

RODA i CRIB. Un repositori digital orientat als serveis

**JOSÉ CARLOS RAMALHO, MIGUEL FERREIRA, LUIS FARIA,
RUI CASTRO, FRANCISCO BARBEDO, LUIS CORUJO**

Introducció

A mitjan 2006, els Arxius Nacionals Portuguesos (Direcció General d'Arxius Portuguesos) van endegar un projecte anomenat RODA (Repository of Authentic Digital Objects) amb l'objectiu d'identificar i reunir la tecnologia, els recursos humans i el suport polític necessaris per preservar a llarg termini els materials digitals generats per l'Administració pública portuguesa.

Un dels objectius originals del projecte RODA era la creació d'un repositori digital capaç d'ingressar, gestionar i proporcionar accés als diversos tipus d'objectes digitals generats per les institucions públiques nacionals. La creació d'aquest repositori s'havia de basar en tecnologies obertes i, sempre que fos possible, en les normes existents, com OAIS (Open Archival Information System)

NOTA: Volem a agrair als autors que ens hagin permès publicar aquest article i a Raimon Nualart la revisió que ha fet de la traducció.

(SYSTEMS 2002), METS (of Congress 2006), EAD (of Congress 2002) i PREMIS (Group 2005).

A un nivell més alt, el model OAIS està integrat per tres grans processos (ingrés, administració i consulta). Al projecte RODA s'especifiquen els cicles de treball de cadascun d'aquests processos.

El procés d'ingrés permet gestionar els paquets nous d'informació que s'afegeixen al repositori digital, és a dir, els paquets d'informació de transferència (SIP per les sigles en anglès). En aquest àmbit, el projecte especifica formalment l'estructura dels SIP.

Durant l'ingrés, els SIP es transformen en paquets d'informació d'arxiu (AIP per les sigles en anglès). Durant el projecte es va especificar un model de dades per emmagatzemar AIP. El procés de consulta és el procés en el qual es responen les sol·licituds dels usuaris lliurant-los paquets d'informació de consulta (DIP per les sigles en anglès). Per a cadascun dels tipus d'objectes digitals emmagatzemats al sistema RODA es van especificar una o més estructures DIP. Actualment, el sistema RODA pot emmagatzemar i donar accés als següents tipus d'objectes digitals: documents de text, imatges fixes i bases de dades relacionals.

La normalització té un paper molt important en el sistema RODA, atès que no tots els tipus de documents de text o imatges fixes es poden arxivar, i fins i tot en el cas de les bases de dades, cada sistema de gestió de bases de dades disposa del seu propi model de dades. Per tant, durant el desenvolupament del projecte es van haver de prendre mesures per normalitzar els formats. Cadascun dels objectes digitals emmagatzemats al sistema RODA se sotmet a un procés de normalització: els documents de text es normalitzen en format PDF; les imatges fixes, en TIF sense comprimir; les bases de dades relacionals, en DBML (Database Markup Language) (Ramalho *et al.* 2007).

El projecte RODA es divideix en diferents components. El component bàsic és el marc Fedora Commons. Fedora aplica les funcions digitals habituals d'un repositori digital, com l'abstracció d'emmagatzematge d'objectes (i metadades) digitals i les relacions entre els objectes, i es pot ampliar amb el motor de cerca Generic Lucene de Fedora. A més a més, els serveis centrals del sistema RODA apliquen tots els serveis bàsics de RODA, als quals es pot accedir programàticament. I, en darrer lloc, la interfície d'usuari web del sistema RODA permet a l'usuari final examinar, cercar, accedir i administrar fàcilment tots els objectes digitals, les metadades i les funcions d'ingrés, preservació i consulta.

Malgrat tots els esforços dedicats al desenvolupament del projecte RODA, encara no hi ha hagut cap acció que fomenti la preservació digital activa real. Un cop els materials s'han emmagatzemat a l'arxiu, es conserven intactes i, per tant, resten susceptibles a l'obsolescència tecnològica, especialment des del punt de vista del format.

Al mateix temps, a la Universitat de Minho, es va començar a desenvolupar un projecte anomenat CRiB (Conversion and Recommendation of Digital Object Formats) amb l'objectiu d'ajudar les institucions de patrimoni cultural i els usuaris a aplicar intervencions de preservació basades en la migració. Entre els serveis que ofereix aquest sistema s'inclouen convertidors de format, eines d'avaluació de la qualitat, plans de preservació digital i producció automàtica de metadades per mantenir l'autenticitat de les representacions.

El sistema CRiB es va dissenyar amb una arquitectura orientada als serveis i pot proporcionar serveis per a les accions que s'enumeren a continuació:

- Identificar el format del fitxer.
- Recomanar les opcions de migració més òptimes tenint en compte els requisits de preservació específics de cada institució client.
- Conversió o transformació del format original dels objectes digitals a uns formats més actuals.
- Valorar el control de qualitat del procés global de migració - pèrdua d'informació, i la idoneïtat del rendiment i el format per a la preservació a llarg termini.
- Generar metadades de preservació en format PREMIS per documentar correctament la intervenció de preservació i conservar l'autenticitat dels objectes.

Després d'obtenir un finançament suplementari per continuar desenvolupant el sistema RODA, l'equip va decidir emprar el projecte CRiB com a unitat de planificació i execució de la preservació.

El projecte RODA segueix una arquitectura orientada als serveis per facilitar el desenvolupament i l'actualització paral·lels i permetre l'ús de tecnologies heterogènies i plataformes independents entre els diferents components. El projecte CRiB també està orientat als serveis per permetre l'ampliació de serveis que només són possibles fent servir plataformes i tecnologies específiques. Aquest informe inclou una descripció de tots dos projectes i analitza la integració de CRiB com a component del sistema RODA, la qual cosa permet emprar les funcions del projecte CRiB per normalitzar les tasques d'ingrés i de generació de metadades, per planificar les accions de preservació, i fins i tot per als serveis de consulta.

El projecte RODA

Els arxius digitals són estructures complexes que solen estar formades per recursos humans, tecnologies de gamma alta, polítiques i informació. El projecte RODA ha obert el camí a una sèrie d'estudis sobre tots aquests àmbits. Els objectius originals del projecte eren bastant ambiciosos i consistien a:

- Definir els requisits funcionals per a un arxiu digital, els seus usuaris i les aplicacions compatibles.
- Dissenyar models conceptuals, lògics i de contingut per a un arxiu digital.
- Identificar el conjunt d'esquemes de metadades necessari per a totes les funcions de l'arxiu digital (descriptives, tècniques, estructurals i metadades de preservació).
- Identificar els requisits tècnics i organitzatius.
- Desenvolupar un sistema d'arxiu digital capaç d'emmagatzemar i preservar objectes digitals durant el període de temps definit per la legislació.
- Desenvolupar mòduls que s'integrin amb les aplicacions o eines disponibles de gestió de documents.
- Desenvolupar una política de transferència i ingrés per als objectes digitals generats per les institucions públiques portugueses.
- Dissenyar una política de preservació digital per a l'arxiu digital.
- Promocionar un estudi sobre els models empresarials capaços de finançar l'arxiu digital.
- Definir unes taxonomies de les propietats significatives per a cadascuna de les classes d'objectes digitals que haurà de gestionar l'arxiu digital per tal d'aplicar mecanismes de control de la qualitat.

Una de les fases d'aquest projecte va consistir a desenvolupar un sistema d'arxiu digital capaç de preservar la documentació digital i garantir l'accés dels usuaris potencials a aquesta documentació sense comprometre en cap moment l'autenticitat dels documents. Aquest repositori hauria de servir de base per crear un arxiu digital totalment funcional en el qual es podrien ingressar i gestionar grans quantitats d'objectes digitals a escala nacional. En la seva primera versió, el repositori permetria treballar amb un petit ventall de tipus d'objectes, com documents de text, imatges i bases de dades relacionals.

Arquitectura

El sistema RODA segueix el model OAIS (Open Archival Information System Reference Model) (SYSTEMS 2002), que determina els principals components funcionals que hauria d'incloure un sistema d'arxiu capaç de preservar materials

digitals a llarg termini. El model proposat està format per quatre unitats funcionals principals —ingrés, gestió de dades, conservació o emmagatzematge i accés— i dues unitats addicionals, anomenades planificació de la preservació i administració. La figura 1 il·lustra la manera com aquestes unitats funcionals interactuen entre elles i amb totes les parts interessades del repositori (internes i externes).



Figura 1. Arquitectura general del sistema RODA

Abans de passar al desenvolupament tècnic, l'equip d'arxiviers del projecte RODA va aplegar diversos requisits funcionals (Barbedo 2006) i va dur a terme un estudi sobre les plataformes utilitzades com a repositoris digitals disponibles en aquell moment. En aquest estudi, DSpace (COMPANY i LIBRARIES) i Fedora (Lagoze *et al.* 2005) es van comparar amb aquest conjunt de requisits.

DSpace va superar Fedora en la majoria de requisits. Però, tot i així, l'equip del projecte va acabar escollint Fedora com a plataforma de desenvolupament. Malgrat que, com va fer evident l'estudi, DSpace ofereix una gamma més àmplia de funcions a punt per ser utilitzades i d'interfícies fàcils d'utilitzar, li manca flexibilitat i capacitat d'expansió. Un exemple molt pragmàtic d'aquesta disposició són els esquemes de suport de metadades, exceptuant el Dublin Core. Perquè DSpace fos compatible amb estructures de metadades descriptives més complexes, com EAD (of Congress 2002), s'hi haurien de dedicar moltes hores de feina.

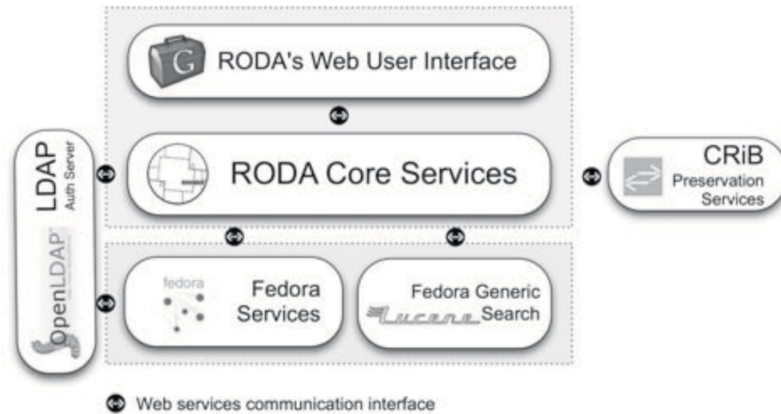


Figura 2. Arquitectura orientada al servei del sistema RODA

La figura 2 il·lustra l'arquitectura general de serveis del sistema RODA. A la part inferior hi ha els serveis bàsics proporcionats per Fedora, que inclouen les tasques elementals en l'àmbit de la gestió de dades i la conservació i emmagatzematge de fitxers. Alguns exemples de les accions que proporcionen aquests serveis són: ingressar, afegir una seqüència de dades a un objecte, obtenir una seqüència de dades, depurar un objecte, cercar objectes i enumerar seqüències de dades. La funció de cerca de Fedora s'obté amb Apache Lucene i els procediments d'autenticació es realitzen a través d'un servidor LDAP (Lightweight Directory Access Protocol).

Els serveis centrals del sistema RODA permeten dur a terme les tasques més complexes, com aplicar tot el conjunt d'accions del flux de treball de l'ingrés, cercar en el repositori amb mètodes més avançats i abstractes i realitzar funcions administratives al repositori. A més, els serveis centrals de RODA fan servir el mateix servidor LDAP que s'ha esmentat anteriorment per a l'autenticació dels usuaris del repositori digital.

Damunt dels serveis centrals de RODA hi ha la interfície web d'usuari (RODA-WUI). Aquesta capa es fa servir per gestionar tots els aspectes de la interfície gràfica d'usuari i està destinada als productors, usuaris, arxivers, administradors del sistema i experts en preservació. Els components de RODA-WUI són compatibles amb Google Web Toolkit i totes les comunicacions es fan a través d'AJAX i tecnologies de serveis web.

Procés d'ingrés. Com s'ha esmentat anteriorment, Fedora només proporciona una sèrie de serveis molt bàsics que els desenvolupadors han d'ampliar per crear un sistema d'arxiu digital que funcioni correctament, la qual cosa inclou el

desenvolupament d'interfícies gràfiques d'usuari i la caracterització de la majoria d'unitats funcionals OAIS que s'especifiquen al principi d'aquest document. El procés d'ingrés fou la primera unitat que es va desenvolupar.

El procés d'ingrés permet afegir documentació nova al repositori i inclou totes les tasques necessàries per descriure, indexar i emmagatzemar correctament aquests documents. Per exemple, en aquesta fase, el repositori pot transformar o migrar els formats dels fitxers transferits a formats normalitzats adequats per a la preservació a llarg termini i sol·licitar a l'usuari que afegeixi metadades descriptives a aquests objectes per facilitar la seva recuperació en el futur fent servir els mecanismes de cerca disponibles. També és una pràctica habitual emmagatzemar les seqüències de bits originals dels documents ingressats amb la versió normalitzada (per si es creés una estratègia de preservació més avançada que permetés gestionar millor la cadena de bits originària).

Les entrades noves s'introdueixen en uns paquets anomenats paquets d'informació de transferència (SIP per les sigles en anglès). Quan el procés d'ingrés ha finalitzat, els SIP es transformen en paquets d'informació d'arxiu (AIP per les sigles en anglès), és a dir, en els paquets que s'emmagatzemaran en el repositori digital. Les metadades estructurals, tècniques i de preservació són fonamentals per dur a terme les activitats de preservació i, per tant, estan vinculades als AIP.

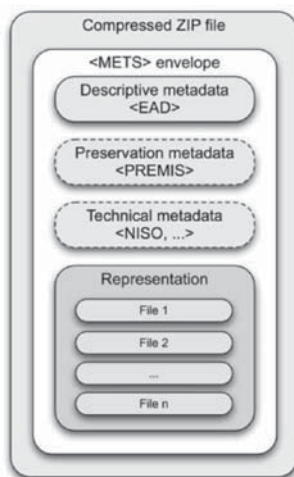


Figura 3. Estructura del paquet d'informació de transferència

El SIP és el format utilitzat per transferir continguts nous del productor al sistema d'arxiu. Està format per una o més representacions digitals i totes les metadades vinculades, emmagatzemades en un sobre METS. A la figura 3 es pot observar l'estructura d'un SIP compatible amb el projecte RODA. El SIP de

RODA és un arxiu ZIP comprimit que conté un document METS, el conjunt de fitxers que integren les representacions lliurades i una sèrie de metadades. En el SIP hi hauria d'haver, com a mínim, un conjunt de metadades descriptives seguint l'esquema EAD¹. Tot i així, en un paquet de transferència també es poden trobar metadades de preservació i metadades tècniques, malgrat que aquest darrer conjunt de metadades no és obligatori, ja que els productors poques vegades el creen. Tanmateix, es va considerar important que el sistema RODA fos compatible amb aquests elements SIP addicionals per a situacions especials, com la successió del repositori, és a dir, quan els elements ingressats pertanyin a un altre repositori que s'hagi de desactivar.

Abans que els SIP es puguin incorporar totalment al repositori, s'han d'enviar a una sèrie de proves per avaluar-ne la integritat, la completesa i la conformitat amb la política d'ingrés.

Si no supera alguna de les fases de validació, el SIP es rebutja i s'envia un informe al grup d'arxivers i al productor. A continuació, el productor pot solucionar el problema i enviar una versió nova del SIP.

Interfície d'accés. El component d'accés estableix una interfície entre l'arxiu i l'usuari final (és a dir, el consumidor). Aquesta unitat funcional pot localitzar un AIP consultant la gestió de dades i recuperar-lo de l'arxiu. A continuació, l'AIP es transforma en un paquet d'informació de consulta (DIP per les sigles en anglès) i es lliura a l'usuari.

Model de dades

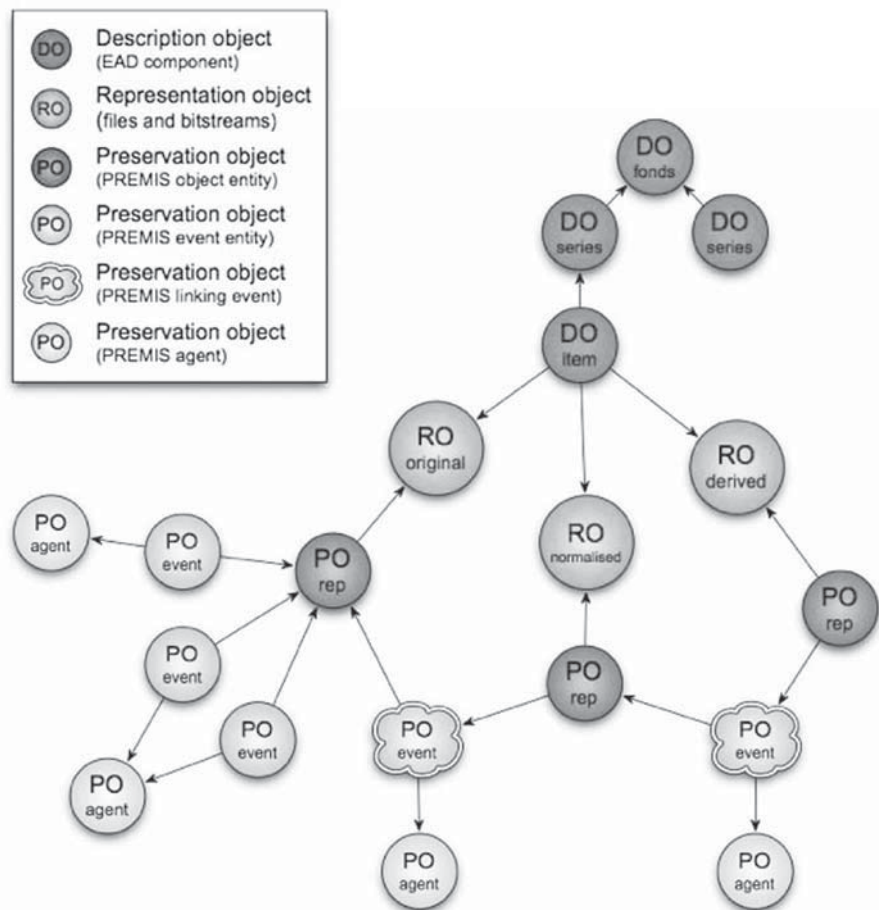


Figura 4. Model de dades de RODA

El model de dades del sistema RODA és un model atomístic i molt orientat a PREMIS (figura 4). Cada entitat intel·lectual es descriu amb un document de metadades seguint l'esquema EAD (nodes DO a la figura 4). Aquests documents s'organitzen jeràrquicament per aconseguir una descripció arxivística completa d'una col·lecció, però es conserven per separat dins del model de continguts de Fedora Commons. Les relacions entre els components EAD es creen mitjançant el mecanisme de vinculació RDF de Fedora.

D'altra banda, cada document que forma part d'una col·lecció jeràrquica (és a dir, un fitxer o un component) està vinculat a un objecte de representació (els nodes RO a la figura), és a dir, un objecte Fedora que inclou tots els fitxers i se-

qüències de bits que formen la representació digital. En darrer lloc, tots aquests objectes estan vinculats entre ells gràcies a una sèrie d'entitats PREMIS que contenen informació sobre la procedència i l'historial dels esdeveniments de l'objecte digital (nodes PO).

Tots els esdeveniments de preservació que tenen lloc al repositori queden registrats en un nou node d'esdeveniment de preservació (els nodes PO d'esdeveniment de la figura). Les accions especials, com les migracions de format, estableixen relacions entre dos nodes de representació i de preservació i, en aquest context, s'anomenen esdeveniments vinculants. Cada esdeveniment de preservació és executat per un agent, que pot ser un usuari del sistema o una aplicació de programari que s'activi automàticament. L'agent que inicia l'esdeveniment queda emmagatzemat als nodes PO d'agent.

CRiB

CRiB² és un projecte que s'està desenvolupant a la Universitat de Minho que ofereix una sèrie de serveis de preservació destinats a ajudar a les institucions durant la planificació i l'execució d'intervencions de preservació basades en un procés de migració (Ferreira, Baptista i Ramalho 2007; 2006). Per facilitar la planificació de la preservació s'empra un servei de suport que permet prendre decisions ben fonamentades sobre quina és la millor estratègia de migració disponible en aquell moment, tenint en compte els requisits individuals de cada institució client. El component que executa l'estratègia de preservació és gestionat per un ampli conjunt de serveis de migració que es poden agrupar per crear un pla d'acció de migració més complex. Per entendre millor com funciona el sistema, cal descriure tots els components que l'integren.

Arquitectura

La figura 5 il·lustra l'arquitectura general del sistema CRiB. A la capa de l'aplicació es pot observar com les aplicacions client aprofiten els serveis que proporciona el sistema CRiB. Aquestes aplicacions poden ser, per exemple, programes personalitzats desenvolupats per usuaris individuals o aplicacions complexes, com repositoris digitals tipus DSpace, Fedora, Eprints or RODA.

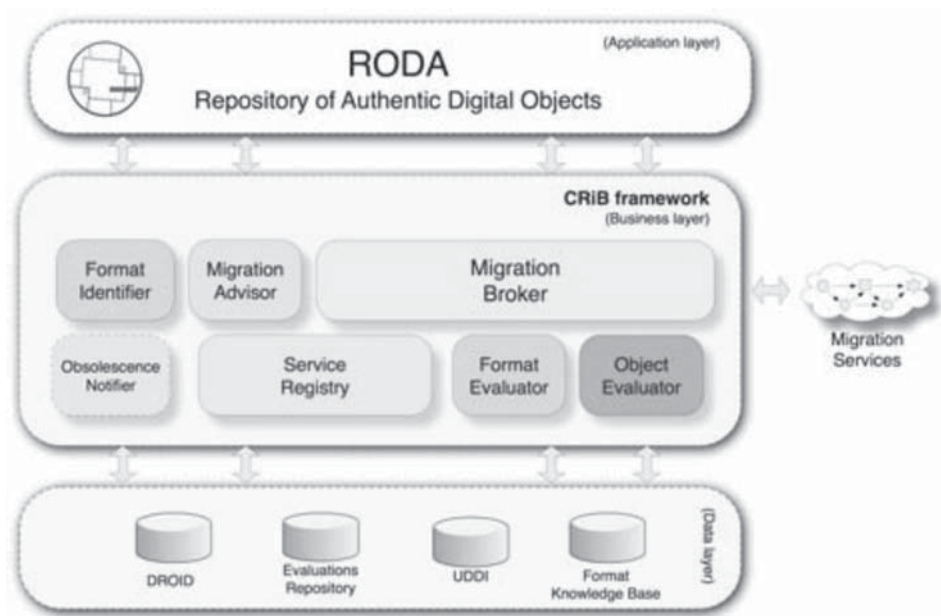


Figura 5. Arquitectura del projecte CRiB

La capa de serveis representa el conjunt de components que actualment ofereix CRiB.

L'identificador de format, tal com suggereix el nom, és un servei que permet identificar la codificació bàsica d'una representació digital. Les aplicacions client responsables de preservar objectes digitals han de poder identificar, caracteritzar i validar la integritat dels objectes que contenen i, si és possible, sense intervencions humanes. Aquest servei és indispensable per aconseguir aquest objectiu. A més a més, permet que les descripcions del format siguin uniformes entre tots els components del sistema CRiB —les descripcions de format les proporciona un vocabulari controlat de formats de fitxers anomenat PRONOM i que ha estat desenvolupat pels Arxius Nacionals del Regne Unit (Darlington 2003).

El notificador d'obsolescència és responsable de controlar el nivell de desús dels formats de fitxers reconeguts. Quan un format de fitxer determinat està a punt d'esdevenir obsolet (per exemple, perquè es publica una nova versió d'un format), aquest component s'assegura d'activar els esdeveniments de preservació adequats. Actualment, el sistema CRiB no aplica aquest component, que ha estat el tema central d'una iniciativa australiana anomenada AONS (Curtis *et al.* 2007).

El sistema CRiB també proporciona una àmplia gamma de serveis de migració per convertir imatges i documents de text a diversos formats. El component

de registre de servei permet emmagatzemar informació sobre aquests serveis. D'aquesta manera, el sistema CRiB pot esbrinar ràpidament quins serveis de migració estan disponibles i es poden utilitzar en aquell moment. Les metadades utilitzades en aquest component es basen en la norma Universal Description, Discovery and Integration (UDDI per les sigles en anglès).

L'agent de migració és un component addicional que garanteix que les migracions programades es duïguin a terme atòmicament des del punt de vista del CRiB. A més, l'agent mesura el rendiment de cada servei de migració per garantir el control de qualitat. El rendiment es mesura segons diversos criteris, com la disponibilitat, l'estabilitat, la capacitat de processament, el cost, la mida de la representació final i el nombre de fitxers resultants a la representació final amb relació a la representació original. Els resultats d'aquestes avaluacions s'emmagatzemen en un component addicional anomenat repositori d'avaluacions. Un assessor de migració utilitza aquests resultats per determinar quins són els serveis de migració disponibles més adequats.

L'avaluador d'objecte és un component que permet detectar la possible pèrdua de dades durant el procés de conversió. Aquesta avaluació és fonamental per determinar l'èxit d'un procés de migració i per documentar correctament la intervenció de preservació. L'avaluador d'objecte compara la representació enviada al procés de migració amb els seus equivalents convertits. Les avaluacions es duen a terme a partir d'un conjunt de criteris fixos però ampliables que se solen denominar propietats significatives, un conjunt d'atributs que s'han de mantenir intactes durant la intervenció de preservació. Aquests atributs caracteritzen la representació digital com una entitat intel·lectual única, independentment del format en el qual estigui codificada la representació.

Les avaluacions realitzades per l'avaluador d'objecte es retornen a l'aplicació client per documentar-les i s'emmagatzemen al repositori d'avaluacions (per facilitar el procés posterior de recomanació).

L'informe d'avaluació que s'envia al client segueix l'estructura de l'entitat d'esdeveniment que es descriu al diccionari de dades PREMIS. Aquesta entitat inclou elements per descriure el tipus d'esdeveniment (per exemple, una migració), la data i l'hora de l'esdeveniment, l'agent que l'ha dut a terme i informació detallada sobre el seu resultat (per exemple, la quantitat de canvis en les propietats significatives que s'han produït durant el procés de migració).

L'avaluador de format proporciona informació sobre l'estat actual dels formats de fitxers. Aquesta informació permet a l'assessor de migració determinar quins formats són més adequats per a la preservació a llarg termini a partir de les ca-

racterístiques tècniques. Per obtenir els resultats oportuns, l'avaluador de format consulta la base de coneixements de format, un magatzem de dades de fets reconeguts sobre formats digitals. En el futur aquest servei podria ser substituït per altres fonts d'informació, com serveis proporcionats per PRONOM o altres serveis externs com Google Trends.

L'assessor de migració permet planificar la preservació generant suggeriments d'alternatives de migració i compara els requisits de preservació assenyalats per les aplicacions client i els seus usuaris amb els coneixements acumulats sobre la qualitat i el rendiment de cada servei de migració independent (o conjunt de serveis de migració). Cal recordar que aquest component s'enriqueix amb cada migració. Durant una migració, el sistema registra la qualitat i el rendiment amb relació a la pèrdua de dades, l'estat dels formats implicats i el rendiment de migració. Amb aquesta informació, l'assessor de migració pot valorar totes les opcions de migració disponibles i oferir un proposta per a una intervenció de migració. Aquests suggeriments solen incloure el format final i el(s) punt(s) d'accés dels serveis de migració més adients i/o les pautes de la migració. Podeu trobar informació addicional sobre el funcionament intern d'aquest procés de recomanació a Ferreira, Baptista i Ramalho (2007; 2006).

Integració entre RODA i CRiB

Com s'ha descrit anteriorment, el sistema CRiB ofereix un ampli ventall de serveis de preservació a qualsevol institució client, aplicació o usuari individual que vulgui conservar les seves col·leccions d'objectes digitals en codificacions interpretables i actualitzades, i reduir al mínim el risc de perdre característiques representatives importants.

La majoria de serveis que ofereix CRiB es poden emprar al sistema RODA. A continuació s'exemplifica l'ús semiautomàtic d'aquests serveis per part del repositori:

Durant el curs de la seva activitat, el sistema RODA ingressa i arxiva una gran varietat d'objectes digitals, la majoria dels quals són enviats pels productors i, en la majoria de casos, codificats en formats diferents. Després d'ingressar aquests objectes, el sistema RODA sol recórrer al servei d'identificació de format del projecte CRiB per determinar si l'objecte digital transferit recentment és adequat i està reconegut com un dels formats de preservació inclosos a la política de preservació de l'arxiu.

Si el document ingressat encara no està codificat en un format de preservació acceptable, es pot consultar el sistema CRiB per trobar els serveis de migració disponibles que permetin dur a terme la normalització corresponent. Després d'obtenir una llista dels possibles serveis de migració, el sistema RODA pot seleccionar un dels punts d'accés suggerits per obtenir una nova representació de l'objecte.

Juntament amb la nova representació, el sistema CRiB ofereix un informe de migració que descriu detalladament el resultat de l'acció de preservació, especialment pel que fa als efectes de la intervenció en les propietats significatives de l'objecte original. A continuació, el sistema RODA pot registrar aquest informe com a metadata de preservació per documentar totalment la intervenció realitzada. Les metadades de preservació proporcionen evidència sobre totes les accions de preservació realitzades amb un objecte determinat en el repositori digital i es consideren una eina fonamental per a la preservació d'objectes digitals autèntics.

El repositori RODA també utilitza els serveis de migració de CRiB per crear representacions derivades dels objectes preservats al sistema, amb l'objectiu de millorar-ne la difusió.

El sistema consulta rutinàriament el notificador d'obsolescència per comprovar si algun dels formats de preservació que utilitza es troba en risc de convertir-se excepcionalment en un format obsolet. En aquests casos, el sistema RODA pot sol·licitar una recomanació a l'assessor de migració de CRiB per obtenir un format de preservació nou i dur a terme un procés de migració de tots els objectes obsolets.

Al sistema RODA, les funcions d'administració de la preservació les duu a terme un planificador de tasques en el qual un usuari especial —l'expert en preservació— pot definir un conjunt de normes que executin accions de preservació específiques. Les accions de preservació es regeixen per un API comú i, per tant, crear i instal·lar accions noves al repositori és tan senzill com copiar l'arxiu del programa al directori corresponent del servidor. Aquestes accions poden invocar serveis remots, com els que proporciona el sistema CRiB, però s'han d'activar localment per complir les normes de conformitat. Les accions activades localment han d'administrar totes les seleccions de serveis remotes i totes les possibles excepcions que es puguin produir.

El planificador de tasques permet a l'expert en preservació configurar les normes que serviran per seleccionar els objectes rellevants per a una intervenció de preservació concreta, a més de planificar la mateixa intervenció.

Conclusió i propers treballs

Com s'ha esmentat anteriorment, el sistema RODA ha estat concebut com un repositori digital que proporciona funcionalitat per a totes les entitats principals que integren el model de referència OAIS. El sistema RODA aplica un cicle de treball d'ingrés que permet validar els SIP i s'encarrega de tot el procés de negociació entre l'arxiu i els productors d'informació. A més, el sistema RODA també inclou funcions d'accés i proporciona diferents mètodes per cercar i navegar entre les metadades disponibles i visualitzar/descarregar els objectes digitals emmagatzemats. També s'han desenvolupat components administratius, que permeten als arxivers modificar les metadades descriptives i definir normes per a les intervencions de preservació, com planificar comprovacions d'integritat en tots els objectes digitals emmagatzemats, iniciar un procés de migració amb determinats tipus/formats de representació, o controlar quins usuaris o grups estan autoritzats a realitzar determinades accions al repositori.

Tot i que el sistema RODA cobreix la majoria de components funcionals que es descriuen al model de referència OAIS, hi havia un element molt important que no s'havia inclòs en el disseny d'aquest projecte: la planificació de la preservació.

Segons OAIS, la planificació de la preservació proporciona els serveis i les funcions per controlar l'entorn d'OAIS i forneix recomanacions per garantir que la comunitat d'usuaris designada continuï podent accedir a la informació emmagatzemada al sistema OAIS a llarg termini, fins i tot, si l'entorn informàtic original queda obsolet.

Les funcions de planificació de la preservació inclouen les tasques següents: avaluar els continguts de l'arxiu i recomanar periòdicament actualitzacions de migració de formats de fitxers, desenvolupar recomanacions per a normes i polítiques en matèria d'arxius i fer un seguiment dels canvis en l'entorn tecnològic i dels requisits de serveis i la gestió de la base de coneixements de les comunitats designades. La planificació de la preservació també permet dissenyar plans de migració detallats, crear prototipus de programari i elaborar plans de prova per aplicar les accions de migració (SYSTEMS 2002).

Per a l'equip de desenvolupament va semblar obvi que la funcionalitat que mancava a RODA seria substituïda fàcilment pel sistema CRiB i els seus components. Per la seva naturalesa orientada als serveis, el sistema CRiB s'integraria sense problemes amb la resta de components i serveis desenvolupats al voltant de Fedora.

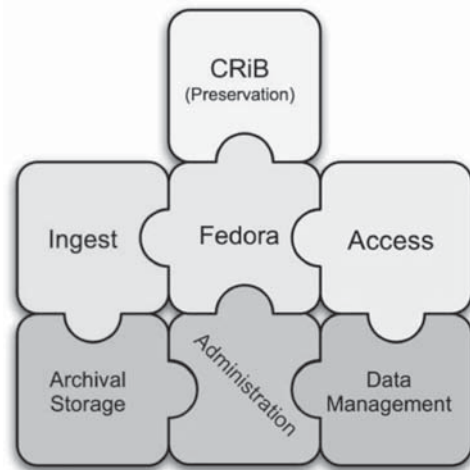


Figura 6. RODA amb el nou component

Per satisfer completament els requisits inicials del sistema RODA, el projecte CRiB hauria de ser capaç de gestionar un tipus de document electrònic addicional: les bases de dades relacionals. Els propers estudis se centraran en el desenvolupament d'una taxonomia de propietats significatives per a les bases de dades relacionals, l'especificació d'un format/esquema de preservació a llarg termini per a aquest tipus d'objectes, el desenvolupament de serveis de migració per a diferents productes de bases de dades (per exemple, Oracle, servidor SQL, Postgresql i MySQL, entre d'altres). Ja s'han dut a terme alguns treballs preliminars sobre aquest tema, que es poden consultar a Ramalho *et al.* 2007 i Henriques *et al.* 2002.

Notes

- 1 Un document EAD no descriu una única representació. De fet, els EAD s'usen per descriure una col·lecció sencera de representacions. El SIP del projecte RODA inclou només un segment d'un EAD, que permet descriure una representació, és a dir, un element i tots els seus subelements. L'equip ha anomenat aquest subconjunt d'EAD «component EAD».
- 2 Conversió i recomanació de formats d'objectes digitals.

Referències

- BARBEDO, F. *Especificação de requisitos*. Informe tècnic 41012-005, IAN/TT. 2006.
- COMPANYY, H.-P.; LIBRARIES, M. Pàgina web de Dspace. <<http://www.dspace.org>>.
- CURTIS, J.; KOERBIN, P.; RAFTOS, P.; BERRIMAN, D.; HUNTER, J. «Aons - an obsolescence detection and notification service for web archives and digital repositories». Dins: *New Review of Hypermedia and Multimedia* 13. 2007.
- DARLINGTON, J. «Pronom - a practical online compendium of file formats». Dins: *RLG DigiNews* 7. 2003.
- FERREIRA, M.; BAPTISTA, A. A.; RAMALHO, J. C. «A foundation for automatic digital preservation». Dins: *Ariadne*. 2006.
- FERREIRA, M.; BAPTISTA, A. A.; RAMALHO, J. C. «An intelligent decision support system for digital preservation». Dins: *International Journal on Digital Libraries*. 2007.
- GROUP, P. W. G. O. O. C. L. C. R. L. «Data dictionary for preservation metadata: final report of the premis working group oclc online computer library center & research libraries group». Informe tècnic. Dublin, Ohio, EUA, 2005.
- HENRIQUES, M.; LIBREOTTO, G.; RAMALHO, J.; HENRIQUES, P. «Bidirectional conversion between xml documents and relational data bases». Dins: *Conferència internacional sobre el treball cooperatiu assistit per ordinador en el disseny*. 2002.
- LAGOZE, C.; PAYETTE, S.; SHIN, E.; WILPER, C. «Fedora - an architecture for complex objects and their relationships». Dins: *Journal of Digital Libraries*. 2005.
- of Congress, T. L. 2002. Pàgina oficial do ead versão de 2002. <<http://www.loc.gov/ead/>>.
- of Congress, T. L. 2006. Pàgina web de la norma METS. <<http://www.loc.gov/standards/mets>>.
- PROJECT, F. «The fedora digital object model». <<http://www.fedora.info/download/2.0/userdocs/digitalobjects/objectModel.html>>.
- RAMALHO, J. C.; FERREIRA, M.; FARIA, L.; CASTRO, R. «Relational database preservation through xml modelling». Dins: *Extreme Markup Languages*. Montreal, Canadà, 2007.
- SYSTEMS, C. C. F. S. D. National Aeronautics and Space Administration. 2002.

RESUM

El 2006, els Arxius Nacionals Portuguesos (Direcció General d'Arxius Portuguesos) van començar a desenvolupar un repositori digital compatible amb el model OAIS per a la preservació de material digital a llarg termini. Paral·lelament, a la Universitat de Minho es va endegar un projecte anomenat CRiB amb la finalitat de desenvolupar un ampli conjunt de serveis creats per contribuir a la preservació digital, com convertidors de format i eines d'avaluació de la qualitat, plans de preservació digital i producció automàtica de metadades per mantenir l'autenticitat de les representacions. Aquest informe inclou una descripció detallada de tots dos projectes i analitza els mètodes per integrar-los en una solució de preservació digital completa basada en les normes d'arxiu i preservació digital disponibles actualment, com, per exemple, OAIS, EAD, PREMIS, METS i ANSI/NISO Z39.87.

RESUMEN

En el 2006, los Archivos Nacionales Portugueses (Dirección General de Archivos Portugueses) empezaron a desarrollar un repositorio digital compatible con el modelo OAIS para la preservación de material digital a largo plazo. Paralelamente, la Universidad de Minho emprendió un proyecto denominado CRiB con la finalidad de desarrollar un amplio conjunto de servicios creados para contribuir a la preservación digital, como convertidores de formato y herramientas de evaluación de la calidad, planes de preservación digital y producción automática de metadatos para mantener la autenticidad de las representaciones. Este informe incluye una descripción detallada de los dos proyectos y analiza los métodos para integrarlos en una solución de preservación digital completa basada en las normas de archivo y preservación digital disponibles actualmente, como, por ejemplo, OAIS, EAD, PREMIS, METS y ANSI/NISO Z39.87.

ABSTRACT

In 2006 the Portuguese National Archives (Directorate-General of the Portuguese Archives) engaged in the development of an OAIS compatible digital repository system for long-term preservation of digital material. Simultaneously, at the University of Minho a project called CRiB was being devised which aimed at the development of a wholesome set of services to aid digital preservation. Among those services were format converters, quality-assessment tools, preservation planning and automatic metadata production for retaining representations' authenticity. This paper provides a detailed description of both projects and discusses how these may be integrated into a complete digital preservation solution based on currently available archiving and preservation standards, e.g. OAIS, EAD, PREMIS, METS and ANSI/NISO Z39.87.

RÉSUMÉ

En 2006, les archives nationales du Portugal (Direction générale des archives du Portugal) ont commencé à développer un système numérique d'accès aux archives compatible avec le modèle OAIS en vue de préserver à long terme le matériel numérique. Parallèlement à ce travail, l'Université du Minho a lancé un projet dénommé CRiB dans le but de développer un large éventail de services visant à contribuer à la préservation numérique, par exemple des convertisseurs de format et des outils pour l'évaluation de la qualité, la planification de la préservation et la production automatique de métadonnées permettant le maintien de l'authenticité des représentations. Ce rapport inclut une description détaillée des deux projets et analyse les méthodes utilisées pour les intégrer dans une solution de préservation numérique complète basée sur les normes d'archivage et de préservation disponibles actuellement, notamment OAIS, EAD, PREMIS, METS et ANSI/NISO Z39.87.