of shared services is the use of a central stylesheet (that may be extended for a specific feature), but there is also a central repository for editing and presenting XML documents (on basis of the Silva extension to Zope).[4]

- notify framework of changes in the project as far as they are of importance to the framework. Some notifications should be done by hand ('update the index'), but most are automatic.

- a project is a Zope container object in our framework. It contains the meta data about the research material and application contained in it and people and paper publications associated with them. At the same time, the database adapter or the documents and all application specific logic also reside in it.

- the projects should display their meta data when they are asked to. This may be to the framework internally for indexing and navigation or in the form of Dublin Core properties to display in the browser or to external search engines.

- To make the framework manageable we have made a division between presentation, logic and data. A few examples may make clear how this works in practice.

- The projects (and some other items) are programmed in Python as a Zope 'Product' as components are called in Zope. This makes it possible to make changes to the definitions of this 'ING Project' that reflect immediately on all of the projects in the framework. When there is a bug, or when new behaviour or features are desired, relatively minimal programming effort is required to update the framework.

- The shared services and Zope's modularity make it possible to start developing an application with the use of an MS Access database backend (through ODBC), because researchers are still working on their database. By the time the application goes live, we can switch to MySQL for performance and maintenance reasons. When sticking to straightforward and standard SQL in querying the database concerned, this is possible with only very minor changes in the application.

- The projects hold a number of menu items that are specific for the project concerned and that serve as visual placeholders for users that navigate through a project. Usually these are items like 'simple search' and 'advanced search'. All navigation goes to the left of the browser frame and the framework has the permanent navigation menu that enables the user to go to the main subdivisions of the site (Institute, Research and Links) as well as more general items such as help, site search and a reaction form. When a project is shown that holds a project menu of its own, the framework will display this under the general menu items.

- In the website everything should be accessible directly through a URL. In this way, a historian can directly refer to materials he found on the site, whether they are documents or database records. We have a commitment to keeping the site structure stable as long as we can keep this under our control. For technical reasons, it is not always easy to keep URLs clean, especially when database queries are involved.

- The meta data that are displayed on each page of the site are now still in the form of Dublin Core elements in 'meta' tags in the HTML header. Above, I have already explained that it is the responsibility of a project to display these data. From the object point of view, this way of representing the meta data is just one possible representation. It is simple to make a template in the framework that make a representation of the same data using another format, for example using XML or another XML schema. In the same vein, using another presentation template, some of our projects use alternative presentations of database queries: a HTML representation for use in a browser and a comma separated values representation for those who want to do further calculations with them. These two representations use (most of) the same queries and logic.

These examples are not a systematic treatment of the framework construction and the research material in it, but they are only meant to give a clue in what ways the use of the framework, of objects and of a separation of data,
logic and presentation has made a complicated construction manageable. Many of the features I have discussed are also present in other content management frameworks or application construction toolkits in one combination or another. It has not been my intention to present Zope as the ideal answer to all web programming tasks, but rather to sketch the way we have used it to solve a particular situation at one humanities research institute. This said, Zope has proven to be an excellent toolkit or platform to address these problems and the framework constructed provides a good technological and organizational basis to grow.

FUTURE DEVELOPMENTS

The division between framework and individual applications has proven to be a good solution for the problem of heterogeneity in data. The framework provides a uniform and structured view of the data in it. The structured and hierarchical constitution of the framework in practice has some disadvantages. In practice, it appears that some users have trouble finding what they want, but we have decided that some degree of complexity is inevitable in a site that is designed to grow. It would be possible to create a prominent place or even subdomains for the most popular sections of the site, of which the biographical dictionary that contains some two thousand biographies of notable Dutchmen from the twentieth century would be the most likely candidate. We have explicitly decided not to do so, because it would break framework structure and because if we make one section more visible, this automatically makes other sections less visible and that is not what we want.

Much of the framework construction is hierarchical as it is and works with hierarchical menus. This is a result of site design, and for most purposes it is a good solution. However, we seek ways to relate data in such ways that increasingly the site will be ‘webbed’ and the hierarchy does not have to be followed to go to related sections of the site or even related data. Zope supports services that allow us to do this, but it is not a simple issue. First, the requirement applies to different levels in the site, as we want to relate different projects to each other, but also make an association between entities that appear in different projects. Zope has the flexibility to make crosslinks using using metadata or by explicitly linking two objects. The use of standardized data structures should also make it possible to cross search research data from different projects. Parts of the necessary infrastructure to make this possible are already there, but most design and all implementation details still have to be worked out.

Another issue that is hard to solve is that of text retrieval in our documents, and, being a humanities research institute, ING is still largely text oriented. For this, we plan to use indexing and querying. The problem is perhaps a bit different than in many cases where a large amount of unclassified documents has to be searched, because the nature of the document collections to access through indexing and querying is well defined because they belong to a single research project. For document meta data or items that formerly went into the index of a paper publication we now increasingly use XML. However, many texts are in Dutch (or other languages) that differ a lot from contemporary Dutch in use, grammar and spelling.

Even more importantly, searching words in texts is an inevitable and required part of many of our applications, but it is also the most low-level, awkward and difficult way to access textual information. Books had the advantage of browsing texts, and this experience is not easily recreated in an electronic environment. At the moment, it is not so much the technical implementation that is the main stumbling block, as the technical infrastructure is flexible enough to accommodate many different solutions, if necessary even in combination. For now, we are still looking for a workable concept and design that has to underly such a browsing strategy. This will be perhaps the largest challenge in the coming years.

CONCLUSION

We only started two years ago with the web publication framework and it has been in production use for a year. Because we used a number of approaches to tackle the challenges from different angles, the overall result was more a direction with small goals than one big final goal consisting of a large ready-made system to solve all our problems at once.

The efforts resulted in an evolving web publishing framework, with a highly modular structure, of which we hope and expect that in the future it will grow and adapt to changing needs. For the moment, it has fulfilled the requirement of flexibility and adaptability in publishing heterogenous research data where they require their own specialized interfaces for optimal access, while maintaining uniformity in presentation and access whenever possible. Using the object oriented features of Zope, the framework also allows for reuse of technology and ‘building blocks’, in presentation, logic and data access. The basic structure is in place and has proven to work and provides a solid foundation for future enhancements.
NOTES AND REFERENCES
1 ING website, which is also its web publishing framework, at <http://www.inghist.nl>. Website is almost entirely in Dutch.
2 Use cases are employed in UML and in the Zope ‘fishbowl process’ <http://dev.zope.org/Fishbowl>. Our employment of use cases is a very loose adaption from these examples.
4 Silva is an open source web-based XML authoring and publication system available from Infrae <http://www.infrae.org/products/silva>