

ARCA: ACCESSING FEDERATED MULTIMEDIA CONTENT

Francisco Cruz Argudo¹, José María Fontanillo², Nicolás Ibáñez Moreno³, Diego R. Lopez⁴

¹ Universidad Carlos III de Madrid, paco@di.uc3m.es

² RedIRIS,jmaria.fontanillo@rediris.es

³ Universidad Carlos III de Madrid, nibanez@di.uc3m.es

⁴ RedIRIS, diego.lopez@rediris.es

1 EXECUTIVE SUMMARY

More and more higher-education institutions are deploying audiovisual and IT equipment to provide network-based multimedia service and generating, as a consequence, a great deal of multimedia contents from classes, conferences and workshops. Most of these contents remain circumscribed within the university that generates them, without reaching the rest of the research and academic community.

ARCA is a platform intended to solve this problem, easing the access to multimedia contents, letting users know which live events are accessible and which contents are available on demand. ARCA is an open solution that we are confident will be enhanced by the collaboration of its user community. Several universities have already joined the ARCA developer community.

The ARCA platform works by aggregating (federating) content descriptions built using RSS. Each participating institution must keep an updated RSS feed announcing its live events and multimedia contents. The ARCA aggregator periodically loads and processes these RSS (actually, XML) documents and incorporates them into a database for further access. The ARCA portal organises the collected metadata and provides simple and intuitive browsing and searching mechanisms, and users can access the multimedia descriptions in a fast and direct way.

ARCA is intended to provide any student, faculty member or researcher with relevant and updated information about live events and media contents available at the participant institutions. Value is added to these media contents, now accessible through an intuitive and simple interface. ARCA references can be easily incorporated into Web 2.0 constructs (such as blogs) using a facility provided by the portal. This allows for the creation of new access services, directly controlled by users, both as producers and consumers of media contents.

2 INTRODUCTION

The University Carlos III of Madrid (UC3M) decided some years ago to make a bold move into the usage of new technologies in teaching support. Technologies such as videoconferencing, synchronous collaboration software, video streaming, and instant messaging are regular and enthusiastically used by faculty and students. When these tools become commonplace, the need for new services and applications related to them is a natural consequence. In this context, the Multimedia Area of the UC3M IT Services started the development of applications able to catalogue, classify, store, publish and distribute the multimedia contents produced by the different teaching and research activities. Therefore, ARCA was born in an environment of new network multimedia service deployment

ARCA (“ark” in Spanish, standing for the Spanish terms for *RSS Aggregator for the Academic Community*) was devised as a space where data about teaching and research multimedia contents available through the network could be retrieved in an easy and intuitive way. All the solutions that the ARCA team reviewed were based on highly centralised specific Web portals (like the TF-Netcast Portal, available at <http://prenosy.cesnet.cz/index.php>), where a specific authentication (username + password) is required to edit the element descriptions (live events only) managed there. From the very first moment, the ARCA team wanted to provide mechanisms as decoupled and decentralised as possible. Mainly for two reasons:

- ◆ Minimize the maintenance burden over the portal administrators.
- ◆ Ease the update of the portal by the participants, allowing them to locally decide what is shown at the portal.

Initially the portal was only intended for live event announcements, but we soon realized the need and relevance of managing more information types: Why not design a portal able to deal not only with live events but also with multimedia files available at the institutions? It appeared evident that in this case the system could only manage metadata related to the latter ones, as it would be completely unrealistic to dimension a central storage system for the whole multimedia contents of the participants (in the style of solutions like YouTube). Therefore, it was necessary to find a way to represent all the information corresponding to the contents of the institutional servers participating in the project. This process allowed us to define three kinds of information objects (that will be further discussed below): live event, video on demand, and podcast/vodcast.

Another important decision was how to describe each one of the different elements going to be announced and accessed through the portal. In a first moment, the SDP and SPDng (Session Description Protocol, see RFC2327 and draft-ietf-mmusic-sdpng-04) formats were considered, but promptly discarded because they did not fulfil the functional goals. We finally selected RSS 2.0, usually employed for podcast and news syndication, since it was the standard best fitted to the system requirements.

At that point we continued defining the information model (and therefore the tags) that the different types of description to be used by the system should contain. Obviously, the data related to a live event are not the same that those pertaining a stored video or a podcast. This led us to employ the concept of independent *namespaces* supported by RSS. Three of these namespaces were initially selected:: *itunes rss*, *yahoo media rss* and *google media base*. Each one of them provides the possibility of modelling (as we will show later) information relevant for each of the different supported content types.

3 ARCHITECTURE

ARCA has been developed on:

- Apache2 as Web server.
- PHP 5 as application server.
- MySQL 5 as database engine.
- Linux as operating system.

The ARCA platform works by aggregating (federating) content descriptions built using RSS. Each participating institution must keep an updated RSS feed announcing its live events and multimedia contents. This feed must comply with RSS 2.0 and use the above-mentioned namespaces. The ARCA aggregator periodically loads and processes these RSS (actually, XML) documents and incorporates them into a database for further access. Metadata transfer is achieved by means of HTTP in the ARCA version described in this paper.

The ARCA portal organises the collected metadata and provides simple and intuitive browsing and searching mechanisms, and users can access the multimedia descriptions in a fast and direct way. Furthermore, any query can be transformed into an RSS feed, allowing users to subscribe to them through any RSS aggregator.

In the process of incorporating feeds into the database, the system creates a specific *channel* for those feeds with a common tag <title> in their definition, associating it with a given institution. This way, all the announced items are grouped into channels, corresponding each one of these channels to a single feed from an institution, or several feeds from the same one. The following example illustrates how a channel element must be represented in a RSS feed:

```
<channel>
    <title>Universidad Carlos III de Madrid</title>
    <link>http://audiovisuales.uc3m.es</link>
    <description>
        Vídeos y retransmisiones de diversos eventos de la
        Universidad Carlos III de Madrid
    </description>
    <language>es</language>
    <copyright>Universidad Carlos III de Madrid</copyright>
    <managingEditor>audiovisuales@uc3m.es</managingEditor>
    <webMaster>audiovisuales@uc3m.es</webMaster>
    <pubDate>Wed, 06 Jun 2007 00:00:00 +0200</pubDate>
    <lastBuildDate>Wed, 06 Jun 2007 00:00:00 +0200</lastBuildDate>
    <category domain="tipo institución">universidad</category>
    <generator>Script PHP</generator>
    <docs>http://blogs.law.harvard.edu/tech/rss</docs>
    <ttl>1440</ttl>
    <image>
        <url>http://marge.uc3m.es/images/4tintas.jpg</url>
        <title>Universidad Carlos III de Madrid</title>
        <link>http://audiovisuales.uc3m.es</link>
    </image>
</channel>
```

It is worth noting that the tag <image> allows the ARCA portal to show a certain image (typically, an institutional logo) associated to the channel and the events announced for it. This has become a requirement for including a channel in the system.

3.1 Live Events

These are organised according to a calendar that can be browsed using day, week or month views. The month view offers the programmed events, their description and the channel where they are announced. From this view it is possible to go to a particular day clicking on it, or directly access the details regarding a certain event.

For a live event description to be included in ARCA, it is mandatory that its RSS description contains its start and end times by means of the <g:event_range> tag, where the formats of the dates and times must follow ISO8601. It is also possible to indicate the physical place for the events using <g:location>. The ARCA portal offers different kind of searches for this type of events, according to different search parameters.

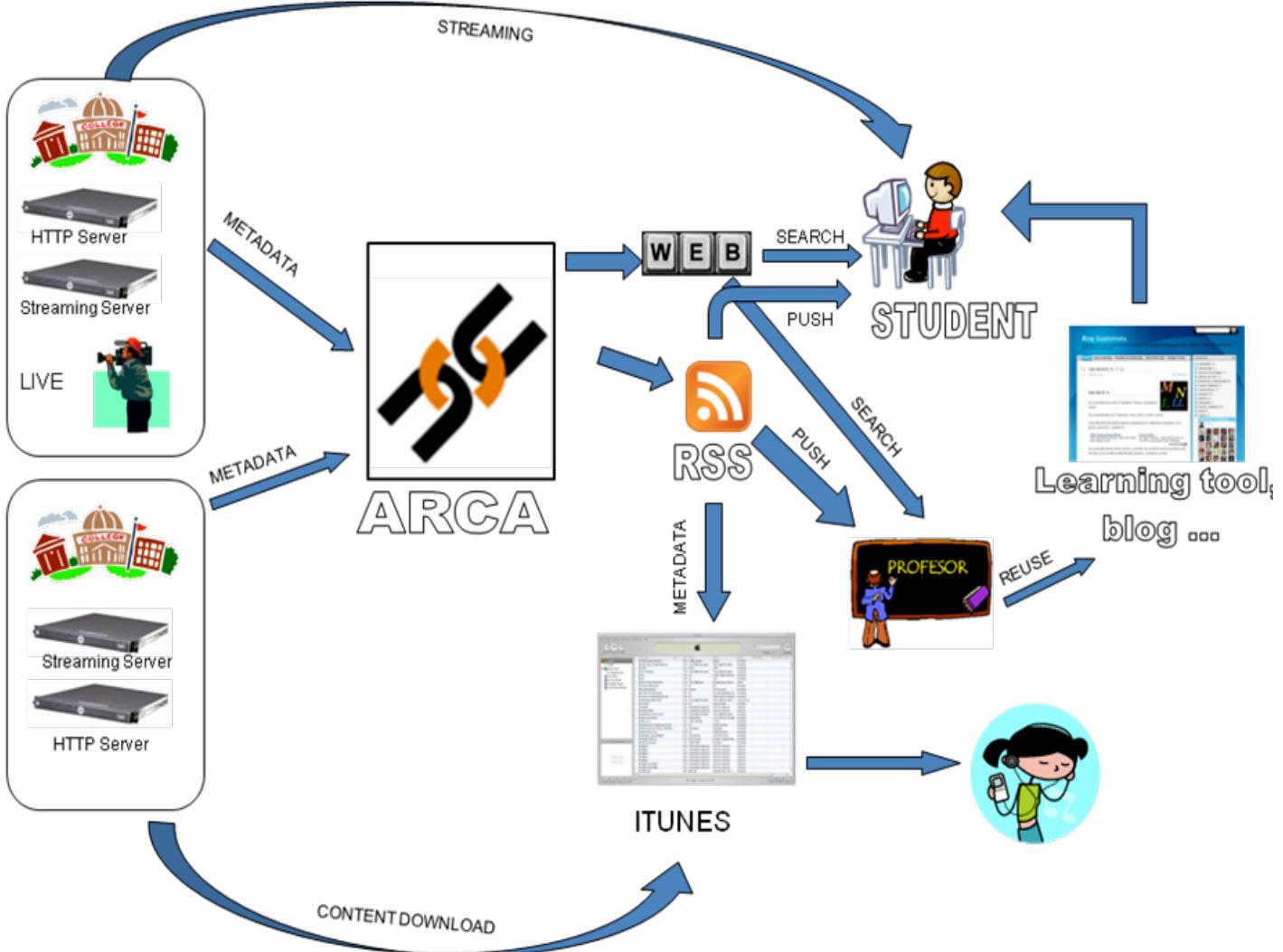


Figure 1 - ARCA Usage scenarios

3.2 Video on Demand

For the management of the contents available on demand, ARCA provides advanced query mechanisms such as:

- Content search filters, on the channel, category or type of the description.
- Search by keywords: ARCA can query its database for the occurrence of a set of words. The search engine is insensitive to uppercase/lowercase or tildes, and can match either all searched words or the occurrence of any of them.
- Both search techniques (filters, keywords) can be combined to improve the results.
- The results can be ordered according to different criteria: title, channel, date of production...

When displaying an item, ARCA can give comprehensive information on the announced media element: still images (up to 11), video and audio formats, language, credits (people involved in its production and role of each one), etc..

The system is also able to internally play from the portal itself, practically any video format. This feature requires the user to have the required players and plug-ins already installed.

In principle, advertising this kind of content only requires to provide a link to the video using the `<link>` tag in the item description. However, you may want to use that tag to point to a link related to the video. The correct use is to always provide these three labels: `<link>`, `<media:content>` and `<g:publish_date>`. Using the second we can also indicate the technical characteristics of the content and indicate the date of production or recording. Finally, the item must be classified according to the existing categories using the `<category>` tag. As an example:

```

<category domain="directriz">Ciencias Sociales y Jurídicas</category>
<media:content
  url="http://www.foo.com/movie.mov"
  fileSize="12216320" type="video/quicktime"
  medium="video" isDefault="true" expression="full"
  bitrate="128" framerate="25"
  samplingrate="44.1" channels="2" duration="185" height="200" width="300"
  lang="en" />
<g:publish_date>2006-01-23</g:publish_date>

```

It is also recommended to use the <g:image> and <media:thumbnail> tags in order to enable the ARCA portal to show images related to the item contents.

Furthermore, the <media:group> tag can be used to provide playing choices for a given item, like different formats, compression quality or languages. This example shows the usage of three different formats (quicktime, windows media and mpeg):

```

<media:group>
<media:content
  url="http://www.foo.com/movie.mov" fileSize="12216320"
  type="video/quicktime"
  medium="video" isDefault="true" expression="full" bitrate="400"
  framerate="25"
  samplingrate="44.1" channels="2" duration="185" height="200" width="300"
  lang="en" />
<media:content
  url="http://www.foo.com/movie.wmv" fileSize="14216312"
  type="video/x-ms-wmv"
  medium="video" expression="full" bitrate="400" framerate="25"
  samplingrate="44.1" channels="2" duration="185" height="200" width="300"
  lang="en" />
<media:content
  url="http://www.foo.com/movie.mpg" fileSize="98216112" type="video/mpeg2"
  medium="video" expression="full" bitrate="2000" framerate="25"
  samplingrate="44.1" channels="2" duration="185" height="200" width="300"
  lang="en" />
</media:group>

```

Currently ARCA supports mp4, wmv, mp3, qt, mpg, rm and 3gp as formats. Anyway, including support for new media formats is practically immediate.

3.3 Podcast/vodcast

The term podcast refers to the possibility of accessing audio/video contents on demand by means of portable devices in an asynchronous way. The requested content is downloaded from the distribution system and stored at the device for a later offline play.

ARCA allows the definition and management of these items for audio and video formats, using the **itunes** namespace. This way, the contents can be played by Apple iPods and the iTunes software. ARCA provides an enhancement over normal podcasts, since it can provide podcasts for almost any video or audio element, with independence of its format (mp4, wmv, mp3, qt, 3gp,...). Furthermore, it offers a richer semantics than the one provided by the **itunes** namespace. In practice, ARCA offers a “universal” podcast player, without format limitations. In other words, any media content that can be played through ARCA is suitable to become a podcast.

The screenshot shows the ARCA portal homepage. At the top, there's a search bar with the placeholder "Search" and a dropdown menu set to "English". Below the search bar is a logo for the Universidade de Vigo. The main content area features a banner for "I Jornadas Software Libre en la Enseñanza" with a cartoon penguin character. To the right of the banner are links to various academic categories: Health Sciences (200), Experimental Sciences (149), Social and Legal Science (647), Technical Sciences (1160), and Humanities (886). Below the banner, there are sections for "Today's events" (empty), "Next week" (calendar grid with Sat 17 highlighted in red), "Newest videos" (links to "Conferencia Concierto Fernando Argneta" and "Firma de convenio con la CAM"), and a "This month" calendar for May 2008.

Mon	Tue	Wed	Thu	Fri	Sat	Sun
			1	2	3	4
5	6	7	8	9	10	11
12	13	14 (1)	15	16	17	18
19	20 (1)	21	22	23	24	25
26	27	28	29	30 (5)	31	

Figure 2: Accessing the ARCA portal

4 NEW FEATURES IN THE CURRENT VERSION

- ◆ GEORSS: ARCA is able to manage this namespace, so it is possible to include geographical data on the recording place for on-demand contents, or the venue for live events. ARCA processes this information and connects to the Google Maps API to make a graphical display of the described location. The supported tag is <georss:point>, describing latitude and longitude coordinates: <georss:point>45.256 -71.92</georss:point>.
- ◆ OPML: OPML (Outline Processor Markup Language) is an XML format for outlines. Its most common usage is to aggregate different RSS sources. It consists of 4 basic tags: the <opml> root tag, <head> and <body> to identify the headers and actual contents of the outline, and <outline> for each of the outline lines. ARCA is able to manage OPML documents that include several RSS sources in a single feed. This way, an institution providing several feeds can point the ARCA collector to a single URL where an OPML document describing those feeds is available, and the collector will seamlessly retrieve them. The following example illustrates the usage of this construct.

```

<?xml version="1.0" encoding="UTF-8"?>
<opml version="2.0">
  <head>
    <title>Universidad Carlos III de Madrid</title>
    <ownerName>Universidad Carlos III de Madrid</ownerName>
    <ownerEmail>audiovisuales@uc3m.es</ownerEmail>
  </head>
  <body>
    <outline type ="rss"
      text="Universidad Carlos III de Madrid"
      xmlUrl="http://audiovisuales.uc3m.es/rss/vod2005.xml" />
    <outline type ="rss"
      text="Universidad Carlos III de Madrid"
      xmlUrl="http://audiovisuales.uc3m.es/rss/vod2006.xml" />
    <outline type ="rss"
      text="Universidad Carlos III de Madrid"
      xmlUrl="http://audiovisuales.uc3m.es/rss/vod2007.xml" />
  </body>
</opml>

```

- ◆ HTTPS: ARCA is now able to access feeds through HTTPS, thus enhancing the trust on the descriptions included in the system by means of a verification of the certificate provided by the RSS source. The trusted CAs are configurable. In HTTPS mode, when certificate validation fails the contents of the feed are not incorporated into the ARCA database.
- ◆ ARCA namespace: ARCA has defined its own namespace, since existing ones were not able to offer all the functionalities required by the portal new features. For example, there were no tags for dealing content series (see below). Since other foreseen functionalities will probably require additional tags, the availability of this namespace guarantees a smooth evolution.
- ◆ Series/Courses: In the new version, ARCA supports the grouping of video on demand contents by series or courses. This allows a better user experience when dealing with teaching materials, since the portal is now able to show integral references to related elements, instead of a simple list of videos. The <arca:course> tag is the mechanism to announce a course or series divided into sessions. This tag is used inside each item description and has the following child tags:
 - arca:title: course title, mandatory.
 - arca:description: course description, mandatory.
 - arca:image: link to an image for the course, mandatory.
 - arca:banner: link to a banner for the course, optional.
 - arca:order: order of the item inside the course, optional.
- ◆ Online translation: The ARCA portal has a module that permits the online translation of pages and messages, taking advantage of the same database engine that supports the basic ARCA functionality.
- ◆ Administration system. It groups a set of modules for the online management of the ARCA portal: classifications, members, feeds, user and languages.
- ◆ Support for module devices. ARCA Mobile is an alternative interface for mobile devices. It is a simplified version of the portal tailored to adapt to the lower graphics resolution available in mobile devices and providing access to the most relevant system services.
- ◆ ARCA intercommunication. The system is able to interconnect different ARCA portals, by means of the exchange of OPML and RSS documents, by simply selecting in the administrative interface which feeds are going to be shared.

5 HOW TO PARTICIPATE

ARCA is based on the federation of contents provided by the participating institutions, using the RSS 2.0 syndication format. In principle, any research or higher-education institution providing multimedia contents supported by the application can join the federation. To participate, the institution must provide one (or several) URL where the feeds are to be retrieved from. There are essentially two ways of organising the information:

- ◆ A single feed with all the contents, of whatever kind among those supported by the system.
- ◆ A feed for each one of the supported content types (live, VoD, podcast). In this case it is important to keep the same channel name for all the feeds, so the portal can identify them as belonging to the same institution.

It is recommended to use HTTPS when transferring these feeds, so the identity of the source can be assessed by the ARCA collector. It is possible to group several feeds by means of OPML.

For live events, the use of the GEORSS namespace provides a way for the portal to show the venue of the event on Google Maps. This way, the ARCA portal can be used to guide people willing to attend the event in the flesh.

The ARCA portal provides several RSS validators allowing feed administrators to check the correctness of their data before making them available. It is also worth noting that ARCA does not perform incremental updates, but full imports of the provided feeds, so it will only show the elements actually present in a given institutional feed at a certain moment. The ARCA collector only invalidates a previous feed once the new one has been downloaded.

ARCA portals can now interconnect, so it is possible to share content descriptions among them, allowing a certain set of elements processed by one portal being seen by whatever other that is sharing the data. Each portal is able to generate an OMPL description of the information it is willing to share.

6 CURRENT STATUS

The current software version is 8.0.2. The project is in a deployment phase and, from November 2007 the common ARCA portal for the Spanish academic and research community is hosted by RedIRIS. At the moment of this writing, 11 institutions are connected to the portal, that offers more than 3000 references, covering around 2300 hours of playtime. ARCA has been presented in several national and international forums. The portal currently supports five languages: Spanish, Catalan, Galician, Basque and English.

Being an open project, several other organisations are contributing to ARCA: URJC is working on a detailed statistics module for the portal, the University of Vigo has incorporated ARCA publishing to their PUMUKIT framework for media content management, UNICAN has built an application for easily build feeds, and the NGI group at DIT-UPM has developed a parser for the Joomla calendar component into ARCA.

7 ARCAMM

As a complementary element to the ARCA portal a client software, ARCAMM (ARCA Media Manager) is under development. ARCAMM is a IPTV portal, intended to simplify the cataloguing, classification and publication of any content formats supported by ARCA (live events, VoD, podcasts, series/courses, associated documents,...). ARCAMM provides a Web interface to manage all the contents at any participating institution, with modules for managing predefined courses, categories, coding profiles, and locations for events and files. All these sections are integrated in modules for adding or editing content descriptions. It also permits the creation of public and hidden elements, where hidden elements are only visible to administrators. Furthermore, it provides the possibility of direct item publication through ARCA. This way, any item marked for ARCA publication is automatically added (by including it in the appropriate feed) to the ARCA portal.

8 FUTURE WORK AND CONCLUSIONS

The ARCA development team is strongly committed to continue developing and enhancing ARCA. Among the future lines for this we can cite:

- ◆ A new browsing interface
- ◆ Mechanisms to show related contents according to user preferences
- ◆ Enhancing the system classification engine

- ◆ Support for other formats, including documents, associated to content descriptions
- ◆ Management of protected contents by means of federated identity
- ◆ Enhancing the admin interface and the software installation process
- ◆ Extend multi-lingual support.

ARCA is intended to provide any student, faculty member or researcher with relevant and updated information about live events and media contents available at the participant institutions. Value is added to these media contents, now accessible through an intuitive and simple interface. ARCA references can be easily incorporated into Web 2.0 constructs (such as blogs) using a facility provided by the portal. This allows for the creation of new access services, directly controlled by users, both as producers and consumers of media contents.

We would like to invite all universities and research centres to join the project, contributing their content and (whenever possible) becoming part of the ARCA development community. To participate, just send an e-mail to: jmaria.fontanillo@rediris.es