

# Solucions de programari lliure en un projecte d'adaptació de dades a XML

Xavier Giró i Neus Camps

juny de 2006

## Resum

Aquesta comunicació explica el desenvolupament d'un projecte d'innovació utilitzant únicament eines de programari lliure. El treball s'ha realitzat en el marc del projecte de Xarxa Audiovisual de Catalunya impulsada per la Fundació i2cat. El resultat fou una aplicació capaç d'adaptar una base de dades en format text a un format XML segons uns esquemes basats en l'estàndard MPEG-7. No només les tasques de desenvolupament han utilitzat eines lliures existents, sinó que la gestió i comunicació de l'equip de treball utilitzat únicament programari lliure.

## 1 Introducció

La digitalització dels continguts audiovisuals ha provocat l'aparició de noves oportunitats i necessitats en el camp de la indexació. L'intercanvi de material entre productors, distribuïdors i usuaris finals requereix també un format de metadades [9] que descriu els continguts en un llenguatge comú per a tots. Arran d'aquesta necessitat va néixer l'estàndard MPEG-7 [11], un llenguatge de metadades orientat a la descripció de continguts audiovisuals basat en el llenguatge XML. Un possible canal d'intercanvi de continguts són les xarxes IP de banda molt ampla, com la que gestiona i impulsa la Fundació i2cat. La xarxa i2cat pretén potenciar la innovació a Catalunya en el camp de la Internet 2 de nova generació. Entre les múltiples aplicacions de la i2cat s'hi troba l'intercanvi de continguts audiovisuals. La Fundació i2cat va crear la Xarxa Audiovisual de Catalunya (XAC) per tal de comprovar les possibilitats i oportunitats de les xarxes IP per a la indústria audiovisual del país. El projecte exposat en aquesta comunicació se situa en aquest marc més ampli que és la XAC. Un dels compromisos de la Fundació i2cat és el de potenciar sempre que sigui possible l'ús de programari lliure en totes les solucions tecnològiques. Per aquest motiu convidà a la Càtedra de Programari Lliure (CPL)[1] de la Universitat Politècnica de Catalunya (UPC) a participar de la XAC. A través de la CPL es contactà amb el Grup

de Processament de la Imatge del Departament de Teoria del Senyal i Comunicacions (TSC) de la UPC [2], que s'incorporà al projecte amb un equip de treball vinculat de l'Escola Universitària d'Enginyers Tècnics Industrials de Terrassa (EUETIT). L'equip contribuí en el projecte amb el desenvolupament d'una aplicació capaç de traduir les bases de dades de dos productors audiovisuals a un format basat en l'estàndard MPEG-7. En aquesta comunicació s'exposen les eines de programari lliure que han permès assolir els objectius marcats des del projecte XAC. El text s'estructura de la següent forma. La Secció 2 planteja els objectius del projecte global. La Secció 3 presenta les eines de programari lliure utilitzades i quins ús se n'ha fet en el marc del projecte. En concret, es parla del sistema operatiu GNU/Linux, el llenguatge de programació Java, el pas de XML a Java amb el XMLBeans, l'entorn de desenvolupament amb l'Eclipse, el control de versions amb el SubVersion, el Subclipse i el repositori de LaFarga, la gestió i comunicació en el marc del projecte amb el Moodle i, finalment, les tasques d'ofimàtica realitzades amb l'OpenOffice. La Secció 4 exposa les conclusions del projecte des del punt de vista del programari lliure.

## 2 Objectius

Els principals objectius del projecte foren els següents:

- utilitzar únicament eines de programari lliure en tots els aspectes del projecte
- adaptar les metadades de continguts audiovisuals de l'Escola Superior de Cinema i Audiovisuals de Catalunya (ESCAC) i Canal Terrassa TV (TVT) al format de dades definit per la XAC.
- facilitar la reutilització del codi desenvolupat.
- realitzar i defensar el Projecte de Fi de Carrera (PFC) de dos estudiants de la titulació d'Enginyeria Tècnica de Telecomunicacions, especialitat en Imatge i So.

Aquesta comunicació se centra en els aspectes del projecte relacionats amb el programari lliure.

## 3 Eines i usos

En els apartats de continuació explicarem les eines, tecnologies i programes que vam utilitzar durant la realització del projecte. Totes les eines de desenvolupament les vam triar sota dos requeriments: ser programari lliure i multiplataform.

### 3.1 Sistema operatiu: GNU/Linux

A l'inici del projecte vam decidir treballar només amb el sistema operatiu GNU/Linux, tot i que cap dels dos desenvolupadors tenien experiència en el seu ús. El treball al laboratori es realitzà amb la distribució GNU/Linux Fedora Core 4 instal·lada al laboratori del TSC a l'EUETIT. Els projectistes intentaren instal·lar la mateixa distribució a casa sense èxit i en aquell entorn optaren pel sistema operatiu Microsoft Windows XP. Donat que la resta d'eines eren lliures i multiplataforma, el traspàs d'informació entre l'entorn domèstic i acadèmic no resultà cap problema.

### 3.2 Llenguatge de programació: Java

Per desenvolupar l'aplicació d'adaptació de les metadades es va optar pel llenguatge de programació Java [13]. Es tracta d'un llenguatge de programació d'alt nivell basat en C/C++ i desenvolupat per l'empresa Sun Microsystems a principis de la dècada dels 90. Sun Microsystems descriu el Java com "simple, orientat a objectes, distribuït, interpretat, robust, segur, d'arquitectura neutra, portable, d'altres prestacions, multitasca i dinàmic". A data d'avui el codi font de Java és propietari, però recentment Sun Microsystems ha anunciat que l'alliberarien utilitzant segurament la llicència OSI (Open Source Initiative).

Una de les característiques més importants del Java és que és un llenguatge totalment portable entre diferents ordinadors, és a dir, un programa escrit en Java pot executar-se sobre qualsevol plataforma, independentment del maquinari, programari i sistema operatiu. Aquesta portabilitat s'aconsegueix compilant el codi font Java (arxius \*.java) a un codi intermig binari especial (arxius \*.class). Aquest codi és interpretat i executat per la Màquina Virtual de Java (Java Virtual Machine, JVM). La JVM és un programa executable específic per a cada plataforma que tradueix el binari Java al codi màquina que pugui ser comprensible pel processador i sistema operatiu sobre els quals s'executa la JVM.

### 3.3 Unió XML-Java: XMLBeans

Per treballar còmodament amb esquemes i documents XML vam buscar una eina per integrar dades XML en aplicacions Java i poder generar així les instàncies XML finals. XMLBeans [8] és una eina de pont entre classes Java i documents XML. El que en llenguatge tècnic s'anomena XML-Java Binding Tool (Eina d'Unió XML-Java) [14]. XMLBeans té com a característica principal un suport complet a XML Schema. El projecte XMLBeans el va desenvolupar BEA Systems i al setembre de 2003 el va migrar a una llicència de Apache Software Foundation, que distribueix lliurement el programa i la seva documentació.

### 3.3.1 Característiques

XMLBeans permet generar i manipular documents XML a partir de classes Java. D'una banda, XMLBeans aprofita les característiques dels esquemes XML, continguts en arxius XSD, per generar classes Java. Un esquema XML (*schema* en anglès) és un document XML que defineix normes per un altre document XML. XMLBeans proporciona un sistema que compila l'esquema XML per generar classes Java que reflecteixin l'esquema. Amb aquestes classes i l'API que proporciona XMLBeans es poden generar i manipular instàncies XML d'acord amb l'esquema. Les classes Java generades contenen mètodes similars a qualsevol classe Java, com *getters* i *setters*. XMLBeans proporciona un sistema per validar les noves instàncies creades amb l'esquema. D'altra banda, encara que no es disposi d'un esquema XML, XMLBeans també permet generar i accedir a documents XML utilitzant les classes Java que proporciona la seva API. En el nostre cas vam utilitzar la primera funcionalitat de XMLBeans per aprofitar els esquemes XML dissenyats des del projecte XAC.

### 3.3.2 Compilació d'un esquema XSD

El primer pas per començar a utilitzar XMLBeans a partir d'un esquema XML és compilar l'esquema i obtenir les classes Java. Aquestes permeten generar instàncies XML d'acord amb l'esquema. Per compilar un esquema XSD ha d'executar-se la comanda *scomp* (Schema Compiler) des de la línia d'ordres. La sintaxi bàsica és la següent:

```
scomp [opcions] [directori] [esquema.xsd]
```

**opcions** -out [nomFitxerJar.jar], d'on *nomFitxerJar.jar* és el fitxer JAR que contindrà un paquet amb les classes Java resultants de compilar l'esquema.

**directori** Directori que conté el fitxer XSD que es vol compilar.

**esquema.xsd** Nom del fitxer XSD.

### 3.3.3 Programació amb XMLObject

Quan es compila l'esquema, els noms de les classes i els mètodes generats són totalment intuitius, ja que es corresponen amb els noms de l'esquema. Les classes s'agrupen en un paquet (package) amb el nom de l'espai de noms (namespace) definit a l'esquema i s'empaqueten en un arxiu JAR. Per accedir-hi tan sols cal utilitzar els mètodes *gets* i *sets* generats automàticament durant el procés de compilació de l'esquema.

### 3.4 Entorn de desenvolupament: Eclipse

Quan s'ha de treballar amb projectes complexos un simple editor de text pot resultar ineficient. Els Entorns de Desenvolupament Integrats (Integrated Development Environment, IDE) són potents programes que permeten treballar amb diferents llenguatges de programació i que ofereixen moltes eines per facilitar la generació de codi font. Per generar el codi font del nostre programa vam decidir treballar amb Eclipse [3], un IDE de codi lliure i extensible. L'Eclipse és una multiplataforma de programació, desenvolupament i compilació de qualsevol tipus de llenguatge, com per exemple Java, C/C++ i XML. La característica clau d'Eclipse és la seva extensibilitat. A diferència d'altres entorns de desenvolupament, on les seves funcionalitats estan totes incloses tan si les necessita l'usuari com no, Eclipse és una gran estructura formada per un nucli i molts connectors que defineixen la seva funcionalitat final. Existeixen connectors per a un ampli ventall d'aplicacions, com accedir a bases de dades, desenvolupar en C/C++, escriure documents XML... El projecte Eclipse es divideix en tres subprojectes:

1. *Plataforma Eclipse*: El cor de l'aplicació, inclou el nucli, l'entorn de treball (workspace), l'àrea de desenvolupament o interfície d'usuari (workbench), l'ajuda a l'equip (Team support) i l'ajuda.
2. *Java Development Tooling (JDT)*: Connector per al llenguatge Java.
3. *Plug-in Development Environment (PDE)*: Proporciona les eines per al desenvolupament de nous mòduls.

Vam decidir utilitzar Eclipse no només perquè és l'IDE de Java més popular, sinó també per les següents funcionalitats:

- Ressalt de sintaxis del codi font.
- Compilació en temps real.
- Control de versions del codi font amb Subversion, via Subclipse.

### 3.5 Control de versions

Durant el desenvolupament del codi ambdós projectistes compartien codi i treballaven en màquines diferents, al laboratori i a casa. El mètode de treball va requerir d'un sistema per compartir els fitxers d'una manera ordenada i segura. Per això es va decidir utilitzar un sistema de control de versions que permetés compartir el codi font. Un sistema de control de versions permet administrar i guardar qualsevol document que canviï amb freqüència, com el codi font d'un programa. Consisteix en un arbre de fitxers i directoris guardats en un repositori i accedits a través d'un programa client amb el

que cada usuari sincronitza la seva còpia local. El repositori actua com a servidor de fitxers i directoris ordinari, amb l'afegit que recorda els canvis fets. Això permet recuperar versions antigues de les seves dades o examinar l'historial de canvis de les mateixes. Els principals avantatges que aporta un sistema de control de versions són els següents:

**Actualització de fitxers modificats** : El client recorre el codi local i el sincronitza amb el del repositori.

**Còpies de seguretat centralitzades** : Només l'administrador ha de preocupar-se de realitzar còpies de seguretat en el repositori.

**Historial de canvis** : El repositori guarda un registre de tots els canvis realitzats. És possible recuperar qualsevol de les versions anteriors de qualsevol fitxer. Si algú esborra tots els fitxers, es pot tornar enrera i recuperar tot el seu contingut.

**Accés remot** : És possible accedir remotament al repositori a través de xarxes, la qual cosa permet utilitzar-lo des de diferents localitzacions.

**Seguretat** : És possible atorgar diferents permisos sobre diferents branques del projecte per mantenir diverses versions del codi, per exemple, una d'estable i una altra en desenvolupament.

### 3.6 Client: SubVersion

Existen diversos sistemes de control de versions. Vam escollir el Subversion [10], un sistema de control de versions de codi obert també i de distribució lliure. Es tracta d'una solució més moderna al popular CVS (Concurrent Versions System), el sistema més extès ara per ara.

### 3.7 Connector: Subclipse

El programa client SVN que vam utilitzar va ser *Subclipse* [7], un connector de l'Eclipse que dóna suport al sistema de control de versions SVN.

### 3.8 Repositori: LaFarga

El repositori SVN va ser proporcionat per La Farga [4], un servei d'hostatge de projectes o continguts lliures que ofereix la CPL. El servei ofereix un ampli ventall d'eines per a gestionar, emmagatzemar i publicar els projectes lliures:

- Sistema de control de versions (CVS i SVN).
- Dipòsit de distribucions descarregables del projecte.
- Gestor de documents lligats al projecte.

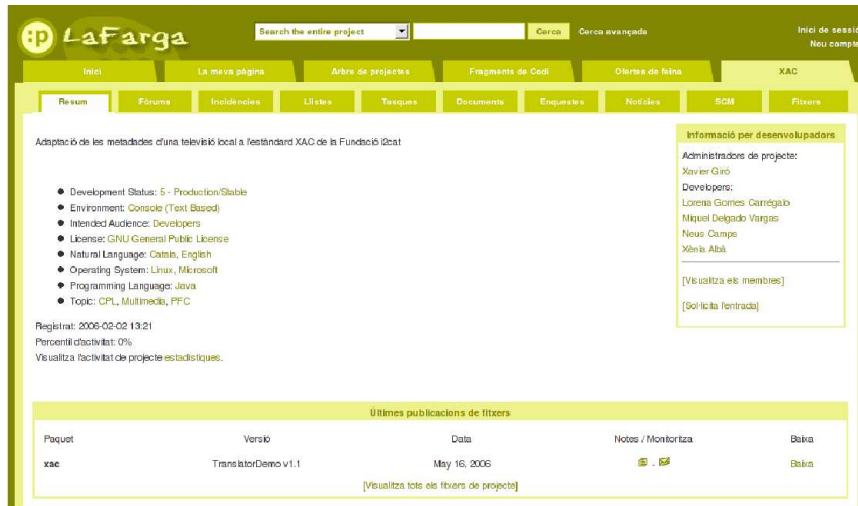


Figura 1: <https://lafarga.cpl.upc.edu/projects/xac/>

- Gestor d'incidències relacionades amb el projecte: errors, suggeriments, etc.
- Fòrums i llistes de distribució de correu.
- Wiki

El servei de La Farga és obert a projectes que es distribueixin sota llicències lliures. Un projecte de la Farga no ha de ser necessàriament de programari. Tot i que les eines estan pensades per a programadors, també són útils per a gestionar altres tipus de projectes, com per exemple la redacció d'un llibre. A més a més, les eines de gestió de llistes de distribució, fòrums de discussió i pàgina web, constitueixen un conjunt de solucions útils per a la gestió de la comunicació entre grups. En el nostre cas ens vam limitar a les funcionalitats de control de versions i de publicació de codi ja que la resta de serveis ja ens els satisfia el Moodle.

Per configurar un sistema de control de versions sols cal indicar al programa client una URL i nom d'usuari per a que es connecti remotament i se sincronitzi amb la còpia de treball local. La Farga també permet la publicació de fitxers i la seva descàrrega gràcies al dipòsit de distribucions descarregables, un servei que nosaltres hem utilitzat per publicar diverses versions del projecte. La Figura 1 mostra una captura pantalla de la interfície web pública del nostre projecte a LaFarga [6].

### 3.9 Gestió i comunicació: Moodle

Abans de començar a desenvolupar el projecte vam creure convenient utilitzar un programa específic per gestionar els continguts i mantenir la comu-

nicació entre el grup de treball. Vam escollir el Moodle [5], un sistema de gestió de continguts orientat al camp docent. El Moodle permet realitzar múltiples activitats d'ensenyança-aprenentatge via Internet. Es pot utilitzar per presentar els continguts d'un curs, col·laborar, fer qüestionaris, assignar tasques, per proporcionar comentaris sobre el material o el desenvolupament del curs. El seu ús com a gestor d'un PFC va resultar una pràctica innovadora i que va obligar a seleccionar quines de les múltiples eines del Moodle ens eren útils pel projecte. L'elecció de Moodle com a eina de gestió va venir donada per diversos motius:

- a partir del pròxim curs 2006-2007 diverses escoles de la UPC començaran a utilitzar Moodle com a interfície virtual de treball per a professors i estudiants. Ens va semblar interessant a tots els membres del grup provar-lo com a mode d'experiència.
- permet tenir un ampli control sobre els paràmetres de configuració, cosa que ens va permetre configurar-lo a mida i adaptar-lo per gestionar el PFC i no a un curs en línia.
- permet crear còpies de seguretat cada dia automàticament i, en cas de pèrdua d'informació, incorpora un sistema de recuperació.

Per tal gestionar el PFC es va crear un curs i s'hi van definir tres tipus d'usuaris registrats: els coordinadors, els projectistes i els assessors externs. No es va permetre l'entrada a cap visitant donada la privacitat d'algunes de les discussions i temàtiques del projecte. Els assessors podien participar a fòrums, els projectistes també podien penjar-hi fitxer i els coordinadors, a més, editar l'arquitectura modular del lloc del curs. La informació al Moodle es va organitzar per temes, la majoria d'ells d'acord amb els capítols finals de la memòria del PFC. Cada tema tenia un seguit de carpetes i documents per guardar els continguts relacionats i un fòrum per discutir-ne. Els fòrums de discussió sobre cada tema foren les activitats més utilitzades i s'hi van discutir d'una manera organitzada que, alhora, va facilitar la posterior redacció de la memòria del PFC. La utilització del bloc d'activitat recent va permetre a cada usuari conèixer les novetats des de l'última entrada i d'aquesta manera estar al dia del treball de la resta de membres de l'equip. El Moodle va resultar una peça clau pel bon desenvolupament del projecte, tant per mantenir l'organització dels continguts, com per mantenir la comunicació no presencial dins l'equip de treball. La Figure 2 mostra una captura de pantalla de la interfície web del nostre projecte al Moodle.

### 3.10 Ofimàtica: OpenOffice.org

La redacció de la memòria del PFC i la seva posterior defensa en una presentació oral es realitzà amb les eines *Writer* i *Impress* del paquet ofimàtic



Figura 2: <http://terrassatsc.upc.es>

OpenOffice. Els documents generats segueixen el format de OpenDocument (ODF) desenvolupat pel consorci industrial OASIS i aplica els principis bàsics del programari lliure al format dels documents d'oficina com, en aquest cas, informes tècnics o presentacions. Tot i que els estudiants estaven familiaritzats amb altres solucions propietàries, no van trobar excessives dificultats per utilitzar l'OpenOffice i descobrir-ne les seves possibilitats. La redacció de la memòria es va basar en el tutorial de Roth [15]. El text es dividí en diferents documents creats a partir d'una mateixa plantilla. D'aquesta manera els dos projectistes van poder treballar en paral·lel en fitxers ODT separats. Un únic document mestre que incloïa les parts serví per generar la memòria final així com generar automàticament una taula de continguts, numeració i referències creuades comuns. El procés fou revisat pel professor a través de l'entrega de diverses versions a través d'un fòrum de discussió del Moodle. Per a la presentació oral també s'utilitzà l'OpenOffice i s'explotà la facilitat d'exportar figures i textos des de la memòria cap al document propi de la presentació.

## 4 Conclusions

En aquesta comunicació s'ha volgut mostrar que avui en dia la comunitat de desenvolupadors de programari lliure ofereix un ventall d'eines suficient per a l'elaboració d'un projecte de fi de carrera de nivell universitari i, alhora, satisfer els requeriments marcats des d'un projecte d'innovació com l'encarregat per la Fundació i2cat. El codi generat s'ha publicat a LaFarga sota la llicència lliure GPL, permetent que d'altres productors audiovisuals que vulguin adaptar les seves metadades a l'estàndard XAC (o a qualsevol format basat en XML) puguin reutilitzar el codi i documentació tècnica as-

sociada. Per altra banda, aquesta comunicació també pretén promocionar l'ús d'eines de programari lliure en l'àmbit acadèmic. És la voluntat dels autors que les guies i experiències proporcionades en aquesta comunicació puguin servir en d'altres activitats similars. Per a un aprofundiment sobre les solucions presentades el lector és referit a la memòria del projecte de fi de carrera de Neus Camps i Miquel Delgado [12].

## Referències

- [1] CÃ tedra de programari lliure. <http://www.cpl.upc.edu/>.
- [2] Departament de teoria del senyal i comunicacions de la upc a terrassa. <http://terrassatsc.upc.edu>.
- [3] Eclipse.org home. <http://www.eclipse.org>.
- [4] Lafarga.cat, idees i projectes relacionats amb programari lliure. <http://www.lafarga.cat>.
- [5] Moodle - a free open source management system for online learning. <http://www.moodle.org>.
- [6] Portal del projecte a lafarga. <https://lafarga.cpl.upc.edu/projects/xac/>.
- [7] A subversion eclipse plugin. <http://subclipse.tigris.org>.
- [8] Welcome to xmlbeans. <http://xmlbeans.apache.org>.
- [9] S.Kalli A. Lugmayr, S.Niiranen. *Digital interactive TV and metadata*. Springer, New York, Estats Units, 2004.
- [10] B.W. Fitzpatrick B. Collins-Sussman and C. Michael Pilato. *Version Control with subversion*. O'Reilly, Sebastopol, CA, Estats Units, 2004.
- [11] P.Salembier B.S.Manjunath and T.S.Sikora. *Introduction to MPEG-7*. Wiley, 2002.
- [12] Neus Camps i Miquel Delgado. *Adaptacio de les metadades audiovisuals l'estandard XAC*. EUETIT (UPC), Terrassa, Catalunya, 2006.
- [13] Y. Daniel Liang. *Introduction to Java programming with JBuilder*. Prentice Hall, Upper Saddle River, NJ, Estats Units, 2004.
- [14] Brett McLaughlin. *Java and XML data binding*. Safari Books, Sebastopol, CA, Estats Units, 2002.
- [15] Werner Roth. *Writing a Thesis with OpenOffice.org*. OpenOffice.org, 1999.