Digital Culture & Heritage
Patrimoine & Culture Numérique

Haus der Kulturen der Welt, BERLIN
Aug. 31st - Sept. 2nd, 2004
31 Août - 2 septembre 2004

BIRTH: BUILDING AN INTERACTIVE RESEARCH AND DELIVERY NETWORK FOR TELEVISION HERITAGE.

Alexander Hecht MA, MAS - ORF-Austrian Broadcasting Corporation - Documentation and Archives (Austria)

Andrew O'Dwyer - British Broadcasting Corporation Information and Archives (UK)

Johan Oomen, MA - Netherlands Institute for Sound and Vision ICT Department (The Netherlands)

DI (FH) Florian Scharinger - Joanneum Research - Institute of Information Systems & Information Management (Austria)

Published with the sponsorship of the French Ministry of Culture and Communication

Actes publiés avec le soutien de la Mission de la Recherche et de la Technologie du Ministère de la Culture et de la Communication, France

Interprétation simultanée du colloque et traduction des actes réalisées avec le soutien de l'Agence Intergouvernementale de la Francophonie

Abstract (EN)

BIRTH is a three-year EU funded project within the MEDIA plus programme that was launched January 2003. It unites the following partners: BBC (GB), Joanneum Research Graz (A), Sound and Vision (NL), Noterik Multimedia (NL), ORF (A), RTBF (B) and Hagenberg University of Applied Science (A)

The main goal is to set up a web portal for the exchange of digitised audiovisual material. The project focuses on the early days of television in Europe. Video content will be stored on streaming servers in the various participating countries following a decentralised architecture. The content provided is not limited to moving images of early television programs, in addition it also includes digitised schedules, stills, statistical figures and much more.

Keywords: Cultural Heritage, Digital repository, Video, Metadata alignment, online exchange, DRM, Television Studies.

Zusammenfassung (DE)

BIRTH ist ein dreijähriges EU-Projekt im Rahmen des *MEDIA plus Programms* und läuft seit Jänner 2003. Folgende Partner bilden das Projektkonsortium: BBC (British Broadcasting Corporation, GB), Joanneum Research Graz (A), Sound and Vision (NL), Noterik Multimedia (NL), ORF (Österreichischer Rundfunk, A) und RTBF (Radio-Télévision Belge de la Communauté Française, B).

Hauptziel des Projektes ist die Errichtung eines Webportals zum Austausch von digitalisiertem audiovisuellem Material. Inhaltlich konzentriert man sich dabei auf die Frühgeschichte des Fernsehens in Europa. Der Videocontent wird dezentral auf verschiedenen Servern der beteiligten Partner gespeichert und ist über ein gemeinsames Interface zugänglich. Das audiovisuelle Angebot umfasst neben den Videofiles von frühen TV Programmen, auch Fotos, digitalisierte Programmabläufe, Statistiken u. v. a. Der vorliegende Artikel umreißt die Grundidee hinter BIRTH, das erwartete Ergebnis des Projekts, die Strategie des multilingualen Zugangs beim Projekt und beschreibt das so genannte „Contribution Tool“, welches den Metadatenimport steuert.

Schlüsselwörter: Kulturelles Erbe, Digitaler Speicher, Video, Metadatenstruktur, Online Austausch, Digitales Rechtemanagement, Fernsehgeschichte

Résumé (F)

BIRTH est un projet financé pour trois ans par l'Union Européenne depuis janvier 2003, dans le cadre du programme MEDIA Plus. La BBC (RU), Joanneum Research Graz (A), Sound and Vision (PB), Noterik Mulitmedia (PB), ORF (A) et la RTBF (B) se sont rassemblés pour créer un portail d'échange de documents audiovisuels dédié aux premières années de la télévision en Europe. Le site BIRTH réunira des vidéos stockées sur les serveurs de *streaming* des partenaires selon une architecture décentralisée, ainsi que des textes scientifiques sur le début de la TV européenn, ou d'autres documents d'étude comme les programmes de télévisions en format PDF, des statistiques, des photos, etc. Mon article décrit la structure du portail, les stratégies multilingues et le système de contribution et sa gestion des métadonnées.

Mots clés: Patrimoine culturel audiovisuel, Structure des métadonnées, Échanges en ligne, Gestion des droits audiovisuels en ligne, Histoire de la télévision.

I. Introduction

Audiovisual material has been a valuable resource for the humanities for ages. In 1951, British academic Andrew Buchanan wrote: “The function of film … is to add to man’s sense of reality by adding the faculty of sight to the abstract apprehension. To see the things of which we learn completes, or almost completes, our study. The eye can often teach that which no words can convey”.(Buchanan 1951) However, due to various reasons, access to audiovisual material has been difficult in the past and was too often a time consuming process.

Luckily, as content suppliers are investing large efforts to make their material accessible in a more flexible way this situation is changing. One of the constructive side-effects concerns the shift in the way archive material is used as a source for academic studies. Digitisation and web-technology are keywords in this new situation where material is embedded in innovative learning environments that allow the integration of hyperlinked information systems, databases, interactive learning resources, communication tools, publication tools and tools to support collaborative and individualised learning.

Audiovisual archives are beginning to harvest the added value that the combination of new media and cultural heritage brings. A large-scale project, funded by the MEDIA programme of the European Commission is showcasing these new opportunities. Within the BIRTH project, major archives across Europe are building a portal to grant seamless access to a collection of meaningful and interrelated content to a wide user group.

Figure 1: State Secretary for Education and Science Joseph Cals during the very first Dutch television broadcast in 1951. © Sound and Vision

II. The road to continuous access to audiovisual content

Gaining access to audiovisual material for scientific research or for teaching purposes has been a difficult task for everyone until recently. The material was stored in archives that were more or less unwilling or able to match the needs of external users. Viewing capacities were very limited and mostly only available at the archives premises, due to the various formats used. Copying video sequences to widespread formats like VHS was causing costs schools or universities were very often not able to bear.
Copyright issues made it difficult to show the video material publicly. Furthermore, due to the various legislation's in different countries, international exchange of audiovisual material in a large scale was hardly possible. From a scientific point of view the audiovisual material often lacked a specific relevance as its documentation was often poor and it was hard to check the original sources: Knowledge indispensable for any scientific source. As a result of all these obstacles, the scientific community itself has only begun in recent years to develop new methodological approaches to audiovisual material and this process is not yet over.

From the mid-nineties new techniques emerged that were to change this situation for the better. Digitization of analogue assets and emerging network technology made it possible to distribute the holdings in the archives in a more flexible way. Material was encoded in low resolution quality, stored on servers and made accessible for previewing. A much cited paradigm in the broadcast environment expresses this new situation, where the audiovisual material and the catalogue description are stored digitally; content = essence + metadata. On the server both the essence (the material itself) and metadata are stored. This paradigm makes clear that without the accompanying metadata, the essence is useless. Connected client PC's enable remote access. This new situation results in various advantages. To name only a few: the digital archive is constantly accessible and is more than previously linked to the rest of the production process. (Borgman 2000, Manovic 2002)

Ongoing reductions of costs in specialized hard- and software and emergence of standards will cumulate in a situation where the whole chain from production to distribution of various media sources (in high resolution) will become digital. All this is still in a continuous flux, but already the production and transmission of television programmes is a largely digitized process. Specialized so-called Media Asset Management Systems are winning ground and

play a major role in transforming the archiving process and the creation of the infrastructure required for new media. (Digital 2002)

These changes in back-office processes also have their effect on end-user services. Already the content from the archives are delivered on-demand using streaming media technology, but this is just the beginning. Groundbreaking new services will be offered to a wide range of end-users: ranging from the general public wanting to look at yesterday's sports event, to academics studying the way media was employed for politics over the past decades. Without question, audiovisual content from the archives will be used more and more as historical source for research. Broadcast archives across Europe earmarked funds for innovative projects that will provide the humanities with unprecedented access to this important part of our cultural heritage. The European funded BIRTH project, is an example where several techniques are combined in order build an oneline service for Europe's audiovisual heritage.

III. The European funded BIRTH project

BIRTH (**http://www.birth-of-tv.org**) is a three-year EU funded project within the MEDIA-PLUS Pilot Programme that was launched January 2003. It unites the following partners: BBC (GB), Joanneum Research Graz (A), Sound and Vision (NL), Noterik Multimedia (NL), ORF (A), RTBF (B) and Hagenberg University of Applied Science (A).

The main project goal is to set up a web portal for the exchange of digitized audiovisual material. The project focuses on the early days of television in Europe, as many of the broadcasters have already celebrated or are going to celebrate their 50th anniversary soon. Video content will be stored on streaming servers in the various participating countries following a decentralized architecture. The BIRTH portal

Figure 2: www.birth-of-tv.org

provides thorough context information around the original sources and presents it in a uniform way, accessible via modern mass media structures to the professional user, the scientific community and the general public. In the first stages of the project, the consortium conducted an extensive user requirements study, in order to map the expectations of end users and to find out which delivery models would fit best.

The content provided in the portal is not limited to moving images of early television programs. In addition it also offers digitized schedules, stills, statistical figures and much more. The results of the BIRTH project also strongly contribute to the preservation of the European cultural heritage, as it showcases the treasures stored in the archives and provides guidelines on best practices concerning preservation strategies.

Particular attention is given to provide language independent search possibilities and to offer possibilities to compare the different development paths in several countries over Europe. In the next chapter, the different ways BIRTH offers language independent access is outlined.

In the first months of the project, an extended user survey and interviews were conducted in order to establish the user group's' preferences on the services offered through BIRTH. Apart from mapping the more or less technical requirements of the portal, the survey made it clear that there is indeed a demand for a website on early TV. The aims of the project and the content being made available from broadcaster partners will make it unique.
The majority of the content - film, printed media and photographs have not been seen widely since original transmission. The bringing together of European archive content in a searchable form in a number of languages makes it a powerful resource for all user groups.

Functionalities offered

Several (existing) technologies and content from the partners are capitalized to provide these functionalities. First of all, existing data in the catalogues was transformed in accordance to the common metadata model defined by the Dublin Core Initiative, and extended with MPEG7 attributes to fit our needs. Secondly, to transcend language boundaries of the heterogeneous content, a multilingual thesaurus was made, using a specialised thesaurus tool. Other techniques that are used deal with digitizing analogue sources, automatic generation of keyframes and mechanisms to support rights administration.

Finally, a peer to peer distribution architecture is put into place in order to meet the needs of the Business to Business profile. Within this architecture, an authentication system validates user permissions (e.g. by requesting payment confirmation) and informs the data source where the requested file is physically located. The one-time key sent to the user as a proof of proper authentication cannot be reused - it is automatically invalidated after the download is finished. No proxy server is involved in the file transfer. The download process is supervised and statistics such as connection characteristics and machine load are gathered and stored for further processing in a central database.

Handling copyrights

Rights issues for (Global) web access to legacy audiovisual content has been problematic. Contracts and rights relating to performances, music and programme distribution were drawn up before the advent of computers, the internet or the digital age. Added to this situation is the diverse legislative landscape in each country relating to copyright and "broadcasting" on the internet. Mostly these rights are undefined and fragmented. As such access to legacy material for internet use has been restrained. Leading to frustration for programme makers, academics and the public who find themselves excluded from really useful content.

However, the situation is slowly changing. The pressure to resolve the issue of internet access has been growing from a number of areas both cultural and financial. Computers and internet costs has been coming down leading to more and more households online. – Broadcasters in general have moved heavily into digital channels many of which rely on archival content. Some state broadcasters have begun to open up their archives as part of cultural and historical celebrations.

To support these growing demands on access the pressure for clarification on the legal definitions has surfaced on an international level. Leading to a number of joint initiatives to establish a legal framework in which interested parties can operate. Examples of this include the "Santiago Agreement" in 2001. (**http://www.gema.de/engl/communication/news/n163/santiago.shtml**). It is hoped that in the near future a clear legal structure will emerge allowing users to easily discover and negotiate rights and reuse efficiently.

Future aims and collaboration

Bringing these technologies together resulted in a website where in the first instance of the early days of television is documented, but this will only be the first step. When a workflow for selecting, cataloguing, uploading and ordering is tested and finalized the BIRTH portal will be extended in time and collections, growing to an ever important source for academic research and the enjoyment of the general public. Already the consortium is in contact with content owners to add the highlights of their collection to the portal.

BIRTH will offer ways for visitors to interact with the contents: to add their own articles and react to contributions from others. The consortium strongly believes that the information provided by professionals should be regarded as valuable information. At the other end of the spectrum, all visitors are potentially experts and should be offered ways to contribute to the information resources. It could stimulate a potentially immense community of avid voluntary cataloguers who will find surprising ways to mine and exploit the treasure trove of material offered. The BBC's Creative Archive initiative (**http://www.bbc.co.uk/pressoffice/pressreleases/stories/2004/05_may/26/creative_archive.shtml**) is a good example of projects already exploring the possibilities of what is often referred to a the 'living archive'. (Mulder 2003)

In order to capitalise the work done during the project, BIRTH conformed to international standards and teamed up with international bodies dealing with the promotion of digital cultural heritage:

- The descriptive metadata format is compliant with both Dublin Core and MPEG-7. For file exchange, digital material was encoded following ISO MPEG standards for moving images and TIFF for stills. Common proprietary formats are used to facilitate streaming media.
- The project linked up with the Open Archives Forum, PrestoSpace, the Moving Image Collection Project of AMIA, European Cultural Heritage Online and various national digitisation initiatives. As these umbrella projects also work with the standards mentioned earlier, the content for BIRTH will be accessible across different platforms.
- The organizations of BIRTH are members of international bodies, such as EBU, FIAT-IFTA, DELOS and DigiCULT. This is an enormous advantage to gain critical mass for the exploitation of the projects outcomes be it lessons learned, the allocation of content or techniques developed.

IV. Providing multilingual access

Contents from several countries can be found on the portal. Due to the different languages used, users need to be assisted crossing the obvious language borders. This was one of the major challenges to be solved.

The BIRTH portal will provide several ways to navigate through the content in a multilingual way. They are listed in this chapter and can be summarised as follows:

- Translation of key elements
- Powerful use of multilingual taxonomies
- Television history timeline
- Knowledge based articles with hyperlinks
- Language independent elements

Translations of key elements

The partners in BIRTH project each have their own proprietary metadata schema, connected to their legacy catalogue. As we needed a common model to map the various elements to a common –searchable- model, we chose the Qualified Dublin Core metadata element set. Several elements were added to the pre-defined set of metadata elements. Some of these new fields will be translated into English and thus act as a uniform access point.
Primarily they include the elements “English title” and “English abstract”. The user survey showed that by far, most users can understand English. By translating key elements in English, users will be able to understand the items they have found and put them in meaningful context.

Powerful use of multilingual taxonomies

There are several ways of entering a search query and navigating through the BIRTH collection. If non-native English users want to use the advanced search to search one of the translated elements (i.e. <description_english>) they will have to translate their query first. This is not ideal, as we would like to offer users the possibility to enter their query in their own language (Dutch, English, French, German,) and subsequently retrieve items from all the collections on the Portal. For structured searching, BIRTH will offer users several tree-structured taxonomies that are easy to navigate.
The first time a user visits the BIRTH portal he or she can set the language in which they would like the front end to be delivered.

The taxonomies are automatically set to the language chosen on the homepage of the portal. Another option might be to automatically detect the user's IP address, determine the country where this address is registered and set the language accordingly. So if a user from Germany selects the preferred language to German, the system will offer him German keywords as entry points, German interfaces and the English abstracts in case the system returns results from non-German archives.

BIRTH will support taxonomy axes for Genres, Keywords and both Heritage and Geographical classification. For the values in these axes, BIRTH studied the possibilities that are already available. For example, for the genre element, BIRTH conforms to the Escort standard, set by the EBU and already in use at several major institutions dealing with audiovisual heritage.

Concerning the keywords, BIRTH will offer the users of the portal the possibility to browse a tree-structured list of thematic keywords, available in several languages. The consortium studied several projects and products providing a multilingual taxonomy to be adopted and used for the project. Experts from the Twente University in linguistics were contacted to make sure no major solutions were overlooked. Eventually, the consortium decided to use the so-called IPTC thesaurus (**http://www.iptc.org/metadata**/) and add television-oriented keywords to the ones that are already part of the IPTC.

Scope notes will accompany all terms, so users understand what is meant by these terms. It will be possible to add new terms to the taxonomy. One of the partners will be responsible for managing this updating process so it will not result in a uncontrolled proliferation of terms.

It is planned to broaden the scope and collections of BIRTH after the formal end of the project, so it will include also more recent television programmes and focus on other events that are beyond the scope of the initial phase. At that point, it would be advisable to look at extending the taxonomy to an extended thesaurus, structured to show relationships between terms and concepts.

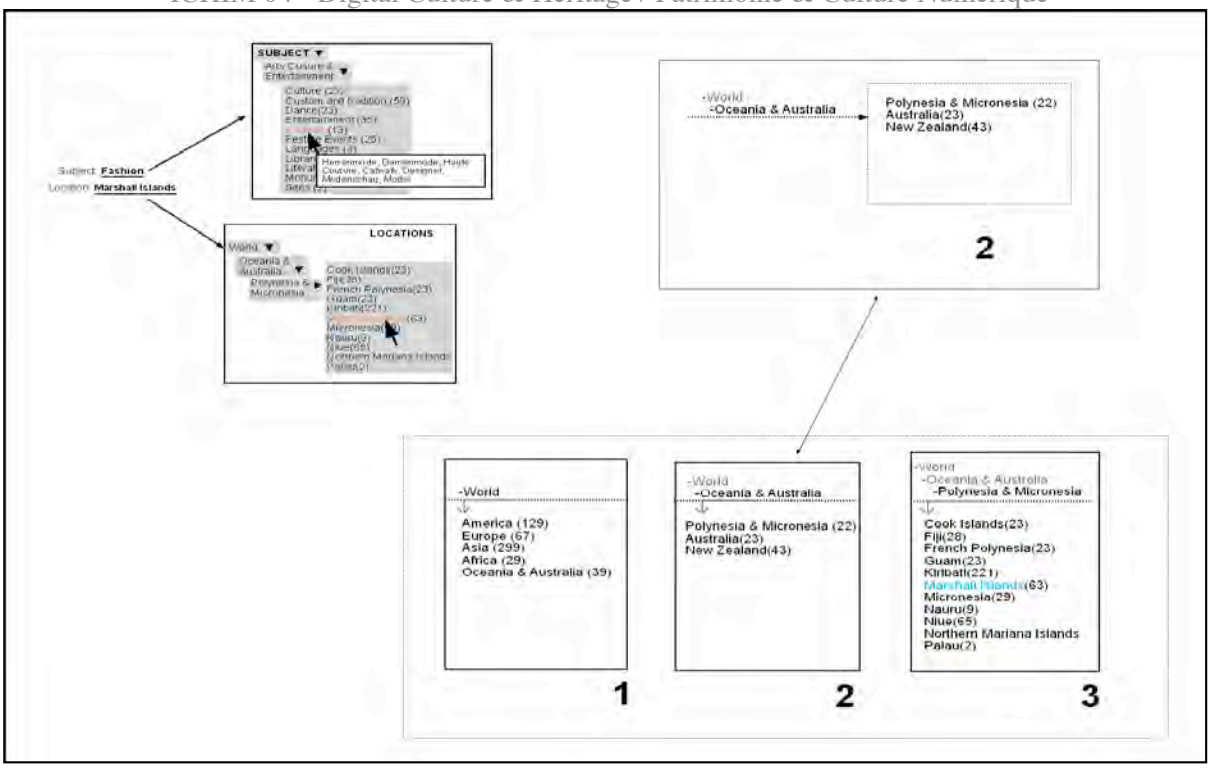

Figure 3: Structure of the thesaurus

Two other controlled vocabularies will be supported. These concern heritage classification and location. As for the heritage classification, a listbox will be included, listing the technical status of the item, i.e.: "preserved", "not preserved", "audio missing or fragmented" etc. This list will also be translated in several languages and in the search screen, users are able to search or exclude items based on these properties by selecting these properties from a list.

As for the classification of locations, Dublin Core uses the <coverage_spatial> element to describe the locations connected the item, i.e. the scene of action of the television programme. This element will be exported directly from the archive catalogue listing, and will only be available in the archive's native language. In addition, BIRTH will add the <coverage_spatial_BIRTH> element, which will basically contain the same information, but here; a controlled vocabulary will be used. There are already lists available, most notably (ISO 3166, see also: **http://www.iso.ch/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html**). As this broadly implemented standard is already available in different languages, BIRTH chose to use this.

These two elements seem to overlap, but this is not the case, as the granularity of the ISO standard is not as specific as the lists in the existing catalogues. The ISO standard only lists

continents, countries, regions and some major cities; the catalogues will typically also list small towns.

The <coverage_spatial_BIRTH> can make dynamic visualizations possible. There are already XML lists available in the public domain linking the listings in the ISO standard to coordinates. Using relatively simple web languages it is possible to generate a clickable map of (for example) Europe – Netherlands – Amsterdam to zoom in to the particular city he will wants to use as a query. The French "Aquitaine Patrimoines" project [project (**http://ajlsm-sdx.hopto.org/sdx-bnsa/pa-portail/**) showcases how this might look.

Handling the translation of well over a thousand terms and the export of these terms in machine readable XML needed a specialised application. Joanneum Research developed the so-called ThesauriX AdminTool; with comfortable interfaces allowing the editing of the thesaurus content in a hierarchical order.
As a matter of course, the ThesauriX AdminTool enables the user to create and manage one or more thesaurus databases by providing functionality like hierarchical ordering of terms, multi-language support, adding relations (e.g. synonym) between terms and more.
Beside the creating and editing of multiple new terms and translations with ease the ThesauriX AdminTool provides the ability to synchronise several thesaurus installations to a master database. This offers the archive partners of the BIRTH project to translate and extend the thesaurus in a simple way - without manually distributing their new terms to the other partners - this is done automatically.

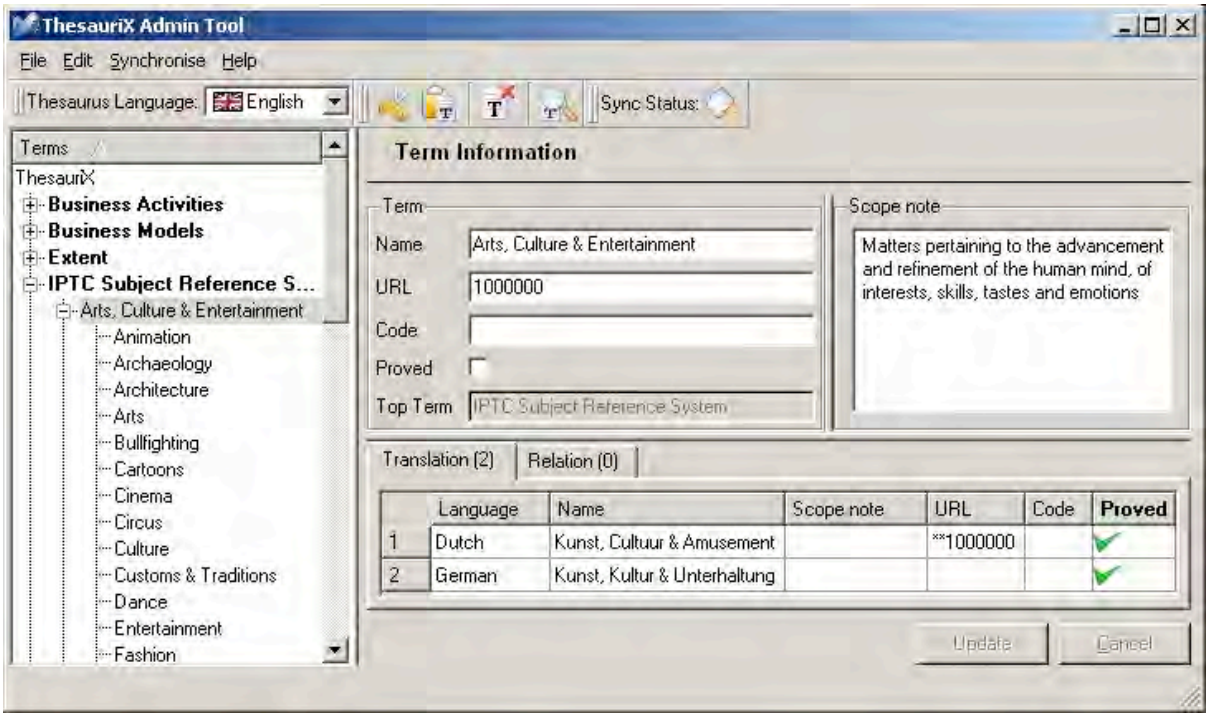

Figure 4: The thesaurus administration tool

Knowledge based articles with contextualised hyperlinks.

The fourth way to access the content in a more or less language independent way are through the knowledge based articles (KBA's). As by far most potential users of the portal understand the English language, these articles and dates (with contextualised links to various content from all participating archives) will also serve as a starting point to access the content. The topics vary and include: government influence, controversial TV and censorship, the early stars, Eurovision broadcasts, major technical developments and highlighted genres.

Language independent elements

Some of the Dublin Core elements are by nature language independent. Most notably this concerns the <creator> field (as person names are usually spelled the same way in different countries) and the <date> and <coverage.temporal> elements.

The date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601and follows the YYYY-MM-DD format. This information will be filled into the <date> Dublin Core element.

Listed below is an example how this might look, taken from the Amicitia project (http://**www.amicitia-project.de**). The sliders on the left and right indicate the time span within which the user wants to query the system. Especially when this is combined with a full text query results in a powerful entry point to the collection.

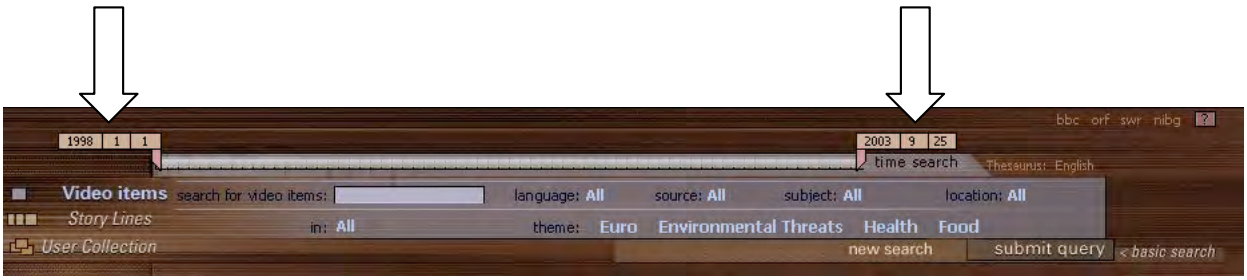

Figure 5: Timeline search

V. Managing content: the Contribution Tool

Obviously, the BIRTH portal can only provide its functionality with content provided by the archive partners. Of course, all the content, which is essence and the corresponding metadata of the archive partners need to be transferred to the BIRTH portal in an easy and effective way – for the archive partners as well as for the portal managers. For these reasons a so called Contribution Tool was developed by a student group of the Hagenberg University of Applied Science (for Software Engineering) with support from the BIRTH team at Joanneum Research. The tool was designed as a consistent entry point for the archives to upload their content.

Generally speaking, the Contribution Tool is a client application installed at every archive partner. Its basic functionality is to add essence, annotate corresponding metadata and make it then available at the portal. Metadata and images (more details about images in the next paragraph) are stored directly at the portal. For performance and legal issues, the video material presented on the BIRTH portal is stored on national servers at every archive partner.

The Contribution Tool supports manifold kinds of essence: Audiovisual material, still images, audio, knowledge-based articles and generic documents (e.g. PDF). All this essence is annotated with the same metadata model to handle it in a consistent way. To make use of already existing metadata, the tool supports advanced conversion mechanisms. As noted above, the archives already have essence and metadata in their legacy catalogues. The

Contribution Tool is able to import metadata in various formats, perform mapping processes and output uniform XML files, conforming to the BIRTH metadata model.

Additional to the functionalities described above, several advanced content analysis features are available for audiovisual essence. To present a quick overview of a video for the portal user, so-called key-frames can be generated automatically. These “screenshots” of the videos can be displayed in the search result list to enhance the usability for the user. Furthermore, the tool supports the transcoding of MPEG formats into different proprietary streaming media formats.

Besides the handling of archive essence, the previously mentioned Knowledge Based Articles can be created and uploaded using templates available in the Contribution Tool. Currently the tool provides a basic editor to enter the text of the KBA. After uploading the KBA (and its metadata) to the portal, the KBA can be enriched with images and additional formatting.

To have a high precision and recall on search results at the portal, the usage of a multi-lingual thesaurus was identified as essential from the beginning of the BIRTH project. To support this aim the thesaurus is integrated in the entire workflow, from the beginning of the annotation until the search process on the portal. The Contribution Tool integrates the thesaurus by letting the user select thesaurus terms for several elements of the metadata. The actual management of the terms (translations, adding new entries etc.) is carried out by the ThesauriX tool that as described earlier. The figure below summarizes the workflow of the BIRTH project.

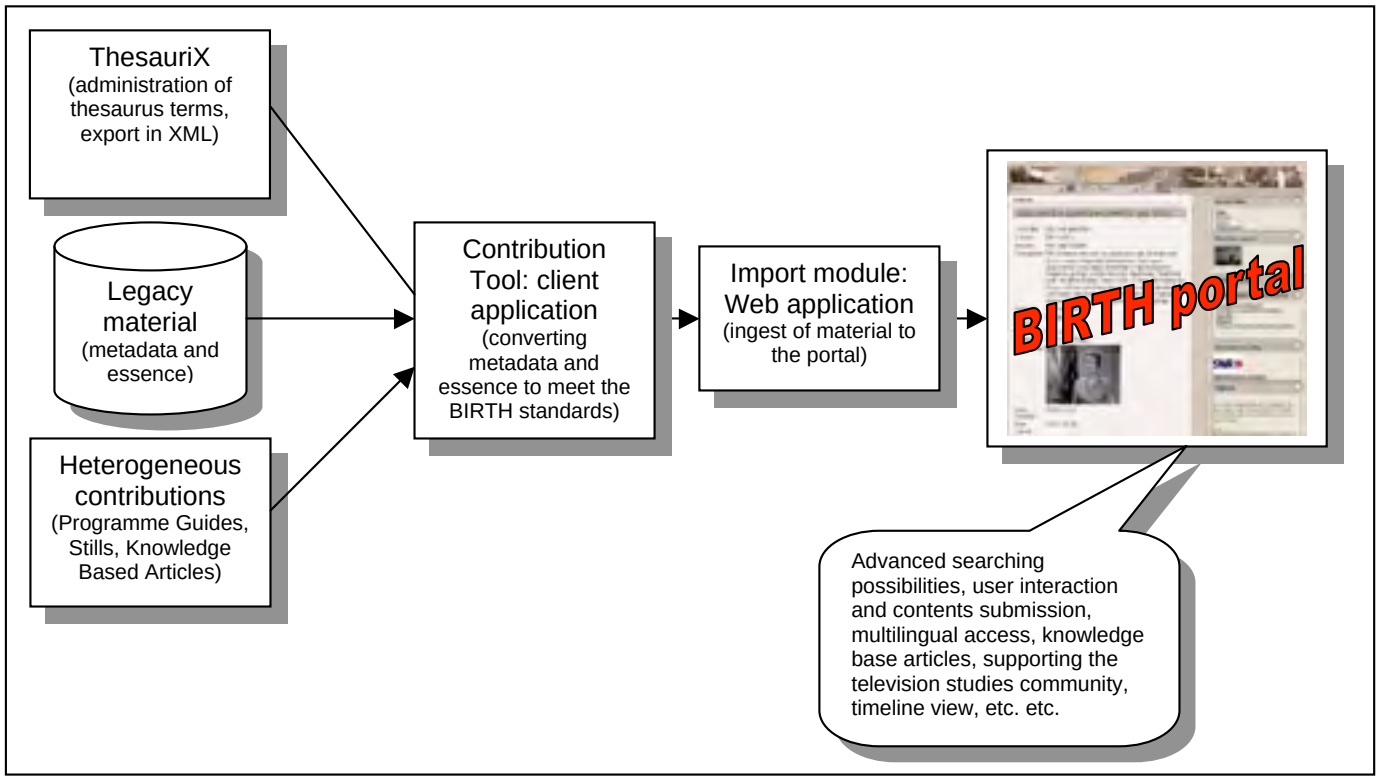

Figure 6: Workflow BIRTH: presenting various content on a web portal

VI. Conclusions

The BIRTH project has been running now for over one and a half years. Within this time, it managed an in-depth user requirements analyses and established a robust and flexible platform for the on-line delivery and exchange of audiovisual material and context information. The publicly accessible portal will be launched in the fall of 2004. Already, BIRTH has set up an impressive network of academics contributing on various ways to the portal, be it by writing articles or providing input for content selection. Apart from expanding the links with academia, BIRTH is also seeking collaboration with content providers to expand the portal with new collections of audiovisual material. A cookbook will be provided for those interested in joining BIRTH, outlining the workflow for publishing material online. In the next stages of the project, BIRTH will also focus on how to tackle existing copyright issues, in order to make it easier to make contributions and to support a business-to-business service, including the distribution of high-resolution material for professional reuse. The solutions can range from using Creative Commons (**http://creativecommons.org**) licenses to deploying advanced Digital Rights Management systems.

Being the premier website on television history, supported by organisations largely responsible for recording history, the BIRTH portal has an enormous potential. It will have an impact being a rich repository of digital content, a promotion platform for member institutions, a best practice example of pan-European collaboration and a showcase of how the use of advanced web technology can increase the accessibility of and interaction with our cultural heritage.

References

Borgman, Christine L. (2000) From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. London: MIT Press, 2000.
Buchanan, Andrew. (1951) The film in Education. London: Phoenix House Limited.
Digital Asset Management Systems for the Cultural and Scientific Heritage Sector (2002). DigiCULT thematic issue 2.
http://www.digicult.info/downloads/thematic_issue_2_021204_low_resolution.pdf
Manovic, Lev (2001) The Language of New Media. London: The MIT Press.
Mulder, Arjan (ed) (2003) Information is Alive: Art and Theory on Archiving and Retrieving Data. Rotterdam, V2.