

La preservació digital segons la tipologia

Alice Keefer Riva

PID_00190063

Índex

1. Les dades científiques	5
1.1. Objectius	5
1.2. Característiques	7
1.3. <i>Digital data curation</i>	7
1.4. Accions	9
1.5. Responsables	10
1.5.1. Gestors de dades	12
1.5.2. Editorials científiques	12
1.5.3. Agències subvencionadores	13
1.5.4. Universitats i les seves biblioteques	14
1.5.5. Els investigadors	16
1.6. Certificació de bones pràctiques	16
2. Material en web	18
2.1. Els reptes principals	18
2.2. Les iniciatives de les biblioteques nacionals	19
2.2.1. El treball col·laboratiu	20
2.2.2. Experiències a Espanya	21
2.3. Preservació web en l'àmbit universitari	24
2.4. La preservació web i la Comissió Europea	26
3. Les noves formes de comunicació en la Xarxa	27
3.1. Web 2.0: característiques i elements	27
3.1.1. La tecnologia del Web 2.0	28
3.1.2. Web 2.0 a les biblioteques	29
3.2. Nous reptes per a preservar informació generada en les xarxes socials	29
3.3. Importància del contingut de les xarxes socials	30
3.4. Aspectes legals de la preservació del Web 2.0	32
3.5. La preservació dels missatges de <i>Twitter</i>	32
3.6. La preservació de les xarxes socials i la Comissió Europea	33
4. Els objectes d'aprenentatge digitals	35
4.1. Definicions	35
4.2. Els objectes d'aprenentatge: propòsits, característiques i elements	36
4.2.1. Els propòsits	36
4.2.2. La creació de material d'aprenentatge digital	37
4.2.3. Característiques i elements	38
4.3. La preservació dels materials d'aprenentatge: la postura dels <i>stakeholders</i>	39

4.3.1.	Arguments a favor de la preservació d'objectes d'aprenentatge	39
4.3.2.	Arguments en contra de la preservació d'objectes d'aprenentatge	39
4.3.3.	Actituds dels autors	40
4.3.4.	Polítiques institucionals a favor de la preservació	41
4.4.	La qüestió dels drets	41
4.4.1.	Els drets intel·lectuals	41
4.4.2.	Els drets dels estudiants	42
4.5.	La gestió dels materials preservats	43
4.5.1.	Propietats significatives	43
4.5.2.	Els objectes d'aprenentatge considerats com a material d'arxiu	44
4.5.3.	L'adopció de tecnologies Web 2.0 per als materials d'aprenentatge	45
Bibliografia		47

1. Les dades científiques

En aquest apartat es parlarà de les dades generades mitjançant un procés de recerca i que prèviament no tenien sortida dintre de la cadena editorial tradicional. Si l'objectiu inicial dels dipòsits institucionals (comentats en el mòdul "La preservació digital des de la perspectiva tècnica" a l'apartat "Els dipòsits institucionals i la preservació") se centrava sobre tot a facilitar la gestió i l'accessibilitat de les publicacions –articles, *e-prints*, etc.– que resulten de les activitats investigadores, ha anat en augment l'interès en l'accés garantit també a les dades científiques –tant les dades en brut (*raw data*) com les elaborades– que es recullen o es generen durant el procés de la recerca. A diferència dels articles i altres documents "publicables", aquestes dades no tenien sortida mitjançant la cadena editorial tradicional: és a dir, no es publicaven en les revistes científiques.

L'any 2011 un document europeu especifica que el concepte d'informació científica cobreix tant els articles publicats en les revistes científiques mitjançant el procés d'avaluació *peer review* (avaluació d'experts), com les "dades científiques" (*research data*), que la Comissió Europea descriu de la manera següent:

"«Research data» or «data» may be numerical/quantitative, descriptive/qualitative or visual, raw or analysed, experimental or observational. Examples are digitised primary research data, photographs and images, films, etc." ("Consultation of scientific information in the digital age", 2011).

En una anàlisi de la pràctica habitual en vuit disciplines diferents, Swan i Brown (2008) observen tres tipus de dades, segons l'origen:

- d'experiments científics,
- de models i simulacions,
- d'observacions de fenòmens específics fetes en un temps i un lloc determinats.

1.1. Objectius

L'arrel de l'interès recent en les dades científiques és la possibilitat de compartir-les actualment (*data sharing*) i de reutilitzar-les (*reusability*) en el futur. A grans trets, els beneficis que s'obtenen són l'avenç científic i la transparència del procés investigador: l'accessibilitat de les dades servirà per a altres investigadors, per a la seva nova recerca i també perquè es confirmen, o no, la metodologia i les conclusions de la recerca original: "*traceability and repeatability*", en paraules del Consell de la Unió Europea.

L'interès en la **disponibilitat** i "**reusabilitat**" de dades científiques va més enllà dels desitjos dels mateixos investigadors i de les seves institucions o col·lectius. També s'hi apunten les agències de finançament de la investigació, les editorials científiques, els governs nacionals i els organismes internacionals com veurem tot seguit. Per una banda, hi ha el reconeixement de la importància de la investigació per a la innovació científica i tecnològica, i per l'altra, hi ha la preocupació pel cost de la recerca (generalment pagat per les arques públiques). Per tant, els organismes –tant públics com privats– que financen la investigació busquen l'estalvi de la reutilització de dades amb vista a projectes de recerca en el futur. JISC, un d'aquests organismes, afirma la necessitat de:

"fomentar la normalització en la creació, tramesa, ús i reutilització [reuse] dels recursos digitals creats pel sector públic i per al bé públic [...] per aconseguir el millor rendiment de les inversions [i] evitar duplicació dels esforços..."

L'any 2010 es va publicar un informe encarregat per la Comissió Europea sobre les dades científiques: *Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data*. Aquest document ressalta tant les oportunitats com els perills de la situació actual. D'una banda, l'informe fa una crida a les entitats europees per a treure profit de les oportunitats presentades per "l'onada" de dades científiques disponibles. De l'altra, el missatge inclou l'avís que si no s'arriba a controlar aquesta allau de dades, es podria convertir en un "tsunami" capaç d'alterar el món de la informació tal com el coneixem¹.

Tot i que aquest informe no estava dedicat a la preservació, el vincle amb la matèria d'aquesta assignatura és evident: les dades científiques s'han de conservar per a posar-les a la disposició de les generacions d'investigadors futures; d'aquesta manera, formaran una mena d'infraestructura: "un bé valuós sobre el qual poden avançar la ciència, la tecnologia, l'economia i la societat" (pàg. 4). Per tant, complir aquesta missió –la creació i manteniment d'una infraestructura de dades– implica la preservació a llarg termini de les dades per a garantir-ne l'accessibilitat i la usabilitat. De fet, l'informe planteja la qüestió següent molt relacionada amb les tasques de la preservació exposades al llarg d'aquesta assignatura: com podrem preservar totes les dades?

Citació

"How will we preserve the data? As we all have seen, the media in which we store information change constantly –from magnetic coils, to tape, to disk, to USB key, to «cloud» storage, and so on in an endless chain of invention and obsolescence. What will be the point of storing all this scientific data if, a century from now, it has degraded, been corrupted, or is simply too difficult for anyone but a well-equipped expert to use?" (*Riding the wave...*, pàg.16).

Web recomanat

Podreu consultar:

<http://www.jisc.ac.uk/whatwedo/programmes/contentalliance/contentframework.aspx>

⁽¹⁾Comentari de Sir John Wood, coordinador de l'estudi, en la conferència inaugural del congrés de l'Alliance for Permanent Access, Hèlsinki, 22 novembre de 2010.

1.2. Característiques

Hi ha nombrosos factors que fan que la gestió de les dades sigui diferent de la gestió de materials produïts amb la intenció de ser publicats –com ara els articles–:

- Les dades científiques no són necessàriament intel·ligibles i, per a poder-les tornar a utilitzar, cal tenir informació especialitzada sobre el seu contingut, context, propòsit, funcionament, etc. (cal pensar que un percentatge molt important de les dades produïdes per l'e-ciència són generades per dispositius tecnològics sense cap intervenció directa humana: sensors, satèl·lits, etc.)
- Les dades són molt variades (a diferència dels formats estandarditzats en què surten les publicacions) i poden ser molt complexes. Segons la coordinadora de la coalició holandesa de preservació digital, es tracta del "tipus d'informació digital més complicat de gestionar (*curate*)".
- Les dades poden incloure informació sensible o confidencial –per exemple, les dades personals o les d'interès comercial– que requerirà un nivell de control d'accés més alt que el que s'estableix per als articles en els quals la informació sensible s'anonimitza o s'elimina abans de ser publicada.

1.3. Digital data curation

En els últims anys, el terme *data curation* (o *digital curation*) es fa servir en el context de la gestió a llarg termini de les dades científiques en lloc de *data preservation*. El concepte de *data curation*, que ja existia en els entorns tradicionals abans d'esdevenir d'ús quasi exclusiu de dades digitals per a la recerca, es refereix al 'procés de seleccionar, normalitzar, anotar i integrar' les dades recollides de diferents fonts, com ara revistes, informes o bases de dades creades per tercers, amb la finalitat d'incorporar-les en una base de dades especialitzada en què es mantenen "actualitzades i pertinents" (Ball, 2010b, pàg. 5).

Un informe sobre *data curation* publicat per la Cornell University (una altra institució molt rellevant en el camp de la preservació digital) avisa sobre la possible confusió:

"In some contexts curators add meaning to data by means of annotations or linking to related data, further adding to the possible confusion surrounding this term" (Digital research data curation, pàg. 33).

A continuació s'ofereixen definicions més recents de *data curation* i de *digital curation*. En alguns casos encara es mantenen les responsabilitats del significat tradicional, és a dir, les d'afegir valor a les dades, de validar-les o de valorar-les (*appraisal*) (vegeu les paraules ressaltades). Però principalment fan referència

Consulta recomanada

Inge Angevaare (2009). "Taking care of digital collections and data: «Curation» and organisational choices for research libraries". *Liber Quarterly* (vol. 19, núm. 1, pàg. 6-7).

Vegeu també

Tal com vam veure en el mòdul "La preservació digital desde la perspectiva tècnica" a l'apartat "Dipòsits digitals fiables", el control curós de les autoritzacions i l'accés s'inclou dintre dels criteris perquè els repositoris es puguin considerar "fiables".

a les activitats relacionades amb la preservació digital, tal com s'han exposat al llarg d'aquesta assignatura. (noteu que algunes consideren la "preservació" una part del procés de *curation*):

- **Data curation:** "Activity of managing the use of data from its point of creation to ensure it is available for discovery and re-use in the future" (Digital research data curation, pàg. 33).
- "Digital curation, broadly interpreted, is about maintaining and **adding value** to a trusted body of digital information for both current and future use: in other words, it is the active management and **appraisal** of digital information over its entire life cycle" (Pennock, 2007).
- "Digital curation is the management and preservation of digital data over the long-term [including] ... planning its creation, best practice in digitisation and documentation, and ensuring its availability and suitability for discovery and re-use in the future [...]. Digital curation can also include managing vast data sets for daily use, for example ensuring that they can be searched and continue to be readable" (Daisy Abbott, 2008. "What is digital curation").
- "Digital curation involves maintaining, preserving and **adding value** to digital research data throughout its lifecycle".
- "Scientific data curation is a means to collect, organize, **validate** and preserve data so that scientists can find new ways to address the grand research challenges that face society" (*Data Conservancy*).

Beagrie, Lavoie i Wollard (2010, pàg. 55) resumeixen els avantatges de la preservació d'aquestes dades en el quadre següent:

Direct benefits	Indirect benefits (costs avoided)
New research opportunities	No re-creation of data
Scholarly communication/access to data	No loss of future research opportunities
Re-purposing and re-use of data	Lower future preservation costs
Increasing research productivity	Re-purposing data for new audiences
Stimulating new networks/collaborations	Re-purposing methodologies
Knowledge transfer to industry	Use by new audiences
Skills base	Protecting returns on earlier investments
Increasing productivity/economic growth	
Verification of research/research integrity	
Fulfilling mandate(s)	

Web recomanat

Podeu consultar:
www.dcc.ac.uk/digital-curation/what-digital-curation

Near term benefits	Long-term benefits
Value to current researcher & students	Secures value to future researchers & students.
No data lost from Post Doc turnover	Adds value over time as collection grows and develops critical mass
Short-term re-use of well curated data	
Secure storage for data intensive research	
Availability of data underpinning journal articles	
Private benefits	Public benefits
Benefits to sponsor /funder of research/archive	Input for future research
Benefits to researcher	Motivating new research
Fulfil grant obligations	Catalysing new companies and high skills employment
Increased visibility/citation	
Commercialising research	

1.4. Accions

La relació següent identifica les tasques relacionades amb *data curation*, entre les quals es combinen accions relacionades amb la preservació amb altres que tenen a veure amb la definició tradicional, és a dir, afegir valor, validació de les dades, etc. (Ball, 2010b, pàg. 5-6):

- "Selection of datasets to curate.
- Bit-level preservation of the data.
- Creation, collection and bit-level (or hard-copy) preservation of metadata to support contemporaneous and continuing use of the data: explanatory, technical, contextual, provenance, fixity, and rights information.
- Surveillance of the state of practice within the research community, and updating of metadata accordingly.
- Storage of the data and metadata, with levels of security and accessibility appropriate to the content.
- Provision of discovery services for the data; e.g. surfacing descriptive information about the data in local or third-party catalogues, enabling such information to be harvested by arbitrary third-party services.
- Maintenance of linkages with published works, annotation services, and so on; e.g. ensuring data URLs continue to refer correctly, ensuring identifiers remain unique.
- Identification and addition of potential new linkages to emerging data sources.
- Updating of open datasets.
- Provision of transformations/refinements of the data (by hand or automatically) to allow compatibility with previously unsupported workflows, processes and data models.
- Repackaging of data and metadata to allow compatibility with new workflows, processes and (meta)data models."

1.5. Responsables

Qui s'encarrega (o qui s'hauria d'encarregar) de la gestió de les dades científiques? Com es fàcil d'imaginar, són els mateixos col·lectius de *stakeholders* que s'han introduït en els mòduls anteriors. El projecte europeu PARSE.Insight (2008-2010) va dedicar els seus esforços a analitzar la problemàtica de l'accés permanent a la producció científica (*records of science*) generada per la recerca a Europa. L'informe final identifica els quatre *stakeholders* següents implicats en la cerca de solucions per a la preservació de les dades científiques:

- els investigadors,
- el gestors de les dades,
- les editorials i
- les agències que financen la recerca.

Per la seva part, la National Academy of Science nord-americana va agrupar els mateixos col·lectius d'una manera diferent i, a més, va afegir un nou grup, el públic. Aquests són els *stakeholders* que, segons el seu informe, haurien de repartir la responsabilitat –de la manera que s'assenyala en el gràfic– per a assegurar la integritat, accessibilitat i gestió curosa de les dades de recerca:

- investigadors
- institucions i biblioteques d'investigació (universitats, etc.)
- subvencionadors de la recerca
- organitzacions professionals

- revistes científiques
- públic

Recomanació	Investigadors	Institucions de recerca i biblioteques de recerca	Patrocinadors de recerca	Societats professionals	Revistes	Públic
Integritat de les dades						
Gestionar projectes per a garantir la integritat de les dades	✓					
Rebre la formació adequada per a la gestió de dades	✓					
Participar en el desenvolupament d'estàndards professionals per a la gestió de dades	✓		✓	✓	✓	
Oferir suport per a la formació en gestió de dades		✓	✓	✓		
Reconèixer la idoneïtat del suport financer per als professionals de les dades			✓			
Ajudar a garantir que les aportacions dels professionals de les dades es reconeixin i es recompensin	✓	✓	✓	✓	✓	
Accés a les dades i la seva compartició						
Assegurar que siguin accessibles de manera oportuna les dades i altres informacions rellevants per als resultats.	✓					
Determinar si les dades són d'accés públic i, si no ho són, veure si les restriccions són adequades	✓					
Si no es pot accedir a les dades, explicar públicament per què	✓					
Si cal, desenvolupar estàndards per a l'accessibilitat de les dades (amb les parts adequades segons el camp)	✓	✓	✓	✓	✓	✓
A menys que les restriccions siguin justificades, exigir que les dades estiguin disponibles		✓	✓		✓	
Fomentar la compartició de dades mitjançant el reconeixement públic dels esforços més destacats i mitjançant el finançament de les polítiques i publicació		✓	✓	✓	✓	
Definir unes polítiques clares per a la gestió de les dades de recerca i el seu accés		✓				

Font: <http://www.nap.edu/openbook/12615/png/116.png>

Una iniciativa europea de molt de relleu és l'Alliance for Permanent Access to the Records of Science (APA), els membres de la qual representen diversos grups de *stakeholders*. L'APA té la missió de desenvolupar un marc per a una infraestructura institucional sostenible per a l'accés permanent a la informació científica. Entre els membres actuals es troben algunes organitzacions, ja esmentades en el material docent de PRID, que fins ara han tingut un paper important en el camp de la preservació digital, com ara JISC, Digital Preservation Coalition, British Library i la Biblioteca Nacional de Holanda. La relació següent agrupa els membres segons el seu col·lectiu:

- biblioteques nacionals: les de la Gran Bretanya, Alemanya i Holanda.
- universitats: representades per JISC (Regne Unit), NESTOR (Alemanya) i DANS (Holanda).

- biblioteques d'investigació: representades per LIBER (Ligue de Bibliothèques Européennes de Recherche).
- Col·lectius d'investigadors, organitzacions productores d'informació científica o els seus representants: CERN, ESA Centre for Earth Observation, CSC-IT Centre for Science, Finnish Academies of Science and Letters, European Science Foundation, Science and Technology Facilities Council, Helmholtz Association of German Research Centres.
- Editorials científiques: representades per la seva associació professional, la International Association of STM Publishers.

A continuació s'exposen algunes iniciatives que aquests col·lectius han portat a terme o, en el cas dels investigadors, dirigides a ells per aconseguir que s'impliquin en el procés de la gestió a llarg termini de les dades digitals.

1.5.1. Gestors de dades

Els gestors de dades que actualment gestionen dades científiques digitals es poden dividir en:

- **Arxius de dades nacionals:**
 - Australian National Data Service (solament guarda les metadades de les dades).
 - UK Data Archive.
 - Holanda. Data Archiving and Networked Services (DANS).
- **Arxius de dades temàtics** (sovint relacionats amb organitzacions professionals de la recerca):
 - GenBank.
 - SIMBAD Astronomical Database.
 - Worldwide Protein Data Bank.
- **Repositoris universitaris:**
 - Cornell University, DataStaR.
 - Oxford, Databank (projecte pilot).
 - University of Edinburgh, DataShare.

1.5.2. Editorials científiques

El paper de les editorials actualment és procurar que els autors es comprometin a dipositar les dades fruit de la seva recerca en arxius accessibles per a altres investigadors. Per tant, moltes suggereixen els arxius més adequats.

A continuació es reproduceix la política de la revista multidisciplinària nord-americana *Science*, en la qual s'indiquen els nombrosos tipus d'arxius aprovats per l'editorial, segons la matèria i el tipus de dades que cal dipositar:

"Approved databases are the Worldwide Protein Data Bank ... BioMag Res Bank, and Electron Microscopy Data Bank (MSD-EBI; [and])

- Synthetic molecules: the Cambridge Crystallographic Data Centre;
- DNA and protein sequences: Approved databases are GenBank or other members of the International Nucleotide Sequence Database Collaboration (EMBL or DDBJ) and SWISS-PROT;
- Microarray data: Data should be presented in MIAME-compliant standard format. Approved databases are Gene Expression Omnibus and ArrayExpress;
- Climate data: Data should be archived in the NOAA climate repository or other public databases."

A més de recomanar aquests arxius disciplinaris, *Science* és de les poques revistes que també accepten arxivar dades que no encaixin amb cap arxiu de dades per assegurar-ne la disponibilitat per als seus lectors:

"Large data sets with no appropriate approved repository must be housed as supporting online material at *Science*..."

1.5.3. Agències subvencionadores

Les agències de finançament estan molt ben situades per a obligar els investigadors a prendre mesures per a assegurar l'ús futur i la conservació de les dades que s'hagin creat durant el procés de recerca. Per tal de protegir la seva inversió en la recerca, moltes ja requereixen una planificació des del moment de la sol·licitud de la subvenció.

Per exemple, a partir de gener del 2011 la National Science Foundation, una de les agències principals de finançament de la investigació nord-americana, obliga l'equip investigador a presentar, juntament amb la seva sol·licitud, un pla de gestió de dades que expliqui com es proposa adherir-se a la política de l'NSF amb aspectes, entre d'altres, com els següents:

- "the standards to be used for data and metadata format and content ...;
- policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- plans for archiving data, samples, and other research products, and for preservation of access to them."

L'any 2011, el JISC i la Strategic Content Alliance van publicar un informe sobre què poden fer les agències de finançament de la recerca per a intentar que siguin accessibles i utilitzables en el futur els materials digitals recollits o

Web recomanat

Podeu consultar:

http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml

Web recomanat

Podeu consultar:

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

generats durant els projectes de recerca (sobretot, les dades científiques i altres recursos complexos): *Funding for sustainability: How funders' practices influence the future of digital resources* (Maron, Loy, 2011).

Els autors de l'informe van fer enquestes a agències de finançament a Europa i Nord-Amèrica per conèixer, entre altres factors, les estratègies que empraven per a obligar o animar els investigadors a mantenir accessibles després del projecte els recursos digitals generats amb el seu finançament. En aquest cas, no es limitaven només a les dades de recerca sinó que incloïen altres recursos complexos i interactius com, per exemple,

"iniciatives de xarxes socials que depenen del creixement i manteniment del contingut afegit pels participants i recursos educatius en línia que es basen en la generació continuada de materials d'aprenentatge pels professors i per altres".

L'informe ofereix el quadre següent per a facilitar el càlcul dels recursos econòmics i institucionals necessaris per a garantir el compliment dels objectius de sostenibilitat d'aquests recursos digitals després de la finalització dels projectes:

Framework for post-grant sustainability planning for digital resources

Definir l'impacte després de la subvenció					
Components de la sostenibilitat després de la subvenció					
	Requisits tècnics	Contingut	Accés i descobriment	Audiència i impacte	Dotació de personal de la iniciativa en curs
Definir els objectius Quins elements s'haurien de mantenir per tal d'assolir l'impacte volgut després de la subvenció? Cal tenir en compte tots els components, ja que els objectius de cada projecte determinaran les seves necessitats.	Què necessitarà el recurs per a la conservació a llarg termini, l'emmagatzematge, l'espai de servidor i la migració a nous formats?	El recurs exigirà una actualització editorial continuada, contingut nou, metadades noves o altres millores?	L'equip del projecte té la voluntat o l'obligació d'oferir un accés obert?	Com defineix el projecte els seus objectius en termes de la seva arribada a una audiència? (O bé: a quanta audiència vol arribar?, i a quina mena d'audiència?, i quins impactes vol causar sobre aquesta audiència?)	Què cal per a mantenir el recurs: tot l'equip del projecte, coneixement en una àrea concreta?
Identificar les activitats Quines activitats continuades caldrà fer per a assolir els objectius anteriors?	Manteniment periòdic i feina dedicada a actualitzar el maquinari i el programari	Feina (del personal o de voluntaris) per a desenvolupar i actualitzar el contingut i les metadades	Manteniment i actualització de la interfície d'usuari i de l'eina de cerca i descobriment: optimització de cercadors	Desenvolupament i execució de plans de projecció o màrqueting	Plans de successió per al lideratge; formació del personal, retenció; contractació de personal nou, experts o voluntaris segons calgui
Determinar els costos Quins recursos es necessitaran com a suport per a aquestes activitats?	Aquí caldria incloure-hi els costos directes i les necessitats de recursos no financers, com el treball de voluntaris i els serveis en espècie.				
Construir el pla d'ingressos D'on obtindrà el projecte els recursos que necessita per a cobrir els costos?	Aquí caldria incloure-hi un pla que abordi totes les fonts d'ingressos possibles, com ara l'ajuda directa i indirecta de la institució amfitriona, els potencials ingressos obtinguts i les subvencions, les donacions, etc.				

1.5.4. Universitats i les seves biblioteques

Les universitats estan interessades en la continuïtat de l'accés i la usabilitat de les dades: els seus investigadors són, a la vegada, generadors i consumidors de dades. D'altra banda, alguns repositoris institucionals –generalment sota la responsabilitat de la biblioteca universitària– comencen a acceptar, a més

a més, altres resultats com articles, les dades de la recerca, etc. Els exemples són els repositoris DataStaR de la Cornell University, Databank de la Oxford University, i DataShare de la University of Edinburgh, esmentats més amunt.

A més, les biblioteques universitàries estan ben posicionades per a donar suport als investigadors en tot el procés. Nicholson i Bennett (2011) observen que l'ajuda prestada als investigadors pel dipòsit de les seves dades en els arxius o repositoris més adequats, també serveix per a potenciar la reutilització d'aquestes dades entre diferents col·lectius de consumidors: investigadors, professors, estudiants i altres. A més, com que els repositoris institucionals són responsabilitat de les biblioteques, els consells que donen als autors sobre la preparació de les dades segons les pautes establertes també contribueixen a reduir el cost d'incorporar-les i mantenir-les.

Tal com s'ha comentat, actualment moltes agències de subvenció exigeixen als investigadors que en el moment de demanar una subvenció presentin un pla que prevegi la gestió de les dades en acabar el projecte i el cost relacionat. Nombroses biblioteques universitàries ja ofereixen suport als investigadors per omplir aquests formularis, per exemple:

- Johns Hopkins University
- University of Michigan
- University of Virginia

Un exemple d'iniciativa cooperativa entre biblioteques universitàries és el Data conservancy (DC), un projecte finançat per la National Science Foundation dels Estats Units i coordinat per la biblioteca de la Johns Hopkins University. La seva missió és "construir una infraestructura per a les dades de recerca, per a gestionar el volum en creixement continu de la informació en format digital generada per a la docència i la recerca". La infraestructura consistirà en un repositori en què "les dades seran dipositades de forma sistemàtica i ben documentades i des d'on podran ser fàcilment consultades i analitzades tant per especialistes com per persones no expertes, que siguin accessibles lliurement i al mateix temps protegides de manera adequada i preservades de manera fiable". Aquesta organització reconeix explícitament la importància de les biblioteques en la sostenibilitat de la infraestructura proposada per a les dades de recerca i afirma la seva predisposició a donar-los suport per a poder portar a terme aquest nou paper:

"Libraries as Cornerstone for Sustainable Infrastructure - DC will represent a blueprint for research libraries in the data age. DC will bolster the resource base and capacity of research libraries toward data curation. In cases where research libraries may not have local capacity, DC could provide services on a fee basis thereby providing another revenue stream."

A continuació es pot constatar la importància que dóna el DC a les activitats de la preservació digital. Noteu sobretot el seu reconeixement del paper de l'OAIS:

"**Data Preservation** - DC architecture maps onto the Open Archival Information System (OAIS) reference model for digital preservation and thus allows the DC infrastructure to ingest data through multiple modes.

Preservation entails preparing content with appropriate representation information, context, metadata, and fixity such that someone other than the producer of the data can access, use, and properly interpret them. It also ensures long-term preservation of data through not only the bits, but also format and media migrations, or other actions consistent with an overall policy framework. DC's approach should allow for remote curation of data, use of preservation services from other providers, and flexibility as research results inform technical future infrastructure development."

Web recomanat

Podeu consultar:
<http://dataconservancy.org/objectives>

1.5.5. Els investigadors

Serà molt difícil garantir la reusabilitat de les dades sense la col·laboració individual dels investigadors. Es per això que els diferents *stakeholders* fan esforços per recomanar, aconsellar i fins i tot obligar els investigadors a generar les dades segons les pautes i a dipositar-les en arxius estables i accessibles per a altres científics.

A més dels exemples que ja s'han exposat en aquest apartat d'editorials, agències i biblioteques universitàries, cal esmentar una publicació de l'any 2011 que orienta els investigadors sobre les "millors pràctiques" per a gestionar i compartir les dades científiques: *Managing and Sharing Data: Best Practice for Researchers*. El text posa en relleu els avantatges de dipositar les dades en un arxiu especialitzat; els avantatges pel que fa a la preservació a llarg termini (1KADATA Archive, pàg. 6) són els següents:

- L'adherència a les normes de qualitat establertes.
- La preservació a llarg termini de les dades en formats normalitzats i accessibles, que inclou la conversió de formats quan sigui necessari per causa de l'evolució del programari, etc.
- La conservació de les dades en un entorn segur, amb controls d'accés si fos necessari.
- Les realització de còpies de seguretat periòdiques.

1.6. Certificació de bones pràctiques

El Data Seal of Approval és una organització formada per institucions que gestionen dades de recerca –entre aquestes UK Data Archive i NESTOR– i que està coordinada pel Data Archiving and Networking Service (DANS) holandès. Aquesta organització va elaborar setze pautes –*Data seal of approval guidelines*– que serveixen per a obtenir una certificació de bones pràctiques en la creació, gestió i ús de les dades. Les responsabilitats són repartides entre el productor de les dades, el repositori que gestiona les dades i els usuaris, o consumidors, eventuals de les dades:

Data Seal of Approval Guidelines

- 1) The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.
- 2) The data producer provides the research data in formats recommended by the data repository.
- 3) The data producer provides the research data together with the metadata requested by the data repository.
- 4) The data repository has an explicit mission in the area of digital archiving and promulgates it.
- 5) The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
- 6) The data repository applies documented processes and procedures for managing data storage.
- 7) The data repository has a plan for long-term preservation of its digital assets.
- 8) Archiving takes place according to explicit workflows across the data life cycle.
- 9) The data repository assumes responsibility from the data producers for access and availability of the digital objects.
- 10) The data repository enables the users to utilize the research data and refer to them.
- 11) The data repository ensures the integrity of the digital objects and the metadata.
- 12) The data repository ensures the authenticity of the digital objects and the metadata.
- 13) The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.
- 14) The data consumer complies with access regulations set by the data repository.
- 15) The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and research for the exchange and proper use of knowledge and information.
- 16) The data consumer respects the applicable licenses of the data repository regarding the use of the research data.

2. Material en web

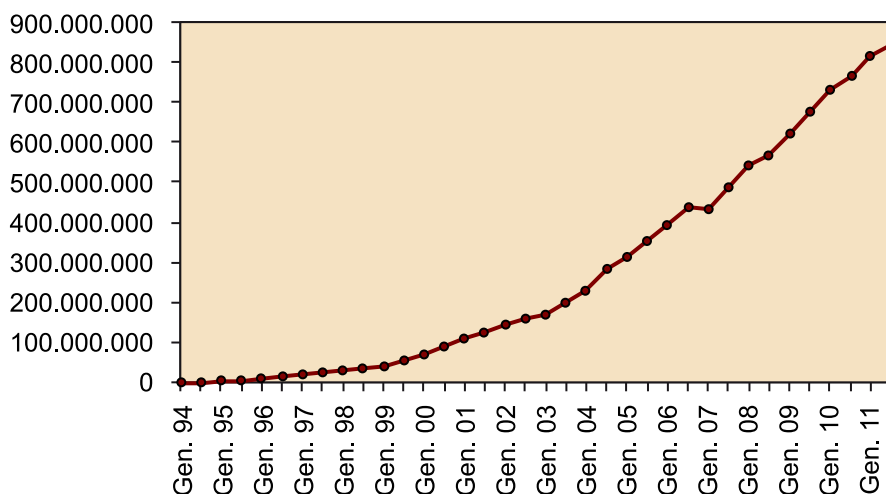
2.1. Els reptes principals

La urgència per a preservar el material del web –o almenys una part d'aquest material– sorgeix principalment de la inestabilitat reconeguda de gran part del contingut. S'ha citat molt el càlcul publicat fa uns anys per l'UK Web Archiving Consortium, segons el qual la vida mitjana d'un lloc web era comparable a la d'una mosca, d'uns quaranta-quatre dies. Aquesta dada s'ha actualitzat recentment i el resultat ha estat una vida mitjana més llarga però no gaire, només de cent dies. És evident que alguns webs són molt estables i podem pensar que aquesta vida mitjana tan reduïda es deu al gran nombre de webs efímers. Però Richard Davis, en un article del 2010, cita un estudi en què s'observa que el 60% dels enllaços a webs de l'Administració pública britànica durant el període 1997-2006 resultaven "trencats", per la desaparició del web o pel canvi sense rastre de l'adreça http.

Al mateix temps hi ha altres factors que dificulten la preservació de molt d'aquest material, entre els quals sobresurten el creixement exponencial que experimenta i l'heterogeneïtat dels tipus de contingut.

El gràfic següent mostra l'augment del nombre de servidors *hosts* que hi ha arreu del món. A principi del 2011 es calculava que n'hi havia uns 850.000.000, la qual cosa representa un increment del 100% en sis anys.

Recompte d'amfitrions segons l'enquesta de dominis d'Internet



Font: <http://www.isc.org/solutions/survey>

Consulta recomanada

Richard Davis (2010). "Moving targets: web preservation and reference management". *Ariadne* (núm. 62).

L'heterogeneïtat també és el resultat de l'avenç tecnològic i respon a factors de subministrament i demanda del mercat o, segons un informe de l'International Internet Preservation Consortium (IIPC), "a l'estira-i-arronsa de les preferències de consumidors, informàtics i venedors de sistemes, i tenen poc a veure amb consideracions de preservació" (M. Davis, 2009, pàg. 5).

Un altre obstacle que dificulta el procés de preservació és la qüestió dels drets d'autor. En la majoria dels països és prohibit capturar i emmagatzemar material i donar-hi accés sense l'autorització dels titulars de la informació. En una enquesta realitzada per a l'International Internet Preservation Consortium, només el 15,6% dels membres enquestats va afirmar que al seu país hi havia una llei que donava suport a la captura i emmagatzematge del material considerat pertinent per aquestes institucions responsables de la conservació del patrimoni cultural nacional (Grotke, 2008, pàg. 13).

2.2. Les iniciatives de les biblioteques nacionals

Sovint el material web va relacionat amb el valor patrimonial i és per això que les activitats principals de preservació del web s'han concentrat en les biblioteques nacionals, moltes de les quals intenten identificar, recollir, emmagatzemar i gestionar el contingut que consideren més rellevant.

La captura de material en web: en general són dues les estratègies que segueixen les biblioteques nacionals per a identificar i capturar material per a preservar-lo, és a dir, l'exhaustiva i la selectiva; a més, hi ha una tercera via, una estratègia híbrida que combina les dues anteriors (Llueca, 2005).

Amb l'**estratègia exhaustiva** un robot recollida (*harvest*) tot el material que compleix uns requisits (per exemple, tot el material sota un domini nacional: .cat o .es). Hi ha altres maneres d'identificar automàticament material en webs d'interès potencial des de la perspectiva patrimonial si aquest s'identifica mitjançant un domini genèric (.edu, .org, .com, etc.): per exemple, la ubicació física del servidor o l'idioma principal del lloc web.

Aquesta estratègia de captura massiva, però, té diversos inconvenients:

- Recull una gran quantitat de material que s'ha de gestionar.
- Recull material que presumiblement té poc valor patrimonial.
- No pot recollir material que no sigui susceptible a l'entrada dels robots, com per exemple informació protegida per contrasenya o incorporada en bases de dades.

- Recull material sense l'autorització dels titulars per a reproduir-lo o per a transformar-lo posteriorment, si no és que hi ha una llei que permeti explícitament aquestes accions a la biblioteca nacional, com és el cas de Nova Zelanda.

Quant a la necessitat de recursos materials i econòmics per a portar a terme el rastreig, emmagatzematge i gestió de quantitats ingents d'informació digital, algunes biblioteques nacionals encarreguen aquesta labor tecnològicament intensiva a altres entitats. Internet Archive, una fundació privada i sense ànim de lucre pionera en la captura de llocs web, presta els seus serveis a institucions tan prestigioses com la Library of Congress, la British Library i, més recentment, la Biblioteca Nacional de Espanya. En el cas de l'arxiu web català, PADICAT, la Biblioteca de Catalunya fa servir la infraestructura de CESCO per a dur a terme la recol·lecció i l'emmagatzematge i gestió posteriors del contingut web català.

Pel que fa a l'**estratègia selectiva**, no ho intenta capturar tot, sinó solament una selecció del material en web que es considera més rellevant. Això es fa mitjançant acords dels productors per al dipòsit o amb el permís per a recollir-la amb un recol·lector automàtic. D'aquesta manera, resol els inconvenients relacionats amb l'estratègia exhaustiva. És a dir:

- Es limita el material a gestionar.
- La selecció es basa en factors de qualitat i pertinència.
- L'acord amb els productors facilita la captura de material no disponible als robots.
- La captura es fa amb l'autorització dels titulars.

La tercera via és la combinació d'aquestes dues estratègies, o sigui, el **mètode híbrid**: la captura exhaustiva de material de webs amb determinades característiques i la selecció curosa mitjançant uns criteris elaborats pel centre i uns acords formals amb els titulars. PADICAT, que es comentarà més avall, és un bon exemple d'un arxiu web que aplica el mètode híbrid.

2.2.1. El treball col·laboratiu

Diversos reptes tècnics, econòmics i legals han conduït a la cerca de maneres de treballar en col·laboració per a compartir coneixements i experiències i contribuir així a un avenç conjunt. Una enquesta a les biblioteques nacionals feta l'any 2007 destaca els aspectes que requereixen una dedicació conjunta per part d'aquestes institucions:

Web recomanat

Podeu consultar:

http://legislation.govt.nz/act/public/2003/0019/latest/DLM191962.html?search=ts_act_national+library_resel&p=1&sr=1

"Posar en pràctica aquest objectiu implica no només qüestions tècniques i de finançament, sinó també qüestions legals com el *copyright*, els materials publicats il·legalment a Internet i la protecció de dades personals, especialment pel que fa a donar accés a materials capturats.

Això requerirà no només una estreta cooperació entre biblioteques nacionals, sinó més aviat col·laboracions en el repartiment dels costos per al desenvolupament tècnic del treball de captura, donant accés, compartint les col·leccions i preservant els materials *online*. Igual d'important és la necessitat de les biblioteques nacionals d'actuar conjuntament per demanar als legisladors el permís per capturar i donar accés als arxius, així com en la negociació amb els propietaris del *copyright*" (Jacobsen, 2007, pàg. 2).

Un dels millors exemples de la col·laboració internacional en qüestions de preservació de material en web és el consorci **International Internet Preservation Consortium (IIPC)**, que es va fundar el 2003 amb dotze membres: les biblioteques nacionals d'Austràlia, Canadà, Dinamarca, Estats Units, Finlàndia, França, Islàndia, Itàlia, Noruega, Regne Unit i Suècia, a més de l'Internet Archive. Actualment hi ha quaranta membres: biblioteques nacionals i universitàries, arxius, museus i altres institucions de patrimoni cultural que han assumit la responsabilitat de preservar material web.

Els seus objectius generals són:

- Facilitar la recollida i la preservació de contingut a Internet del món sencer i l'accés a llarg termini a aquest material.
- Promoure el desenvolupament i l'ús d'eines, tècniques i normes per a la creació d'arxius internacionals.
- Donar suport a les iniciatives i les lleis que contribueixin a la recollida i preservació del contingut i el seu accés a Internet.
- Donar suport a les biblioteques, arxius, museus i institucions de patrimoni cultural arreu perquè puguin plantejar la recollida i preservació de material a Internet.

2.2.2. Experiències a Espanya

Dues institucions espanyoles són membres de l'IIPC: la Biblioteca de Catalunya, que des de 2005 recopila i emmagatzema material web en el seu arxiu Patrimoni Digital de Catalunya (PADICAT) i, des de l'any 2009, la Biblioteca Nacional d'Espanya (BNE).

El PADICAT té la missió de "capturar, preservar i difondre l'arxiu web de Catalunya". En els paràgrafs següents reproduïm dades sobre els objectius, el contingut i l'estratègia del sistema PADICAT, què són a la vegada representatives de la política i les actuacions d'altres membres de l'IIPC:

Web recomanat

L'enllaç següent mostra la llista dels arxius gestionats per membres de l'IIPC. Per qüestions legals, tècniques i/o institucionals, el material web capturat no sempre és accessible:

<http://www.netpreserve.org/about/archiveList.php>

"La Biblioteca de Catalunya, responsable de recollir, conservar i difondre el patrimoni bibliogràfic de Catalunya, i per extensió el patrimoni digital, compta amb la col·laboració tecnològica del Centre de Serveis Científics i Acadèmics de Catalunya (CESCA) per preservar i donar accés a versions antigues de pàgines web publicades a Internet, des de 2005.

Els recursos digitals susceptibles de formar part de PADICAT són els documents culturals, científics i de caràcter general produïts en format digital i

- creats a Catalunya, o
- que contenen informació relativa a Catalunya, o
- d'interès majoritari per al poble català."

Els objectius del PADICAT són:

- Compilar massivament el domini .cat.
- Impulsar el dipòsit sistemàtic de la producció web de les entitats i les empreses de Catalunya.
- Promoure línies de recerca processant de manera monogràfica els recursos d'esdeveniments de la vida pública catalana, com ara campanyes electorals a Internet, el fenomen de la música en línia o els museus a Internet.

Queda explícit el seu compromís amb la preservació del material a llarg termini:

"La preservació digital consisteix en l'aplicació de tècniques i mètodes que permetin garantir que la informació emmagatzemada digitalment en qualsevol tipus de format, programari, maquinari o sistema continuï sent accessible en el futur."

S'empra el mètode híbrid per a identificar material per a ser incorporat al PADICAT, és a dir, la combinació de les estratègies exhaustiva i selectiva:

- La captura sistemàtica de webs sota domini .cat.
- La captura de webs de les institucions amb les quals la Biblioteca de Catalunya ha signat un conveni de col·laboració.
- La captura dels webs que es consideren pertinents a partir de la cerca per navegació.
- La captura dels webs que (comprovada la seva pertinença) s'incorporen a la col·lecció mitjançant les recomanacions dels usuaris.

Com s'esperava, el volum de recursos arxivats en el PADICAT va en augment. Al final de 2011, les més de 44.300 pàgines web emmagatzemades ocupaven uns 9,4 TB i consistien en més de 300.000.000 fitxers.

Les xifres sobre els formats gestionats a PADICAT fins al gener del 2011, que mostrem a continuació, serveixen per a apreciar la complexitat del compromís de mantenir un accés permanent a material en formats variats i en plena evolució.

Tipus	Fitxers	
text/html	204.476.266	81,90%
imatge/pega	22.431.482	8,98%
imatge/GIF	5.402.622	2,16%
Applications/PDF	3.360.105	1,35%
imatge/panga	2.876.602	1,15%
Applications/atom+xml	1.745.564	0,70%
text/xml	1.369.180	0,55%
text/css	1.063.851	0,43%
Applications/rss+xml	1.012.506	0,41%
Applications/javascript	914.282	0,37%
text/plain	895.300	0,35%
Applications/x-shockwave-flash	563.300	0,23%
text/dns	518.623	0,21%
no-type	363.638	0,15%
Applications/x-javascript	354.966	0,14%
Applications/octet-stream	269.399	0,11%
Applications/xml	226.696	0,09%
Applications/msword	211.210	0,08%
Image/tiff	188.378	0,08%
Image/pjpeg	164.580	0,07%
Altres	1.252.363	0,50%

Font: <http://www.padicat.cat/ca/coneix-nos/webs-capturats>

Quant a la previsió de creixement, el web del PADICAT explica:

"[...] a partir del 2012 es persegueix sistematitzar la capacitat de creixement, amb la fita d'incorporar anualment unes 75.700 versions d'aproximadament 32.000 pàgines web, procedents de:

- Compilació semestral de 30.000 recursos del domini .cat.
- Compilació semestral de 550 recursos de les 450 entitats amb què s'ha arribat a un conveni de cooperació.
- Compilació semestral dels 800 recursos procedents de recomanacions dels usuaris.
- Compilació única de 1.000 recursos de col·leccions monogràfiques.
- Compilació diària d'una part substancial de 30 publicacions seriades en línia."

El projecte **Preservació digital del domini .es** de la **Biblioteca Nacional de España** segueix l'estratègia exhaustiva per a la identificació i captura massiva de material web del domini .es, a diferència de l'estratègia híbrida emprada per la Biblioteca de Catalunya. Per a aquesta labor, la BNE disposa de l'experiència i la infraestructura de l'Internet Archive.

La missió i els objectius s'expliquen a continuació:

"La Biblioteca Nacional de España está trabajando en la **conservación de los productos culturales «nacidos digitales»**, para que puedan convertirse en un patrimonio cultural e intelectual perdurable para las generaciones presentes y venideras. [...] en su compromiso de garantizar el derecho de acceso a la cultura de todos los ciudadanos, ha emprendido la tarea de **recolectar e indizar automáticamente el dominio .es** para que, una vez archivado, pueda ser consultado por nuestros usuarios.

La Biblioteca Nacional de España... tiene entre sus deberes identificar, preservar, conservar y difundir el patrimonio documental español. Reunir el «**Archivo web del dominio .es**», es parte de esta misión.

[...] la Biblioteca Nacional de España aspira a crear un marco propicio para el desarrollo de una estrategia global de **custodia a largo plazo del patrimonio digital** que se produce en España. Para el desarrollo de este proyecto, la BNE cuenta con la colaboración de la organización internacional Internet Archive."

La primera acció que es va prendre va ser un rastreig del domini .es fet per l'Internet Archive el 2009, seguit de tres de complementaris addicionals el 2010. En total, es van recollir 875.000.000 URL i es va emmagatzemar un total de 30 TB de contingut. Per al 2011 i 2012 es programen més captures sobre el domini .es i altres enfocats a les eleccions municipals y autonòmiques. Aquests rastrejos més enfocats es programen sobre una selecció de vint-i-tres adreces de partits polítics, premsa i mitjans de comunicació audiovisual (Pérez Morillo i Muguerza López, 2011).

La recollida automàtica, com s'ha comentat prèviament, comporta riscos legals. En aquest sentit, la postura de la BNE és la següent:

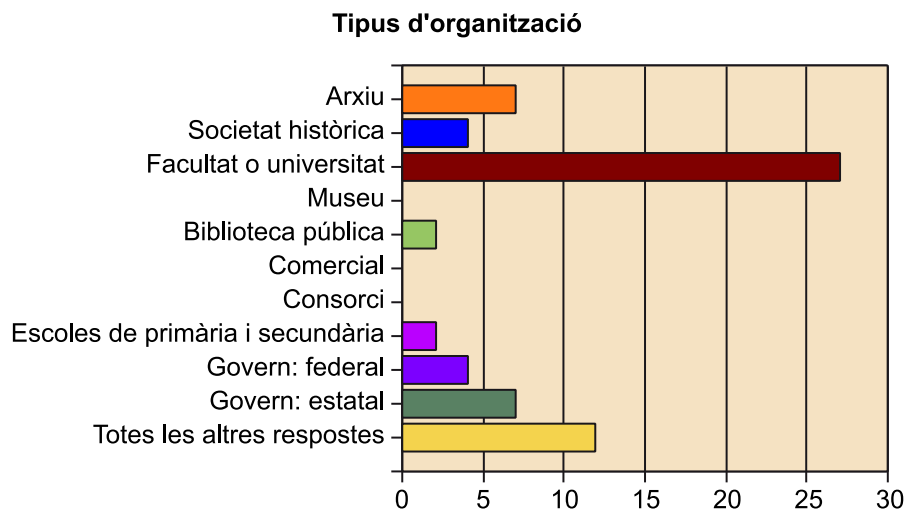
"La Biblioteca Nacional de España no difundirá ningún contenido fruto de esta recolección automática sin el permiso de sus respectivos dueños o de otros garantes legales."

2.3. Preservació web en l'àmbit universitari

Si les iniciatives per a la preservació de material web han estat liderades des de les biblioteques i els arxius nacionals per l'obligació que tenen de preservar el patrimoni nacional, també es detecta un gran interès entre les universitats. En els últims anys, per exemple, dues universitats –Harvard University i University of North Texas– s'han incorporat com a membres de l'IIPC.

Durant el segon semestre de 2011 la National Digital Stewardship Alliance ha dut a terme una enquesta als Estats Units sobre l'estat actual de la captura i gestió de llocs web per part de diferents tipus d'organitzacions més enllà de la Library of Congress (LC) i la National Archives and Records Administration (NARA).

Els resultats finals no estaran disponibles fins a 2012, però els preliminars destaquen l'interès creixent per part de les universitats:



Font: <http://blogs.loc.gov/digitalpreservation/2011/10/early-returns-on-web-survey/>

A diferència de les biblioteques nacionals, la preocupació de les universitats se sol centrar en com preservar el contingut web creat per membres de la seva pròpia institució per a finalitats administratives, docents o d'investigació.

L'any 2008 l'agència britànica JISC va publicar un manual per a la preservació de recursos web: *POWR: the preservation of web resources handbook*. Es tracta d'una recopilació de les bones pràctiques sobre la captura i l'emmagatzematge segur de material web, dirigida sobretot als gestors de webs de les universitats britàniques. El manual destaca la importància de fixar la prioritat del material susceptible de ser preservat i suggereix, com a primer pas, l'agrupació en diferents categories de contingut segons la classificació MSCoW, que representa les inicials en anglès de les paraules *must*, *should*, *can* i *won't*:

- M: els materials que la institució ha de preservar.
- S: els materials que la institució hauria de preservar.
- C: els materials que la institució pot preservar.
- W: els materials que la institució no preservarà.

El manual identifica els tipus següents de material que sovint es troba en els webs institucionals universitaris:

- Documents d'arxiu de la institució i dels seus departaments: aquests sovint tenen requisits legals o institucionals quant a la durada de la seva retenció.
- Materials que afecten els alumnes, com objectes d'aprenentatge digitals.
- Informació generada per l'administració institucional.
- Informació generada per l'activitat d'investigació.
- Informació generada per l'activitat docent.
- Informació generada pels projectes.
- Evidència d'altres tipus d'activitats (per exemple, congressos).

A més d'identificar els tipus i formats dels materials en el web institucional, també suggereix que s'identifiquin:

- Els sistemes que actualment els gestionen.
- El motiu de la seva presència en el web.
- El lloc on s'emmagatzema (servidors, dominis, etc.).

Després del procés d'identificació, té lloc el de valoració i selecció. Però, abans de procedir, el manual recomana que els responsables tinguin un coneixement aprofundit dels factors següents:

- L'estructura i la missió de la institució.
- Les polítiques i els factors que promulguen la preservació.
- Els requisits legals relatius a la retenció de documents d'arxiu.
- L'ús dels recursos web.
- Els *stakeholders* i les seves necessitats.
- El valor potencial dels recursos en el futur (la possibilitat de la seva reutilització).

2.4. La preservació web i la Comissió Europea

El LiWA - **Living web archives** (2008-2011) va ser el projecte dedicat al desenvolupament d'eines per a permetre la captura i l'emmagatzematge de material en web, assegurant que aquest material es mantingués autèntic, coherent i intel·ligible mitjançant el desenvolupament de components i altres eines. El va coordinar la Universitat de Leibniz, a Hannover, Alemanya.

Les aplicacions generades durant aquest projecte estan disponibles en codi obert (*open-source*) en la pàgina *liwa-technologies*.

- *The rich media capture module*, un connector (*plug-in*) dedicat a la captura de contingut de vídeo en temps real (*streaming*).
- *The temporal coherence analyser*, un connector (*plug-in*) dedicat a l'anàlisi de la coherència temporal del contingut web arxivat.
- *The Spam Assessment Interface*, un servei web que permet valorar la qualitat del contingut web arxivat.
- *The semantic analyzer*, un component dedicat a la detecció de l'evolució de la terminologia.
- *The web archive UI framework*, un marc basat en el client que facilita la creació d'interfícies d'usuari per a la navegació del contingut web arxivat.

3. Les noves formes de comunicació en la Xarxa

A l'apartat anterior s'han exposat les iniciatives referents a la preservació del material web, principalment entre les biblioteques nacionals i tot seguit amb la intervenció més recent de les universitats. Aquest apartat pot ser considerat com una extensió de l'anterior, ja que se centra en el contingut creat i tramès per les tecnologies de l'anomenat Web 2.0, que representa reptes nous i, a més, generalment queda fora de l'abast dels sistemes desenvolupats per a la identificació i captura de material en web.

3.1. Web 2.0: característiques i elements

Les definicions del Web 2.0 varien sensiblement. Aquest terme es va començar a utilitzar en els mitjans de comunicació per a descriure les noves aplicacions que començaven a sortir al final de la dècada dels noranta i que es caracteritzaven per la generació i difusió descentralitzades de contingut pels usuaris, en lloc dels creadors o dels proveïdors tradicionals de contingut com són les editorials, institucions, etc. A més, tenen en compte la funcionalitat de compartir dades i informació i el treball col·laboratiu com a característiques bàsiques d'aquesta nova generació d'eines web personals i socials.

La versió catalana de *Viquipèdia* descriu el Web 2.0 de la manera següent:

"Web 2.0: S'associa habitualment amb les aplicacions web que faciliten la compartició interactiva d'informació, el disseny centrat en l'usuari i la col·laboració dins el World Wide Web. Alguns exemples del Web 2.0 serien les comunitats basades en web, les xarxes socials, [...], els wikis i els blocs. Un lloc Web 2.0 permet als seus usuaris interactuar amb altres usuaris o canviar el contingut del lloc; en front dels llocs web no interactius on els usuaris es limiten a mirar passivament la informació que se'ls proporciona" (*Viquipèdia*, 12 de desembre de 2011).

L'any 2007, l'agència JISC va encarregar un informe sobre les tecnologies que permeten compartir contingut en el marc de les universitats britàniques. L'informe va identificar els set tipus de sistemes que formen la categoria de Web 2.0, tot i afirmar que molt sovint alguns d'aquests tipus de sistemes estan agrupats en els serveis actuals, com ara *Twitter* o *Facebook* (Van Harmelen, 2007):

- *Blogs.*
- *Wikis.*
- *Social bookmarking.*
- *Media-sharing services.*
- *Social networking systems.*
- *Collaborative editing tools.*
- *Syndication and notification technologies.*

L'any següent JISC va publicar el manual *PoWR handbook* (comentat en l'apartat anterior), sobre la preservació de material en web en les universitats. L'obra dedica un capítol a qüestions relacionades amb la preservació

d'informació generada amb aquestes aplicacions. A més a més dels set tipus de sistemes identificats en l'informe de Van Harmelen, al manual, se n'hi afegeix un de nou: la **missatgeria instantània**. Tot i que va aparèixer abans de la creació del WWW, els autors argumenten que el protagonisme que ha adquirit aquesta funcionalitat en l'entorn Web 2.0 fa que es mereixi ser inclosa entre les set aplicacions de nova generació identificades.

3.1.1. La tecnologia del Web 2.0

A continuació oferim algunes definicions de les aplicacions més relacionades amb la funcionalitat del Web 2.0:

- **Blog** (també *weblog*): "Un diari accessible en web que conté entrades curtes, que surten en ordre cronològic invers [les més recents, en primer lloc]. Sovint se centren en un tema i s'actualitzen molt sovint. Les entrades normalment tenen enllaços a altres pàgines webs externes..." (OCLC, 2007),
- **Marcadors socials** (*social bookmarking*): "El sistema que permet que els usuaris emmagatzemin enllaços als seus llocs web predilectes i els comparteixin amb altres usuaris, normalment amb la introducció d'etiquetes [*tags*] per part de l'usuari per descriure el contingut" (OCLC, 2007).
- **Library 2.0**: "L'aplicació de tecnologies basades en la web (interactives, col·laboratives i de multimèdia) als serveis i fons de biblioteques disponibles en la web" (Maness, 2006).
- **Llocs socials de multimèdia**: "Pàgines web que permeten que les persones comparteixin contingut que elles mateixes han creat, com per exemple *YouTube* (vídeos) i *Flickr* (fotos)" (OCLC, 2007).
- **Missatgeria instantània**: "La tecnologia que permet «converses» curtes o xats [*chats*] privats que es realitzen en temps real entre dues o més persones. Els usuaris reben informació sobre quins altres usuaris estan en línia quan entren en el sistema" (OCLC, 2007).
- **Podcàsting**: "Distribució de contingut àudio (tant en temps real [*streamed*] com de descàrrega) mitjançant els serveis tipus RSS via Internet" (OCLC, 2007).
- **RSS** (*real simple syndication*): "Un format web que permet la recepció per part dels usuaris de contingut nou de llocs web determinats, mitjançant la descàrrega directa als seus ordinadors" (OCLC, 2007).
- **Sindicació de continguts**: "Un procés que facilita la transmissió d'informació d'un lloc a un altre, on pot ser reutilitzada" (OCLC, 2007).

- **Wiki:** "Un lloc web dissenyat per a l'ús col·laboratiu, que admet les contribucions de nombrosos usuaris per a la seva creació o modificació" (OCLC, 2007).
- **Xarxes socials:** "Llocs web dissenyats principalment per a facilitar la interacció entre usuaris que comparteixen interessos, actituds i activitats, com *Facebook* o *MySpace*" (OCLC, 2007).

3.1.2. Web 2.0 a les biblioteques

El quadre següent mostra com s'ha aplicat la tecnologia del Web 2.0 a les biblioteques (Maness, 2006):

Library 1.0	Library 2.0
Serveis de referència per correu electrònic	Referència per xat
Programes d'aprenentatge textuais	Programes d'aprenentatge multimèdia amb bases de dades interactives
Llistes de distribució per correu electrònic	<i>Weblogs</i> , <i>wikis</i> , RSS
Esquemes de classificació controlats	Assignació d'etiquetes (<i>tags</i>) paral·lelament a l'ús d'esquemes controlats
OPAC	Interfície de xarxa social personalitzada

3.2. Nous reptes per a preservar informació generada en les xarxes socials

En tractar-se de comunicació personal i efímera, segurament es produiran nous reptes per a la identificació, recollida i preservació d'aquest contingut innovador tramès pels nous mitjans. Alguns dels reptes principals que els experts han identificat són els següents:

- **Servidors externs.** Una gran part d'aquest contingut (que inclou material d'interès per a les universitats com ara contingut relacionat amb les activitats d'investigació i docència) no resideix en servidors institucionals, ja que hi ha un ús molt significatiu de serveis de tercers: *YouTube*, *Facebook*, *Twitter*, etc. Per tant, resulta molt difícil identificar el material d'interès institucional, nacional, etc. A més, en el cas dels serveis que funcionen amb contrasenya, el seu contingut queda fora de l'abast dels recollidors que actualment capten una part de la Xarxa (determinats dominis, etc.).
- **Autors múltiples.** Un altre obstacle a la preservació eventual del material identificat i capturat està relacionat amb una de les característiques bàsiques del Web 2.0: el treball en col·laboració i el contingut compartit. Aquest fet complica –i molt– la qüestió dels drets intel·lectuals, ja que no

es tracta de l'obra d'un sol autor o de diversos tots ben identificats, com és el cas dels articles científics.

- **Actualitzacions i modificacions contínues.** L'actualització contínua i íntegra de la informació –i la facilitat per a modificar el contingut ja publicat– fa difícil identificar amb precisió i fiabilitat les versions a diferència, per exemple, de les versions d'un treball d'investigació enviat a una revista científica en què cadascuna és identificable –normalment amb data de revisió– com a obra independent: l'esborrany, la còpia enviada a l'editor, la còpia que inclou les modificacions recomanades per la revisió d'experts, la versió definitiva de l'article tal com surt publicat (per exemple, un PDF). La falta d'estabilitat d'aquestes fonts noves representa un problema sobretot pel que fa al rigor científic. Què passa, per exemple, quan se cita un weblog mantingut col·lectivament si posteriorment no es pot comprovar que existeix la versió esmentada a causa dels canvis experimentats com ara dades afegides, canviades o eliminades?
- **Falta de còpies "mestres".** És evident que tota la informació en web és vulnerable per les raons que s'han explicat al llarg d'aquesta assignatura. Però el contingut generat en les aplicacions de comunicació social ho és encara més. Si tornem a l'exemple del paràgraf anterior, a la cadena editorial digital, l'autor crea el contingut en el propi ordinador, per exemple com a document Word, i posteriorment l'envia a l'editor; a més, possiblement, el dipositarà en el repositori digital de la seva institució en format PDF. És a dir, hi ha còpies. En el cas de les aplicacions socials, la còpia original se sol generar directament dintre de l'entorn informàtic del *host*, fet que sens dubte augmenta el risc de pèrdua.
- **Informació sense context.** Com ja s'ha afirmat, les xarxes socials es caracteritzen per la comunicació entre persones, entitats, etc. Per tant, el context dels missatges i les seves connexions amb altres fonts de contingut seran imprescindibles per a poder-los interpretar en un futur. Però encara sembla impossible resoldre aquest problema. És especialment greu en el cas dels missatges curts de *Twitter*, limitats a cent quaranta caràcters, una extensió semblant a la d'un titular de diari que no és fàcilment comprensible si no es llegeix l'article.

3.3. Importància del contingut de les xarxes socials

Moltes de les aplicacions relacionades amb les xarxes socials donen eines d'autopublicació i comunicació pública als usuaris. Com tots sabem, en molts casos –potser en la majoria– porten informació d'importància patrimonial qüestionable. Fins a quin punt ens poden interessar els detalls de la festa d'aniversari de la filla dels veïns exposats a *Facebook* o comunicats mitjançant *Twitter*?

Però tal com ha observat el periodista nord-americà Christopher Beam citant l'historiador Paul Freedman: "Els historiadors estan interessats en la vida quotidiana i *Twitter* és una font d'informació increïble sobre la vida quotidiana" (Beam, 2010).

El fet que aquestes aplicacions han penetrat en totes les facetes de la vida i que han envaït tots els sectors comercials i socials, fa que ara es reconegui la importància potencial de la informació generada, bé individualment, bé en agregació. Veiem-ne alguns exemples:

- **Entorn polític.** La pàgina web del projecte europeu ARCOMEM (del qual parlarem més avall) observa que alguns ministres del Govern britànic fan servir *Twitter* per a contestar preguntes parlamentàries.
- **Entorn social.** Els participants en les revolucions populars que l'any 2011 van sacsejar el món àrab i que van acabar amb diversos règims polítics, feien servir els nous canals de difusió d'informació per a informar dels esdeveniments i també per a mantenir la comunicació entre les persones i grups involucrats. Un estudi de Lotan i altres va analitzar l'ús de *Twitter* i el flux d'informació entre els diferents participants (els activistes del carrer, els que informaven dels esdeveniments mitjançant els seus weblogs, els periodistes i els mitjans de comunicació tradicional) durant aquests dies caòtics. L'informe va notar que aquests mitjans socials van guanyar encara més protagonisme perquè alguns dels règims intentaven frenar l'accés dels periodistes professionals als llocs dels incidents, censuraven la premsa nacional i anul·laven l'accés a Internet. Per tant, en esdeveniments imprevisibles com aquests, hi ha hagut un canvi transcendental entre els creadors i els usuaris tradicionals de la informació: els mitjans tradicionals arriben a dependre en gran part de la informació generada *in situ* pels "periodistes ciutadans" i divulgada per les xarxes socials.
- **Entorn científic.** En el camp científic, altres autors han analitzat l'ús de weblogs i *Twitter* per a la comunicació entre investigadors. L'expert en els processos de la investigació, Michael Nielsen (citat per Richard Davis), observa:

"És fàcil ignorar l'impacte dels *weblogs* sobre la investigació, ja que la majoria d'ells se centren més en la divulgació dels resultats. [...] Estem experimentant una expansió espectacular de l'abast dels *weblogs* com a mitjans [de comunicació]" (R. Davis, 2010).

Un estudi fet el 2010 identifica *Twitter* com un dels tres serveis principals que fan servir els investigadors de les ciències socials per a intercanviar informació. Específicament, els autors analitzen com els científics l'utilitzen per a comunicar les seves idees en temps real des dels congressos professionals (Letierce i altres, 2010).

Consulta recomanada

Gilad Lotan i altres (2011). "The revolutions were tweeted: information flows during the 2011 Tunisian and Egyptian revolutions". *International Journal of Communication* (núm. 5, pàg. 1375-1405).

3.4. Aspectes legals de la preservació del Web 2.0

Tal com s'ha comentat durant l'assignatura, actualment les lleis sobre drets intel·lectuals –i principalment les restriccions sobre la reproducció i transformació de les obres protegides pels drets d'autor– frenen les iniciatives per a preservar el material digital en web². En un discurs pronunciat a Gijón l'any 2010, Laura Campbell, sotsdirectora de la Library of Congress, exposa aquest problema des de la perspectiva de la seva institució:

⁽²⁾les restriccions sobre la divulgació i l'emissió de les obres afecta la capacitat de les institucions de donar accés al material arxivat

"Les lleis de drets d'autor obsoletes creen el repte principal per a les biblioteques en el seu intent de complir amb la seva missió històrica: col·leccionar, gestionar i preservar el coneixement per donar-li accés a aquesta generació i a les del futur [...].

Quan parlem de contingut a la web oberta com són els *weblogs*, *YouTube*, les xarxes socials com *Facebook* i *Tuenti*, les dificultats augmenten. Tot aquest contingut té el potencial de ser de gran importància històrica. És el resultat de la creació de moltes persones. És un fenomen històric nou: la creació col·lectiva a l'escala actual. Per aquesta raó, el contingut demana a crícs ser preservat. El problema, ens fa l'efecte, és que no podem preservar-lo perquè no hi ha una titularitat clara" (Campbell, 2010).

Tal com queda implícit en aquesta declaració, la solució és canviar la llei per a autoritzar que determinades institucions, com ara les biblioteques nacionals, capturin, gestionin i preservin aquest material. Però la via legislativa és un procés llarg i molts països encara no l'han encetat.

Per exemple, a la Gran Bretanya el canvi del dipòsit legal per a incloure materials digitals en línia es va aprovar l'any 2003. Però el 2011 encara no s'havien aprovat els reglaments necessaris per a implementar la llei, en part per causa de l'oposició del sector editorial que tem que tindria efectes negatius sobre el seu model comercial actual (Corbyn, 2011).

A Espanya la Llei 23/2011, de 29 de juliol, de dipòsit legal estableix la inclusió del dipòsit del material digital en el web, cosa que facilitarà a la Biblioteca Nacional la recol·lecció d'aquest material i que també eliminarà els dubtes actuals sobre les accions que la BNE pot portar a terme per a garantir la preservació d'aquests materials per al futur.

3.5. La preservació dels missatges de *Twitter*

De les aplicacions de la xarxa social esmentades, la més recent i amb un èxit més aclaparador és el servei *Twitter*. Es va iniciar l'any 2006 i actualment passen pel sistema més de cinquanta-cinc mil missatges al dia! Segons *Viquipèdia*:

"Twitter és un servei de *microblogging* que permet als seus usuaris enviar i llegir missatges de text d'una longitud màxima de 140 caràcters (denominats *tweets*, *tuits*, o *piulades*) per mitjà del mateix web de *Twitter*, SMS, missatgeria instantània o bé aplicacions de tercers com [entre d'altres] *Facebook* [...]" (*Viquipèdia*, 12 de desembre de 2011).

Segons estadístiques del 2010, les converses privades representaven el tipus de missatge més freqüent (27%), seguit pels enllaços a notícies i weblogs (10%), política, esports i actualitat (6%), recomanacions i reclamacions de productes (4%) i enllaços a altres continguts web, imatges i vídeos (6%). També s'hi tramen anuncis (4%) i correu brossa o *spam* (4%).

Web recomanat

Podeu consultar: "Twitter Facts & Figures; history & statistics. Website-Monitoring Blog" (2010, abril).

A l'abril de 2010 es va formalitzar la donació de tot l'arxiu de *Twitter*, des del seu començament l'any 2006, a la Library of Congress. La finalitat d'aquesta donació era la preservació del fons i la seva disponibilitat per a la investigació per part de la biblioteca nacional nord-americana. La LC reconeix que *Twitter* serveix de plataforma tant per als mitjans tradicionals (*Reuters*, *New York Times*, etc.), que el fan servir per a divulgar els seus titulars, com per als "periodistes ciutadans". Per tant, "*Twitter* forma part de la comunicació històrica al segle XXI, mitjançant la captura de notícies, esdeveniments i tendències socials." L'aspecte espinós dels drets intel·lectuals sembla menys important en aquest cas, ja que els terminis que accepten els usuaris de *Twitter* li cedeixen el dret d'arxivar i fer públic els seus missatges (Campbell, Dulabahn, 2010).

3.6. La preservació de les xarxes socials i la Comissió Europea

Tal com s'ha vist en els altres apartats d'aquest mòdul, la Comissió Europea es dedica a promoure la preservació dels materials digitals, tant per a assegurar-ne l'accessibilitat en el futur com per a la transferència de coneixements tecnològics entre les diferents esferes sectorials i institucionals a Europa, necessàries perquè es compleixi aquesta missió.

Els seus esforços per a preservar la informació generada pels nous mitjans de comunicació són més recents: els dos projectes comentats a continuació es van aprovar l'any 2011 i, per tant, encara no han produït cap resultat. Però la descripció del que proposen fer il·lustra el grau d'implicació de la CE a favor de la preservació d'aquestes noves vies de comunicació d'interès social, cultural, científic i governamental, actualment i en el futur.

ARCOMEM- Archive Communities' Memories (2011-2013).

Aquest projecte s'adreça a les institucions de memòria cultural (com ara els arxius, museus i biblioteques) per a cercar solucions als reptes que representen les xarxes socials i els canvis radicals que afecten els processos de creació i comunicació d'informació, caracteritzats, sobretot, per la implicació directa dels ciutadans.

El projecte proposa crear eines per a ajudar els responsables a identificar, seleccionar i gestionar la informació que generen les xarxes socials, amb les finalitats següents:

- Mostrar com els mitjans de comunicació socials poden ajudar els arxivistes a seleccionar materials per a preservar, mitjançant eines de valoració basades en les xarxes socials.
- Mostrar com la mineria de dades (*data mining*) dels mitjans de comunicació socials pot enriquir els arxius, procedint a la preservació estructurada al voltant de categories semàntiques.

- Estudiar els mètodes de creació d'arxius, sigui quin sigui el seu origen (comunitat, usuari, etc.).

Tot i que aquest projecte s'adreça a la selecció i ús posterior de la informació, la seva missió es basa clarament en la necessitat de preservar aquests fons. Per exemple, un dels resultats del projecte és la reducció del risc de la pèrdua d'informació efímera irremplaçable.

Una altra mostra de la seva implicació en la preservació d'aquest material és el seu objectiu d'analitzar les diferents maneres d'assegurar la "preservació semàntica", és a dir, com es pot garantir la capacitat d'entendre el contingut mitjançant la preservació del context original.

Blog forever (2011-2013).

La missió d'aquest projecte és crear una plataforma informàtica capaç de preservar, gestionar i difondre el contingut dels weblogs. El seu repte principal és trobar la millor manera de capturar weblogs, que són dinàmics i evolucionen contínuament, juntament amb l'estructura de les seves xarxes i connexions i l'intercanvi d'informació que promouen.

El model que es crearà servirà per a definir una política de preservació dels weblogs que té en compte aspectes com la interoperabilitat i les qüestions de la gestió dels drets intel·lectuals. El model no solament vol preservar el contingut dels weblogs, sinó que també vol que els usuaris del futur puguin entendre l'evolució del context social al llarg del temps. El resultat final serà la producció d'una solució senzilla per a arxivar weblogs en format digital, que n'assegurarà l'autenticitat, integritat, exhaustivitat, usabilitat i accessibilitat a llarg termini.

4. Els objectes d'aprenentatge digitals

4.1. Definicions

- **Objecte d'aprenentatge (*learning object*).** "Una entitat informativa digital desenvolupada per a la generació de coneixement, habilitats i actituds que té sentit en funció de les necessitats del subjecte i que es correspon amb la realitat" (*Viquipèdia*, 16 de desembre de 2011).
- **Aprenentatge virtual (*e-learning*).** "[...] és la manera d'accedir al coneixement mitjançant les Tecnologies de la Informació i la Comunicació (TIC). [...] Dins d'aquestes tecnologies estarien Internet, el multimèdia o els simuladors. D'aquesta manera, s'aconsegueix tenir un alt nivell d'accessibilitat i fa desaparèixer qualsevol barrera espai-temporal i, com a conseqüència, pot fer arribar els continguts a molta més gent" (*Viquipèdia*).
- **Material d'aprenentatge.** "Aquells materials desenvolupats pel personal d'una universitat per a facilitar i donar suport a l'aprenentatge per part dels alumnes:
 - Documents (apunts, exemples, llistes de lectura recomanada).
 - Fotografies.
 - Gravacions àudio i vídeo (presentacions, tutorials, demostracions, actuacions escèniques i pràctiques, tant en format de fitxer com visualitzats en temps real –*podcasts*, etc.).
 - Simulacions, animació i models informàtics.
 - Programari informàtic interactiu.
 - Blocs."

(Emmerson, 2008, pàg. 3.)

- **Entorn virtual d'aprenentatge (*virtual learning environment*).** "És un programari específicament dissenyat per facilitar l'ensenyament-aprenentatge utilitzant sistemes informàtics. En general es tracta de sistemes que treballen en un entorn web i que possibiliten la comunicació entre els diferents subjectes del fet educatiu (entre el professorat, entre l'alumnat i entre el professorat i l'alumnat).
Els entorns virtuals d'aprenentatge permeten al professorat presentar tota mena de materials educatius en format digital (interactius o no), fer el

seguiment de les activitats realitzades pel seu alumnat, intercanviar amb el mateix informacions, consells, etc. i, finalment, portar un registre de les seves activitats i gestió de les qualificacions. L'alumnat, per la seva banda, pot consultar en tot moment els materials dels cursos, comunicar-se amb el professorat i amb la resta de l'alumnat" (*Viquipèdia*, 16 de desembre de 2011).

- **Recursos educatius oberts (*open educational resources*).** "Recursos per a l'ensenyament, l'aprenentatge o la recerca que es troben al domini públic o bé que han estat publicats amb una llicència de propietat intel·lectual que permet que se'n faci un ús lliure. Inclouen cursos complets però també objectes d'aprenentatge, mòduls, llibres de text, vídeos, tests, programari i qualsevol altra eina, material o tècnica que es pugui utilitzar per facilitar l'accés al coneixement.
[...] Els recursos educatius oberts segueixen les lleis de propietat intel·lectual ja que la major part dels recursos educatius estan protegits per drets d'autoria que s'han de respectar. Les llicències Creative Commons [...] esdevenen essencials per al moviment dels recursos educatius oberts ja que faciliten el mecanisme per alliberar certs drets" (*Viquipèdia*, 16 de desembre de 2011).

4.2. Els objectes d'aprenentatge: propòsits, característiques i elements

4.2.1. Els propòsits

El motiu de l'interès recent en la preservació dels objectes d'aprenentatge va estrictament lligat a la possibilitat de poder-los compartir (*share*) i reutilitzar (*reuse*), també per a nous propòsits (*repurposing*). El material pedagògic digital creat dins d'un entorn institucional, per a activitats docents institucionals i amb recursos institucionals, s'hauria de considerar patrimoni intel·lectual institucional. És per aquesta raó que durant l'última dècada les universitats s'han interessat per l'estalvi que representaria poder estendre la vida dels recursos pedagògics mitjançant la seva reutilització per part d'altres membres de la institució. Així es podrien evitar les despeses –en temps i diners– relacionades amb la creació repetida dels mateixos recursos.

Actualment el tema de la preservació de materials d'aprenentatge sovint està relacionat amb les iniciatives dels recursos educatius en obert. Un dels primers exemples és el repositori OpenCourseWare de la universitat nord-americana MIT, posat en marxa el 2002 amb la missió de donar accés gratuït als materials docents de la MIT "per a docents i estudiants d'arreu del món i estendre l'abast i l'impacte tant d'aquest repositori de la MIT com del concepte dels recursos educatius oberts".

Exemple

Un fons fotogràfic amb imatges d'esglésies gòtiques creada per a una assignatura d'arquitectura podria servir per a assignatures d'altres disciplines, sense haver d'invertir en la compra o encàrrec de noves imatges.

A Catalunya el repositori Materials Docents en Xarxa (MDX), gestionat per CBUC i CESCO, dona accés als materials i objectes docents produïts en les universitats participants: nou universitats catalanes (entre elles, la UOC), més la Universitat Jaume I de Castelló. La finalitat d'aquesta iniciativa és aconseguir una millor visibilitat dels materials docents per a contribuir tant a la "innovació educativa" com a "l'accés lliure" al coneixement. Els seus objectius es reproduïxen a continuació; es ressalta l'objectiu (el penúltim) directament relacionat amb la preservació d'aquests materials:

- Facilitar la gestió dels materials i els objectes docents produïts en les universitats, integrant-los i organitzant-los en un servidor comú.
- Oferir al personal acadèmic de les universitats participants un servidor de recursos per a arxivar els materials docents que produeixen i per a reutilitzar-los.
- Proporcionar un accés senzill, ràpid i permanent a la producció docent de les institucions membres.
- Afegir valor als materials aplegats amb elements com ara l'adreça permanent, les citacions normalitzades o les estadístiques de consulta.
- **Establir i aplicar mecanismes de preservació per a assegurar la perdurabilitat dels materials.**
- Potenciar la publicació i l'edició de la producció docent en suport electrònic.

4.2.2. La creació de material d'aprenentatge digital

Tal com s'ha exposat en els mòduls anteriors, és aconsellable prendre decisions sobre la preservació de materials digitals al més aviat possible. Però, actualment, sovint els objectes d'aprenentatge es creen de manera aïllada i sense pautes. Per tant, els docents fan servir "una multitud de formats i metodologies" dels quals resulta un conjunt de materials no normalitzats i que, per això, requereixen una gestió més complexa i, per tant, més costosa. Un informe encarregat per JISC recomana:

"Sempre que sigui possible, les decisions sobre la preservació s'haurien de prendre en el moment de crear l'objecte d'aprenentatge, per facilitar la identificació dels mètodes de preservació per a aquells objectes que s'hauran de retenir durant un període superior a cinc anys" (Conyers i Dalton, 2008, pàg. 46).

Web recomanat

Podeu veure també: Open Educational Resources Common.

De tota manera, els autors d'aquest informe reconeixen que l'estat actual "poc madur" de l'aprenentatge virtual a la Gran Bretanya fa "poc probable" que s'aconsegueixi la intervenció institucional en la creació dels recursos pedagògics.

El mateix any, un altre informe va fer una recomanació més concreta sobre les mesures que les institucions poden prendre:

"Per a facilitar la retenció dels materials d'aprenentatge, les institucions han de requerir l'ús de formats i estructures tècniques normalitzades per a assegurar que tots els materials creats i presentats dins de la institució facin servir una mateixa gamma de formats bàsics de text, àudio, gràfics, vídeo i material en web. Això reduirà la complexitat i el cost del repositori institucional i facilitarà qualsevol migració o emulació posterior" (Emmerson, 2008, pàg. 15).

4.2.3. Característiques i elements

La vida útil dels objectes d'aprenentatge varia molt segons la disciplina (uns amb coneixements en plena evolució i altres amb conceptes estables sense grans canvis) i el grup i nivell a què van dirigits. Els experts parlen d'una vida mitjana d'uns cinc anys, un període que correspon aproximadament a la duració de la carrera universitària. Però la vida tecnològica d'un recurs digital es calcula que dura aproximadament entre tres i cinc anys. Per tant, diversos experts argumenten a favor de la presa de mesures de preservació, tot i que no hi ha el propòsit de preservar el material a llarg termini (Conyers i Dalton, 2008, pàg. 33).

Emmerson recomana les mesures mínimes següents que les institucions haurien de prendre per a garantir l'accessibilitat dels materials a curt termini:

"la combinació de la neutralitat de plataforma i les estratègies de migració tan sols per assegurar que el material sobrevisqui més enllà de l'ús immediat, per a satisfer els requisits de gestió administrativa i institucional."

Segons els autors Ashley, Davis, i Pinsent, els objectes d'aprenentatge poden servir les finalitats següents:

- Objectes d'instrucció (eines d'aprenentatge interactives, presentacions en línia, etc.).
- Objectes de col·laboració (qualsevol aplicació que permeti la interacció humana en línia).
- Objectes d'aplicació (simulació de tasques basades en casos reals).
- Objectes d'avaluació.

Al final de 2011, el directori de repositoris d'accés lliure OpenDOAR incloïa 1.779 repositoris institucionals en la seva base de dades, 274 dels quals arxivaven objectes d'aprenentatge. Dels 83 repositoris institucionals espanyols que hi ha en el directori, 12 indiquen la inclusió d'objectes d'aprenentatge, cap amb arxius exclusivament de materials d'aprenentatge.

4.3. La preservació dels materials d'aprenentatge: la postura dels stakeholders

4.3.1. Arguments a favor de la preservació d'objectes d'aprenentatge

Hi ha diversos motius a favor de la preservació dels materials digitals:

- A curt o mitjà termini
 - La reutilització i compartició de la informació en els objectes d'aprenentatge
 - La capacitat de respondre a tràmits administratius o polítiques institucionals
- A llarg termini
 - L'interès històric en una disciplina o en les activitats d'un docent
 - L'interès pedagògic en l'evolució dels mètodes didàctics i d'aprenentatge

4.3.2. Arguments en contra de la preservació d'objectes d'aprenentatge

Alguns experts, però, no veuen avantatges per a les institucions i argumenten en contra dels esforços per a preservar aquests materials. Els seus motius, que recull el projecte JORUM (Stevenson, 2006, pàg. 35), són els següents:

- "Possiblement seria més barat tornar a crear els materials en lloc d'invertir en la infraestructura necessària per a la gestió d'un procés de preservació.
- Possiblement seria més fàcil i més pràctic crear nous materials –sobretot en el cas d'objectes senzills.
- La creació de nous objectes d'aprenentatge tindria el benefici de la funcionalitat del programari més recent, i els usuaris potser es trobarien més a gust.
- Els nous objectes d'aprenentatge poden estar més ben adaptats a les noves tendències pedagògiques.
- Si no es preserva el material a llarg termini, s'eviten els possibles problemes amb les qüestions complexes relatives als drets d'autor."

Val a dir que la majoria dels experts que afavoreixen els mecanismes de preservació també recomanen una selecció prèvia dels materials susceptibles de ser preservats a mitjà o llarg termini. En el mòdul "Nou paradigma en preservació: la preservació digital" a l'apartat "Quins recursos cal preservar?" s'han comentat la selecció de material per a preservar i també els criteris de selecció recomanats pels autors Jones i Beagrie.

Per a la selecció de material d'aprenentatge, Conyers i Dalton suggereixen criteris que són molt semblants als de Jones i Beagrie:

- valor dels materials

- qüestions tècniques
- metadades
- drets i responsabilitats

4.3.3. Actituds dels autors

Els docents discrepen tant sobre la compartició dels seus materials com pel que fa al seu valor a llarg termini i, per tant, a la necessitat de preservar-los, a diferència de les actituds generalment favorables a la preservació de material digital patrimonial o d'investigació. El projecte Rights and rewards in blended repositories va dur a terme una enquesta per conèixer l'opinió dels professors sobre la compartició dels seus materials docents amb altres col·legues (Bates i altres, 2006). Alguns dels factors que en frenen la col·laboració són:

- El seu interès a controlar el material i el seu ús posterior.
- La possible inclusió de material de titularitat aliena o fins i tot desconeguda.
- La possible inclusió de material confidencial.
- El neguit que els altres puguin jutjar la qualitat dels seus materials.
- Els dubtes sobre la manera de compartir el material i els mecanismes per a permetre la reciprocitat.

Tal com afirmen Conyers i Dalton, no es pot comparar aquesta actitud amb la dels autors de material científic, els quals ja tenen previst compartir els resultats mitjançant la publicació en revistes o la presentació en congressos professionals d'aquests materials. D'altra banda, el fet de divulgar els seus treballs té uns beneficis molt clars: la visibilitat de l'autor i l'impacte dins de la seva disciplina.

Però quan es tracta de material pedagògic, els mateixos docents reconeixen que hi ha pocs incentius per a dipositar i compartir els seus materials i, a més, hi ha poc reconeixement per al seu esforç si ho fan. Per tant, sembla que caldrà resoldre les qüestions culturals, legals i institucionals abans de poder aconseguir la participació generalitzada del professorat en iniciatives per a compartir els recursos docents i dipositar-los en un repositori central.

Hi ha, però, expectatives de canvi en les actituds dels docents, amb l'evidència de l'ús de plataformes externes (*Flickr*, *Slideshare*, etc.) que alguns ja fan servir per a mostrar i donar accés als seus materials. Val a dir, però, que aquestes plataformes que faciliten la distribució de la informació no són adequades per a la seva gestió i, molt menys, per a la seva preservació a llarg termini.

4.3.4. Polítiques institucionals a favor de la preservació

Un document de JISC recomana als administradors acadèmics que elaborin una estratègia per a la gestió i l'ús de materials d'aprenentatge digital i que recullin aquests materials en el repositori institucional en què s'arxiven principalment els treballs relacionats amb la recerca (Rothery, 2008).

Segons aquest text, hi ha una manca de visió estratègica de la postura institucional relacionada amb la gestió de materials docents i d'aprenentatge; troba a faltar, per exemple, polítiques explícites sobre els drets d'autor i la propietat del material. També pot ser necessari introduir millores en la infraestructura que permetin compartir materials pedagògics entre els docents i altres membres institucionals. Els avantatges per a la institució d'una gestió normalitzada dels recursos d'aprenentatge i la compartició posterior entre els seus membres serien:

- Estalviar temps i costos com a resultat de la reutilització de materials.
- Posar a l'abast del professorat recursos d'alta qualitat.
- Donar suport al desenvolupament col·laboratiu dels estudis, tant dintre de cada institució com amb d'altres institucions.
- Donar accés als resultats de projectes de desenvolupament pedagògic.
- Promulgar la transferència dels coneixements resultants de la recerca i innovació institucional per a l'aprenentatge i la docència.
- Agilitzar el procés de reemplaçar professors que deixen la institució.
- Promoure la reputació professional de la universitat i dels seus membres.

4.4. La qüestió dels drets

4.4.1. Els drets intel·lectuals

Tots els estudis sobre aquest tema identifiquen la qüestió dels drets d'autor com un dels obstacles principals per a compartir i preservar els materials d'aprenentatge digital. Com era previsible, és un tema complex. Hi ha diversos aspectes que cal considerar, entre els quals:

Drets institucionals enfront de drets del creador: els estudis britànics han comprovat que la majoria dels docents pensen que són ells mateixos els titulars del material docent que han creat. Aquests estudis també exposen el fet que hi ha poques polítiques, tant en relació amb els materials didàctics tradicionals com amb els digitals. Entre les polítiques identificades hi ha una variació considerable. A més, s'ha vist que quan hi ha un dret institucional sobre el material creat pels docents, normalment no es comunica adequadament al professorat.

Si es pot entendre que els autors reclamin els drets sobre el material que han creat, tampoc no es poden negar els interessos legítims de la institució en casos com els següents:

- si es requereix el material d'aprenentatge per a respondre a una petició administrativa o judicial per part d'un alumne;
- si cal passar els materials a un professor nou que substitueix l'autor del material.

Fins que no s'aprovin i s'implementin polítiques per a aclarir i comunicar la situació dels drets dels materials docents, serà difícil promoure la compartició i la reutilització d'aquests materials, tal com afirma un full informatiu del JISC:

"La confusió que rodeja el tema de drets d'autor se cita com una barrera que frena l'ús de repositoris per part del professorat; per tant, aclarir les polítiques sobre la titularitat és una prioritat alta" (Rothery, 2008).

L'ús d'elements aliens a l'objecte: el mateix full informatiu del JISC també esmenta el cas de professors que reconeixen haver utilitzat material extern – per exemple, contingut o aplicacions informàtiques– sense demanar prèviament l'autorització del titular. Tal com observen Conyers i Dalton, "no s'havia considerat la qüestió de drets quan es va crear el material" (Conyers i Dalton, 2008, pàg. 26). El risc legal era més baix quan el material no sortia de l'entorn original de l'aula, però és molt més alt si es preveu incorporar-lo en un repositori perquè altres el puguin reutilitzar.

4.4.2. Els drets dels estudiants

A més d'utilitzar els recursos pedagògics quan cursen l'assignatura, els estudiants tenen dret a accedir al material d'aprenentatge "durant el temps que quedi pedagògicament útil per a ells" (Emmerson, 2008, pàg. 3). Davant d'una reclamació o un requeriment que impugni els resultats de l'avaluació per part d'un estudiant, la universitat pot necessitar recuperar l'objecte d'aprenentatge tal com era durant la temporada en què l'alumne va estar matriculat a l'assignatura.

Aquest punt, però, introdueix la necessitat del manteniment de les diferents versions de l'objecte pel seu valor probatori, un esforç complex, ja que aquests materials solen experimentar modificacions iteratives i, fins ara, no documentades formalment.

4.5. La gestió dels materials preservats

4.5.1. Propietats significatives

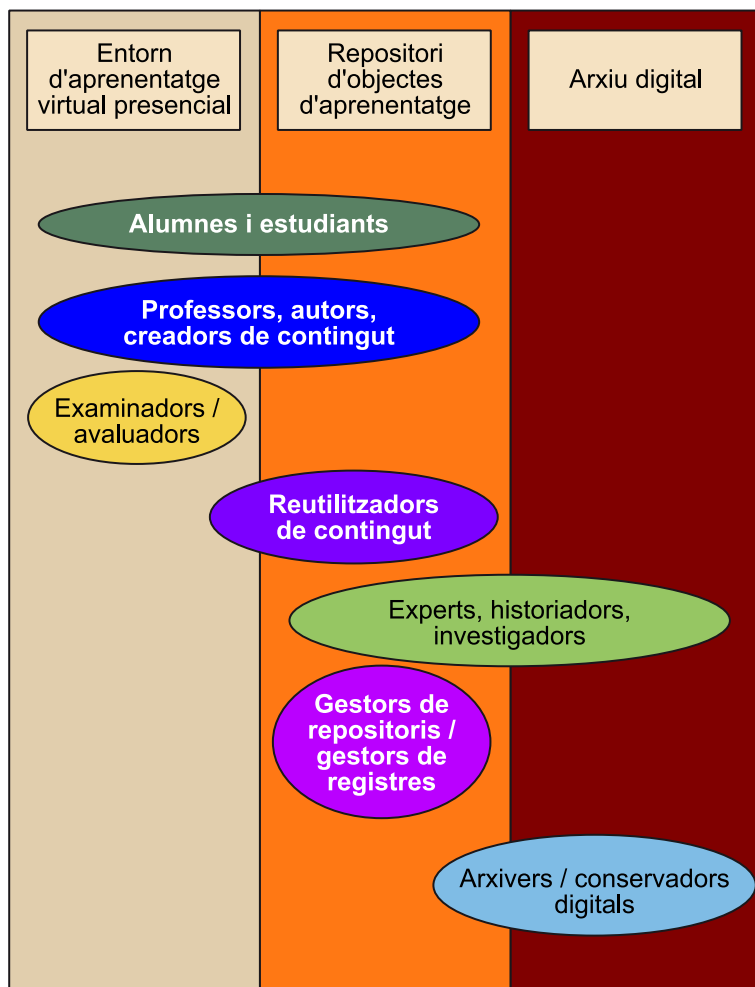
En el mòdul "La preservació digital des de la perspectiva tècnica" s'ha introduït el concepte de propietats significatives ("elements essencials" segons la UNESCO). Hi ha hagut diverses iniciatives per a identificar les propietats dels recursos d'aprenentatge digitals que caldria procurar preservar. Ashley, Davis i Pinsent proposen set grups de propietats significatives:

- Funció
- Context
- Reusabilitat
- Tramesa
- Gestió de drets
- Objecte digital
- *Look and feel* ('aspecte i sensació' segons la UNESCO)

Però com que la determinació de quines propietats s'intentaran preservar depèn de la comunitat designada d'usuaris i de les seves necessitats pedagògiques, aquests autors també identifiquen els possibles usuaris/*stakeholders*. Les comunitats d'usuaris/*stakeholders*, entre la comunitat acadèmica (1-5) i la comunitat de preservació (6-7), que designen són:

- 1) Estudiants.
- 2) Docents, autors i creadors de contingut.
- 3) Avaluadors.
- 4) Altres docents que reutilitzen el material.
- 5) Investigadors, historiadors.
- 6) Gestors d'arxius administratius i de repositoris.
- 7) Arxivers digitals.

El gràfic següent mostra la relació d'aquests usuaris amb les diferents etapes de la vida de l'objecte:



Font: Ashley, Davis, Pinsent, 2007, pàg. 23

4.5.2. Els objectes d'aprenentatge considerats com a material d'arxiu

Sovint, la bibliografia professional sobre aquest tema considera el material d'aprenentatge com a documents administratius i històrics de la institució en què s'han creat. Un dels primers a alertar sobre la pèrdua de materials pedagògics digitals va ser Ross Harvey. L'any 2002 va recomanar que s'apliquessin els procediments arxivístics a la gestió dels materials d'aprenentatge com, per exemple, la identificació, la valoració i la programació de la retenció (Harvey, 2002).

Seguint en aquesta línia, Ashley, Davis i Pinsent recomanen adoptar un cicle vital per als objectes d'aprenentatge i per a la seva gestió. Les etapes previstes són:

- Es creen els objectes.
- S'emmagatzemen en un EVA (VLE en anglès) en què es gestionen mentre tinguin un valor actiu (per exemple, mentre s'utilitzen en una assignatura).
- S'accedeix a aquests objectes, es reutilitzen i es reproduïxen i aquestes accions es capten en l'històric del seu ús.

- S'inclouen en productes administratius com ara avaluacions, resultats d'exàmens, registres d'assistència.
- Es migren a un arxiu digital per a la seva preservació.
- Es tornen a utilitzar dins d'altres objectius d'aprenentatge per a l'objectiu original o per a altres propòsits.

D'altra banda, JISC va redactar uns terminis de retenció de documents en les institucions d'educació superior al Regne Unit, entre els quals s'inclouen el material d'aprenentatge. Les seves recomanacions són que es mantinguin disponibles com a mínim durant un període de cinc a set anys (JISC, 2007).

4.5.3. L'adopció de tecnologies Web 2.0 per als materials d'aprenentatge

Ja s'ha esmentat la problemàtica de la preservació de contingut generat per les tecnologies de Web 2.0; per exemple:

- la qüestió dels drets d'autor en relació amb material generat de manera col·lectiva;
- la utilització de servidors externs i, en molts casos, comercials per a generar el contingut;
- el caràcter dinàmic i canviant del contingut; etc.

Però també l'adopció d'aquestes tecnologies pot representar un element positiu per a la reutilització per tercers dels materials docents. Si actualment la majoria dels estudis observen una actitud poc favorable a la compartició dels materials docents, l'evolució cap a material generat i difós mitjançant les xarxes socials pot marcar un canvi d'actitud, ja que, en el moment d'adoptar aquestes xarxes per a transmetre informació als alumnes, els docents ja comencen a compartir el seu material més enllà de la seva aula. Es preveu que la reticència a compartir material docent desapareixerà amb el proper relleu generacional.

Bibliografia

Abbott, Daisy (2008). "What is digital curation".

Alliance for Permanent Access.

Angevaere, Inge (2009). "Taking care of digital collections and data: «Curation» and organisational choices for research libraries". *Liber Quarterly* (vol. 19, núm. 1, pàg. 6-7).

Ashley, Kevin; Davis, Richard; Pinsent, Ed (2008). "Significant properties of e-learning objects". Draft version 0.91. JISC.

Ball, Alex (2010a). *Preservation and curation in institutional repositories*. Edimburg: Digital Curation Centre.

Ball, Alex (2010b). *Review of the State of the Art of the Digital Curation of Research Data*. Project Report. Bath: University of Bath.

Barker, Ed i altres (2004). "Long-term retention and reuse of e-learning objects and materials; report commissioned by the Joint Information Systems Committee". Ver.1.4.

Bates, Melanie i altres (2006). *Rights and rewards project: academic survey; final report*.

Beagrie, Neil; Chruszca, Julia; Lavoie, Brian (2008). "Keeping research data safe-1: a cost model and guidance for UK universities".

Beagrie, Neil; Lavoie, Brian; Wollard, Matthew (2010). *Keeping research data safe-2: final report*.

Beam, Christopher (2010). "#Posterity: how future historians will use the Twitter archives". *Slate*.

Campbell, Laura (2010). "Preservación digital y derechos legales". A: *5º Congreso Nacional de Bibliotecas Públicas*. Gijón.

Campbell, Laura; Dulabahn, Beth (2010). "Digital preservation: the Twitter archives and NDIIPP"

Campbell, Lorna M. (2005) "Repository issues from a teaching and learning perspective".

Charlesworth, Andrew i altres (2007) *Sharing eLearning content: a synthesis and commentary; final report*.

Conyers, Angela; Dalton, Pete (2008). *Preservation of e-learning materials: an attitudinal study*. Birmingham City University.

Corbyn, Zoë (2011, setembre), "Memory failure detected". *Times Higher Education* (núm. 1).

Data centres: their use, value and impact (2011).

Data Conservancy: A Blueprint for Research Libraries.

Data Conservancy NSF Data Management Plan Template.

Data management plans.

Data Repositories blog. Simmons College.

Data Seal of Approval.

Data Seal of Approval: an overview.

Data Seal of Approval Quality Guidelines.

Davis, Maxine (2009). "Preserving access: Making more informed guesses about what works". Report to the IIPC Preservation Working Group.

Davis, Richard (2010). "Moving targets: web preservation and reference management". *Ariadne* (núm. 62).

Digital Curation Centre.

Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library. (2008)

DMP Online: DCC Data Management Planning Tool.

Emmerson, Peter (2008) "Retention of learning materials: a survey of institutional policies and practice"

Ensuring the integrity, accessibility and stewardship of research data in the digital age (2009). National Academy of Sciences.

Ferran Ferrer, Núria; Minguillón Alfonso, Julià (editors) (2011). *Content management for e-learning*. Nova York / Londres: Springer.

Franklin, Tom; Harmelen, Mark van (2007). *Web 2.0 for content for learning and teaching in Higher Education*.

Grotke, Abbie (2008). *International Internet Preservation Consortium 2008. Member Profile Survey Results*.

Guidelines for Effective Data Management Plans.

Hank, Carolyn; Sheble, Laura; Choemprayong, Songphan (2009) "Considerations for the preservation of blogs: briefing paper of Digital Preservation Europe."

Harmelen, Mark van (2007). "Briefing paper on Web 2.0 technologies for content sharing: Web 2.0; An introduction". Version 2.0.

Harvey, Ross (2002) "Now you see it, now you don't: maintaining digital learning objects for the future".

Insight into issues of permanent access to the records of science in Europe (2010).

Jacobsen, Grethe (2007). "Archivo web internacional: ¿interoperabilidad en el futuro?". IFLA World Library and Information Congress, 73rd. Durban. (Traducció a l'espanyol: Ma. José Llovera Lozán)

JISC InfoNet (2007). *HEI records retention schedule: user guide*.

Joy, Mike (2011). "Blogforever".

Legal status of research data in the Knowledge Exchange partner countries.

Letierce, J. i altres (2010). "Understanding how Twitter is used to spread scientific messages". *Web Science Conference*. Raleigh, NC.

Llei 23/2011, de 29 de juliol, de dipòsit legal. BOE (núm. 182, 20 de juliol de 2011, pàg. 86716-86727). <http://www.boe.es/boe/dias/2011/07/30/pdfs/BOE-A-2011-13114.pdf>

Llueca Fonollosa, Ciro (2005). "Webs sempre accessibles: les biblioteques nacionals i els dipòsits digitals nacionals". *BiD: textos universitaris de biblioteconomia i documentació* (núm. 15).

Llueca Fonollosa, Ciro i altres (2011). "PADICAT: el archivo de internet". 12as Jornadas Españolas de Documentación (Fesabid). Màlaga.

Lotan, Gilad i altres (2011). "The revolutions were tweeted: information flows during the 2011 Tunisian and Egyptian revolutions". *International journal of communication* (núm. 5, pàg. 1375-1405).

Lord, Philip; Macdonald, Alison (2003). "E-Science curation report: data curation for e-science in the UK: An audit to establish requirements for future curation and provision". *JISC*.

Maness, J. (2006). "Library 2.0 theory: web 2.0 and its implications for libraries". *Webology* (vol. 3, núm. 2, art. 25).

Maron, Nancy L.; Loy, Matthew (2011). *Funding for sustainability: How funders' practices influence the future of digital resources*.

Meyer, Eric T.; Thomas, Arthur; Schroeder, Ralph (2011). *Web archives: the future(s)*. Oxford: Oxford Internet Institute.

Milligan, Colin (2007). *Community dimensions of learning object repositories: CD-LOR final report*. Ver. 1.0.

Nicholson, Sean W.; Bennett, Terrence B. (2011) "Data sharing: Academic libraries and the scholarly enterprise" *Portal: Libraries and the Academy* (vol. 11, núm. 1, pàg. 505-516).

OCLC (2007). "Sharing, privacy and trust in our networked world: a report to the OCLC membership".

PARSE.Insight Project (2008-2010).

Pennock, Maureen (2007). "Digital curation. A life-cycle approach to managing and preserving usable digital information". *Library & Archives Journal* (núm. 1).

Pérez Morillo, Mar; Muguerza López, Icíar (2011). "El patrimonio del futuro y el futuro del patrimonio: el archivo web de la BNE". (Presentació PPT.)

POWR: the preservation of web resources handbook (2008).

Research Data: Unseen Opportunities. (2010)

Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. October 2010. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

Rothery, Andrew (2008) "Managing and sharing e-learning resources: how repositories can help". *JISC briefing paper*.

Stevenson, Jane (2005). *Preservation watch report*.

Stevenson, Jane (2006). *Jorum preservation watch report*.

Stewardship of digital research data - a framework of principles and guidelines.

Sustaining digital resources: an on-the-ground view of projects today (2009).

Swan, Alma; Brown, Sheridan (2008). "To share or not to share: Publication and quality assurance of research data outputs". *Research Information Network*.

To stand the test of time: Long-term stewardship of data sets in science and engineering (2006).

UK Data Archive (2011). *Managing and sharing data: best practice for researchers*.

What is digital curation?

