

Representació i visualització d'informació a Internet

Pablo Lara Navarra
David Maniega Legarda

PID_00193315



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

| | |
|---|----|
| Introducció..... | 5 |
| Objectius..... | 6 |
| 1. L'organització i representació del coneixement electrònic.... | 7 |
| 1.1. Les dades, la informació i el coneixement | 7 |
| 1.2. De la informació al coneixement | 8 |
| 1.3. La gestió del coneixement | 10 |
| 1.4. La societat del coneixement | 11 |
| 1.5. Les dades en la cadena de valor del coneixement | 12 |
| 1.5.1. Dades en obert, una filosofia i un fet | 14 |
| 1.5.2. <i>Big data</i> / gran dada | 16 |
| 2. Els llenguatges de marcatge en la representació i l'organització del coneixement..... | 18 |
| 2.1. La família de normes SGML/XML | 18 |
| 2.2. Les DTD de SGML | 19 |
| 2.3. El llenguatge XML | 20 |
| 2.4. Mecanismes per a representar el coneixement | 21 |
| 2.5. Serveis d'informació web basats en XML | 22 |
| 2.6. L'XML en l'organització del coneixement intern | 23 |
| 3. Arquitectura de la informació, navegació i recuperació..... | 25 |
| 3.1. L'arquitectura de la informació | 25 |
| 3.2. Sistemes de classificació del coneixement electrònic | 26 |
| 3.3. Navegació i recuperació de la informació web | 28 |
| 4. Tecnologies, programes i formats..... | 30 |
| 4.1. Els servidors | 30 |
| 4.2. <i>Cloud computing</i> : Internet/intranet | 31 |
| 4.2.1. Visió retrospectiva d'intranet documental | 32 |
| 4.2.2. SharePoint, una solució híbrida | 34 |
| 4.2.3. Tecnologia en el núvol | 38 |
| 4.2.4. Avantatges i inconvenients principals del <i>cloud computing</i> | 40 |
| 4.2.5. Entenent el <i>cloud computing</i> | 41 |
| 4.2.6. Intranets en el núvol | 43 |
| 4.3. RIA (<i>rich internet applications</i>) | 45 |
| 4.4. Els programes de programari | 47 |
| 4.5. El programari lliure | 48 |
| 4.6. Els formats de documents | 49 |

| | | |
|---------------------|---|-----------|
| 4.7. | Els documents multimèdia | 51 |
| 4.8. | El format PDF | 54 |
| 4.9. | Els formats ODF i OpenXML | 55 |
| 5. | Visualització d'informació: Casos | 56 |
| 5.1. | Visualització d'informació | 57 |
| 5.2. | Exemples de models de visualització | 65 |
| 5.3. | Recursos web complementaris | 69 |
| 6. | Cas model SocialNet: gestió del coneixement amb dades en obert | 70 |
| 6.1. | L'aplicació SocialNet per dins | 71 |
| Bibliografia | | 77 |

Introducció

En aquest material de l'assignatura de *Representació i visualització d'informació* s'aclareixen els conceptes de *dada*, *informació* i *coneixement*, s'especifiquen els tipus de coneixement que hi ha, s'introdueixen les característiques bàsiques de la gestió del coneixement i s'aclareix la influència de la societat del coneixement en la producció de continguts electrònics i la necessitat de representar-ne el coneixement.

A continuació s'estudia com la gestió del coneixement i els models d'accés en obert han permès una revaloració de les dades com a elements fonamentals per a entendre els canvis que s'estan produint en aquesta àrea de coneixement. Conceptes com *dades en obert* (open data) o *gran dada* (*big data*) apareixen cada vegada més sovint. Es destaca l'abast i la importància que aquests conceptes tenen en les organitzacions i les societats intensives en informació. En la gestió de dades obertes, que avui dia es poden consultar massivament, cal generar noves eines que permetin la representació i l'organització del coneixement en obert. Finalment, es presenta una solució de sistema que permet representar i gestionar dades existents en repositoris en obert a partir de recollidors de dades per a generar una nova informació prenent com a base elements quantitius i qualitius.

A continuació s'analitzen les possibilitats que ofereixen els llenguatges de marcatge per a la classificació i l'organització del coneixement, es descriuen els tipus de llenguatges de marcatge per a estructurar la informació, es presenta la família de normes SGML/XML i les relacions amb altres llenguatges de marcatge, s'analitzen les possibilitats dels llenguatges de marcatge en la representació del coneixement, i més tard es presenten alguns exemples d'utilització d'XML per al desenvolupament de serveis d'informació i la gestió del coneixement organitzacional.

En l'apartat següent es tracten les principals eines i vocabularis estructurats per a l'organització del coneixement electrònic: llistes de categories, taxonomies, tesaurus, classificacions facetades i ontologies. Es destaca la funció dels vocabularis estructurats per a l'organització de continguts, el foment de la interoperabilitat de la informació i el desenvolupament de sistemes de navegació i recuperació de la informació en els serveis electrònics.

Finalment, s'introdueixen la tecnologia, els programes i els formats que formen els sistemes d'informació entorn d'Internet, ja que evolucionen amb gran rapidesa, fins i tot apareixen i desapareixen sense percebre els canvis que han implicat en la construcció de nous paradigmes a la Xarxa de xarxes.

Objectius

- 1.** Com es representa, organitza i visualitza la informació.
- 2.** Quina és la diferència entre dades, informació i coneixement.
- 3.** Quines són les característiques principals de la gestió del coneixement electrònic i de la societat del coneixement.
- 4.** Quin és el nou paper de les dades en la societat de la informació.
- 5.** Quina evolució de paradigma estan succeint en les dades.
- 6.** Com s'han d'utilitzar els llenguatges de marcatge per a representar el coneixement en els serveis web.
- 7.** Què és l'arquitectura de la informació web i quines eines i vocabularis controlats es poden utilitzar per a millorar els sistemes de navegació i recuperació de la informació electrònica.
- 8.** Com evoluciona el camp de la visualització per mitjà de casos.

1. L'organització i representació del coneixement electrònic

1.1. Les dades, la informació i el coneixement

Les dades són els resultats d'observacions humanes sobre l'estat del món i estan associades a un objecte o fet concret. És a dir, les dades són expressions mínimes d'informació que aïllades no tenen sentit en elles mateixes, però que adquireixen valor dins d'un context determinat. Les dades posseeixen tres característiques essencials.

- En primer lloc, es poden identificar amb claredat, sense possibilitat de confusió, per conjunts de símbols.
- En segon lloc, tenen un nivell d'estructura elevat, és a dir, la possibilitat d'error en la comunicació entre emissor i receptor és mínima.
- Finalment, les tecnologies de la informació són molt útils quan s'apliquen al processament de les dades.

La informació es compon de dades organitzades, agrupades o classificades en categories que les doten de significat. A més, la informació està associada a un context que en facilita la interpretació. Per tant, *informació* són dades elaborades amb un significat per al receptor. D'aquesta manera, la informació es pot entendre com la significació que adquireixen les dades com a resultat d'un procés conscient i intencional d'adequació de tres elements: les dades de l'entorn, els propòsits i el context d'aplicació, i també l'estructura de coneixement del subjecte.

El coneixement és la informació interpretada, personalitzada, que té valor i que està orientada a l'acció, això és, que propicia la presa adequada de decisions. El coneixement està associat a l'individu i a l'acció, el qual constitueix un instrument per a la presa de decisions en el marc d'una organització.

Es poden identificar dos tipus de coneixement en una organització, i que és precisament la conversió constant d'un tipus a l'altre el que explica la generació de nou coneixement.

El **coneixement tàcit** és aquell que difícilment pot ser formalitzat, expressat en paraules (la destresa d'un músic o la inspiració d'un artista, entre d'altres), i també les habilitats personals més simples que són molt difícils de transmetre d'una persona a una altra (utilitzar l'ordinador, muntar amb bicicleta o conduir un cotxe, entre d'altres). Per tant, el coneixement tàcit resideix indi-

vidualment en cada persona. Les organitzacions disposen de coneixement tàcit “encapsulat” en forma de persones i difícilment es pot emmagatzemar en altres formes.

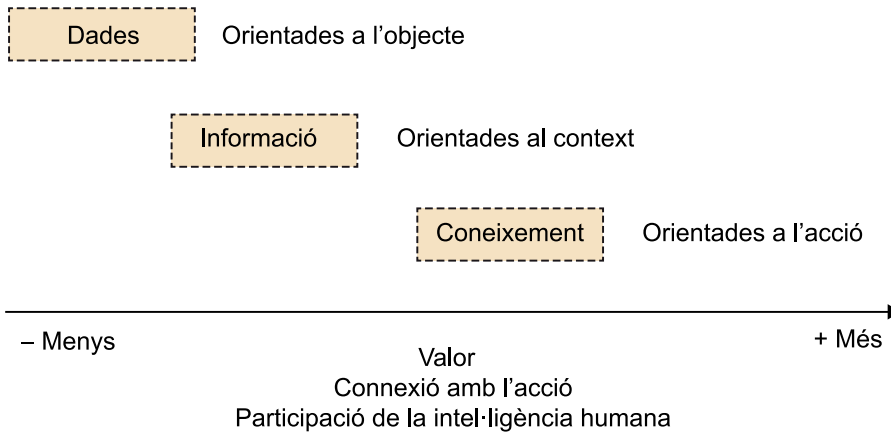
El **coneixement explícit** és aquell que es pot expressar en forma de paraules i nombres, i que és compartit entre persones per mitjà de dades, fórmules, especificacions, etc. (un programa d'ordinador, una patent, un catàleg, un manual, etc.), per tant, el coneixement explícit és més fàcilment transmissible entre persones. Les organitzacions disposen de grans quantitats de coneixement explícit, en forma de processos estandarditzats, procediments, manuals, guies i altres documents.

1.2. De la informació al coneixement

Les diferències essencials entre *dades* i *informació* es poden sintetitzar en dues. En primer lloc, la informació és només allò que l'usuari considera que li informa; les dades, encara que siguin rellevants, no li aporten res nou o res que es pugui predir a partir del que ja sabia, l'usuari no les considera informació. En segon lloc, les tecnologies de la informació poden manejar molt bé les dades, però convertir les dades en informació és un procés bàsicament humà, encara no es pot automatitzar.

La diferència essencial entre *informació* i *coneixement* consisteix en el fet que el coneixement es refereix a l'habilitat d'entendre el context i descobrir les connexions i el significat de la informació. És a dir, *informació* i *coneixement* són en principi conceptes diferents, però no pot existir l'un sense l'altre. La informació no té valor per si mateixa, el valor resideix en l'habilitat d'un individu per a aportar-li significat i crear coneixement nou. El coneixement, igualment, manca de valor per si mateix, el valor apareix quan el coneixement s'utilitza per a solucionar un problema o generar informació nova. El coneixement seria l'experiència i la capacitat dels individus, unides als processos i a la memòria corporativa; i la informació seria la matèria primera que el coneixement utilitza per a generar coneixement nou.

En la figura següent es representen les interrelacions que hi ha entre els conceptes de *dada*, *informació* i *coneixement*.

Figura 1. Relacions entre *dades*, *informació* i *coneixement*

En un sentit genèric i seguint la tendència en l'evolució qualitativa del trinomi *dades*, *informació* i *coneixement*, es pot construir un model piramidal d'aquesta evolució. Les dades, que són entitats independents i sense significat intrínsec, es converteixen en informació quan són:

- **contextualitzades** (es coneix el propòsit de les dades recollides),
- **categoritzades** (es coneixen les unitats d'anàlisi o els components clau de les dades),
- **calculades** (les dades són analitzades matemàticament o estadísticament),
- **corregides** (s'eliminen els errors que hi ha en les dades) o bé
- **condensades** (les dades es resumeixen en una forma més concisa).

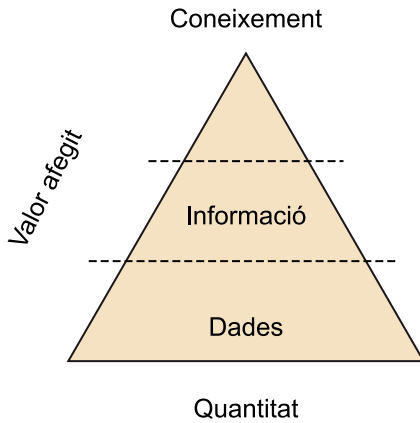
Per tant, usant un o més dels cinc criteris anteriors, les dades es converteixen en informació.

Igualment, la informació es pot transformar en coneixement quan es produeixen una o diverses de les situacions següents:

- **comparació** (la informació es compara amb altres informacions ja conegudes),
- **conseqüència** (la informació suposa certes implicacions per a la presa de decisions o l'acció),
- **connexió** (la informació està o no relacionada amb altres coneixements), i
- **conversa** (l'opinió d'altres persones sobre la informació).

Es tracta, per tant, d'una progressió lineal en què les dades es converteixen en informació i aquesta informació, en coneixement mitjançant un procés d'agregació de valor, que es pot representar mitjançant un model piramidal.

Figura 2. Model piramidal de la informació



El coneixement és la informació internalitzada, és a dir, integrada en les estructures cognitives de la persona; per tant, si no hi ha persona, no hi ha coneixement. Però en els últims anys, amb la influència de les tecnologies de la informació, ha sorgit un nou punt de vista que considera que el coneixement pot ser generat mitjançant la informació emmagatzemada en la memòria humana o en la memòria artificial (electrònica).

Quan el coneixement es produeix a partir del processament de la informació continguda en la memòria humana, s'obté el denominat *coneixement simbòlic*.

Quan el coneixement es produeix a partir del processament de la informació codificada i organitzada documentalment (tècniques d'emmagatzematge, processament i recuperació de la informació), s'obté el coneixement figuratiu.

Un pas més enllà, partint del coneixement figuratiu, quan generem informació nova per mitjà de tècniques d'intel·ligència artificial, s'obté nou coneixement, denominat *coneixement productiu*, que es podria descriure com un tipus de coneixement artificial o informació electrònica útil.

1.3. La gestió del coneixement

La gestió del coneixement s'ha considerat des d'una simple moda fins que és el principi de la solució a tots els problemes d'incompetència organitzacional. L'anàlisi del concepte de *gestió del coneixement* s'inicia amb l'estudi dels tipus de coneixement que es poden produir en una organització. Una classificació tradicional diferencia el coneixement per la facilitat de comunicació i explotació que té. D'aquesta manera, en una organització hi ha dos tipus de coneixement:

- el **coneixement explícit**, entès com les habilitats, les experiències i els fets que estan escrits o es poden escriure, i es comuniquen amb facilitat; i
- el **coneixement tàcit**, entès com les habilitats, les experiències i les intuïcions de les persones, que s'emmagatzemen en la seva ment i que no es poden transmetre amb facilitat.

Una altra manera de classificar el coneixement es fonamenta en la valoració econòmica. En aquest cas, diferenciem entre tres tipus de capital:

- el **capital humà** es refereix al coneixement útil per a l'empresa que posseeixen les persones i equips d'aquesta empresa, i també la seva capacitat d'aprenentatge,
- el **capital estructural** està format per coneixements estructurats com sistemes d'informació i comunicació, tecnologia, procediments de treball i patents, entre d'altres,
- el **capital relacional** fa referència al valor que per a l'organització tenen el conjunt de relacions establertes amb l'exterior, com aliances, contractes i col·laboracions, entre d'altres.

Des d'una perspectiva basada en la facilitat de comunicació i explotació, la gestió del coneixement pretén identificar, organitzar i explotar racionalment el coneixement explícit (el registrat o susceptible de ser-ho per l'organització) i transformar la major quantitat possible de coneixement tàcit (el que tenen els individus de l'organització) en explícit. Per contra, des d'una perspectiva basada en la valoració econòmica dels actius, la gestió del coneixement és l'art de crear valor a partir dels actius intangibles d'una organització.

1.4. La societat del coneixement

Quatre factors impulsen l'aparició de la societat i economia del coneixement:

- 1) El fenomen de la globalització, que interrelaciona les economies de zones geogràficament disperses mitjançant la internacionalització de les empreses, el flux de capitals, béns, serveis i persones, i l'obertura de nous mercats.
- 2) El fenomen de les TIC, i especialment d'Internet, que ha comportat un increment significatiu en les possibilitats de comunicació i transmissió d'informació i coneixement.
- 3) La tendència de les empreses i les institucions a organitzar-se de manera més distribuïda, fomentant l'aparició de xarxes geogràficament disperses i descentralitzades.

4) La intensitat creixent en l'aplicació del coneixement en la producció de béns i serveis.

A partir de la segona meitat de la dècada dels anys noranta del segle xx, les tecnologies digitals han permès, fomentat i ampliat notablement la importància econòmica del coneixement, mitjançant dues vies principals. La primera via ha estat l'espectacular millora de l'accés i la gestió dels fluxos d'informació i coneixement, que ha minimitzat les barreres per a la difusió d'aquests dos recursos, i, per tant, s'ha produït un increment notable del coneixement explícit. La segona via, estretament vinculada a la primera, ha estat la millora de les possibilitats d'accés i difusió dels elements que incideixen en el coneixement tàcit, bàsicament, les habilitats formatives i l'experiència.

En aquesta nova societat/economia, el coneixement passa a ser un recurs tan significatiu o fins i tot més que el capital i la mà d'obra. Això és, l'augment de coneixement explícit, la transformació de coneixement tàcit en explícit i el desenvolupament d'habilitats noves han generat un cercle virtuós en la producció de coneixement, que es constitueix com un dels recursos estratègics de l'activitat econòmica de l'actualitat. En el nou context social i econòmic entorn del coneixement, s'atorga una gran importància a la generació, la difusió i l'ús de la informació i el coneixement en les organitzacions. El bon ús del coneixement determinarà el nivell d'èxit tant de les organitzacions com de les economies nacionals.

Les organitzacions s'enfronten al repte de projectar-se i adaptar-se a un procés de canvi que avança molt ràpidament. Es tracta d'un procés dinàmic, caracteritzat essencialment pel desenvolupament de noves tendències en la generació, la difusió i la utilització del coneixement, les quals demanen la revisió i l'adequació de moltes organitzacions i procuren, així mateix, la creació d'altres de noves amb capacitat per a assumir i orientar el canvi.

En aquest context, la **societat del coneixement** és una tendència social i econòmica amb capacitat per a generar, apropiat i utilitzar el coneixement per a atendre les necessitats del seu desenvolupament i així construir el seu propi futur, i converteix la creació i la transferència del coneixement en una eina de la societat per al seu propi benefici. En la societat del coneixement, les comunitats, les empreses i les organitzacions avancen gràcies a la difusió, l'assimilació, l'aplicació i la sistematització de coneixements creats o obtinguts localment, o aconseguits de l'exterior. El procés d'aprenentatge es potencia en comú, per mitjà de xarxes, empreses, comunicació interinstitucional i intrainstitucional, entre comunitats i països.

1.5. Les dades en la cadena de valor del coneixement

Des de final del segle xx, la gestió del coneixement és un tema central d'estudi i excusa perfecta per a executar grans inversions en el desenvolupament de mètodes, models i eines que aconseguixin explicitar el coneixement tàcit

d'individus i organitzacions. L'aplicació d'aquest tipus de tècniques permet estalviar esforços i temps en tasques rutinàries, i també augmentar la productivitat (Castells, 1996; Davenport i Prusak, 1998; Choo, 1998; Chong i altres, 2000; Carrión, 2001; Saz, 2003; Navarro, 2003).

Per la seva banda, en l'àmbit empresarial també podem trobar nombrosos estudis fets per consultores en què estableixen que l'ús d'eines de gestió del coneixement constituïrien una revolució per a la millora de l'eficàcia en les organitzacions (*Knowledge Management Research Report*, 2000).

La gestió del coneixement, per la seva forta irrupció en els àmbits acadèmics i empresarials, ha constituït la porta d'entrada a altres camps com la gestió de continguts, comunitats de pràctiques, semàntica web o coneixement intel·ligent, entre d'altres, vitals per a entendre què succeeix actualment amb les dades. Mentrestant, el moviment obert emergeix amb força, com una nova manera d'entendre la societat de la informació basada en programari, dades o serveis, i en obert (que desenvolupem més detalladament en l'apartat 4.5, "El programari lliure", d'aquest mòdul). Cal indicar que, segons la disciplina, és difícil determinar si són conceptes amb identitat pròpia o s'han utilitzat com a sinònims indiscriminadament.

El moviment obert inicia els passos quan Richard Stallman, fundador de la Free Software Foundation, va començar a treballar en el projecte GNU en els anys vuitanta. A Stallman se li atorga la creació del programari lliure i del concepte *copyleft* (Wikipedia, 2011), que va desenvolupar per a proporcionar més llibertat als usuaris i per a restringir les possibilitats d'apropiació del programari. Per tant, l'auge per a explotar el coneixement en totes les dimensions, al costat del moviment basat en l'accés obert, desencadena la revaloració de la primera baula de l'esmentada cadena de valor de la informació, la dada. Avui dia podem afirmar amb rotunditat que la dada és el rei en el paradigma del coneixement, com a continuació s'abordarà.

Què succeeix amb les dades?

Com s'ha exposat anteriorment, els models oberts han modificat certes regles aparentment tancades i fins i tot, però equivocadament, impossibles d'alterar, com per exemple, gestionar dades de diferents fonts i de manera massiva. Models que ratifiquen aquest fet els trobem en les dades geolocalitzades del cadastre, conèixer qui i quantes ajudes atorguen els governs, l'ús i la disposició de caixers automàtics, el tractament d'imatges satel·litals, les freqüències horàries de trens, el seguiment de cotitzacions d'empreses en diferents borses mundials, els termes més llegits en xarxes socials, etc.

Les administracions, coneixedores d'aquest canvi, ja han començat a legislar (Directiva 2003/98/CE i, específicament, a Espanya amb la Llei 37/2007, de 16 de novembre) dissenyant un marc regulador que:

“persegueix harmonitzar l'explotació de la informació en el sector públic, especialment la informació en format digital [...] amb la finalitat de facilitar la creació de productes i serveis d'informació basats en documents del sector públic, i reforçar l'eficàcia de l'ús transfronterer d'aquests documents per part dels ciutadans i de les empreses privades perquè ofereixin productes i serveis d'informació de valor, amb la qual cosa s'afavoreix la transparència i la millora de l'accés a les dades.”

Així doncs, tenim dades de diferents tipus i fonts diverses que es troben a la disposició de qualsevol ciutadà del món i, curiosament, moltes en temps real, la qual cosa en l'argot informàtic significa que es poden treballar en calent, subministrant informació d'alt valor espaciotemporal. En conseqüència, totes aquestes dades necessiten nous paradigmes basats en l'*open data* o *big data* i, avui dia, ningú no pot arribar a conèixer-ne la dimensió i l'abast tant en l'àmbit industrial en concret com per a la ciutadania en general, com bé expressava Tim Berners-Lee (*TEDtalks*, 2009), que afirma el següent:

“si la gent col·loca les seves dades en el web, al costat de dades governamentals, científiques, i les pròpies de la comunitat, les podran usar altres persones en formes que no es podrien haver imaginat mai.”

1.5.1. Dades en obert, una filosofia i un fet

La capacitat de capturar, tractar, analitzar, modificar, reutilitzar o difondre dades, en definitiva, gestionar-les sense continent perquè tinguin un nou valor informatiu, és el marc conceptual que cobreix el concepte de *dades en obert*, i del qual es nodreixen diferents definicions depenent de l'àmbit d'actuació com es pot deduir en les definicions següents.

Alguns autors defineixen el terme a partir de la comparació d'universos semblants, com Rodrigo (2012) quan indica que:

“*Open data* s'assembla al moviment *open source*, però a diferència d'aquest moviment no apunta a desenvolupar programari sinó a possibilitar l'accés i la reutilització de dades.”

Per la seva banda, Peset, Ferrer-Sapena i Subirats-Coll (2011) estableixen el concepte de *dades obertes* des d'un punt de vista basat en estàndards tècnics com un moviment que promou l'alliberament de dades, generalment no textuals i en formats reutilitzables com CSV (*comma separated values*), procedents d'organitzacions. Al seu torn, també es refereix a la seva exposició o altres dades en el web, descrites en RDF (*resource description framework*) i amb indicació de la relació que hi pot haver entre aquestes dades i d'altres.

Un tercer punt de vista és l'ofert pel Govern balear, amb una visió de transparència administrativa envers el ciutadà, que entén l'*open data* com una nova manera de publicar les dades que requereix que siguin de lliure accés per a tots els ciutadans, sense limitacions tècniques o legals i publicades en formats oberts i estructurats. Pel que fa a l'Administració pública, l'accés a les dades garanteix un marc de transparència i eficiència com demostra el portal datos.gob.es, de caràcter nacional, que organitza i gestiona el Catàleg d'Informació Pública de l'Administració General de l'Estat.

Hem d'indicar l'existència de nombrosos grups de treball dedicats a les dades en obert, d'entre els quals cal esmentar l'aportació del Grup Open Data-RISP (2012), que recull deu punts bàsics, més un d'essencial, que ha de tenir qual-sevol iniciativa d'obertura de dades:

- 1) Harmonització entre administracions.
- 2) Publicar dades en formats oberts i estàndards.
- 3) Usar esquemes i vocabularis consensuats.
- 4) Inventari en un catàleg de dades estructurat.
- 5) Dades accessibles des d'adreces web persistents i amigables.
- 6) Exposar un conjunt mínim de dades relatives a les competències de l'organisme i la seva estratègia d'exposició de dades.
- 7) Compromís de servei, actualització i qualitat de la dada, mantenint un canal eficient de comunicació reutilitzador AAPP.
- 8) Monitoritzar i avaluar l'ús i el servei mitjançant mètriques.
- 9) Dades amb condicions d'ús no restrictives i comunes.
- 10) Evangelitzar i educar en l'ús de dades.
- 11) Recopilar aplicacions, eines i manuals per a motivar i facilitar la reutilització.

Amb tot el que hem afirmat, i des d'una visió de l'àrea de la documentació i informació, cal destacar que les dades en obert, independentment de la font i la seva naturalesa, una vegada alliberades i posades a la disposició del consumidor/usuari, necessiten una narració per a ser convertides en informació i que, al seu torn, derivi en coneixement col·lectiu. Així, aquesta capacitat d'establir un argument perquè les dades llancin la informació necessària, es torna vital per a la presa de decisions. D'aquesta manera, en aquesta narrativa, el narrador seria el professional/sector que necessita interpretar les dades en obert, i el gestor d'informació seria l'encarregat de la narració en sentit estricte, tal com indica la RAE, que defineix el concepte de *narració* com l'encarregada d'esclarir els fets per a aclarir l'assumpte que es tracta i per a facilitar l'assoliment de les finalitats de l'orador.

1.5.2. *Big data* / gran dada

La primera pregunta que sorgeix davant la lectura d'aquest terme és quin és l'abast del mateix concepte?, seguida de, què entenem per *moltes dades*? Per a poder donar una resposta ràpida, en la taula 1 es fa un resum d'unitats de mesurament binàries i la correspondència en un univers d'informació concret, i es destaca com la "mesura" de les dades té el seu rang d'importància, com ho demostren els sorprenents 15 TB que formen el catàleg de llibres de l'America's Library of Congress.

Taula 1. Inflació de dades

| Unitat | Mida | Significat |
|----------------|--------------------------|--|
| Bit (b) | 1 o 0 | Abreviatura de <i>binary digit</i> (dígit binari), del codi binari (1 o 0) que utilitzen els ordinadors per a emmagatzemar i processar dades. |
| Byte (B) | 8 bits | Prou informació per a crear una lletra o un número anglesos en codi informàtic. És la unitat de computació bàsica. |
| Kilobyte (KB) | 1.000 o 2^{10} bytes | De <i>mil</i> en grec. Una pàgina de text ocupa 2 KB. |
| Megabyte (MB) | 1.000 KB; 2^{20} bytes | De <i>gran</i> en grec. L'obra completa de Shakespeare ocupa un total de 5 MB. Una cançó pop normal ocupa uns 4 MB. |
| Gigabyte (GB) | 1.000 MB; 2^{30} bytes | De <i>gegant</i> en grec. Una pel·lícula de dues hores es pot comprimir en 1 o 2 GB. |
| Terabyte (TB) | 1.000 GB; 2^{40} bytes | De <i>monstre</i> en grec. Tots els llibres catalogats a la Biblioteca del Congrés dels Estats Units ocupen un total de 15 TB. |
| Petabyte (PB) | 1.000 TB; 2^{50} bytes | Totes les cartes lliurades pel servei postal dels Estats Units durant aquest any ocuparan uns 5 PB. Google processa aproximadament 1 PB cada hora. |
| Exabyte (EB) | 1.000 PB; 2^{60} bytes | Equivalent a deu mil milions d'exemplars de <i>The Economist</i> . |
| Zettabyte (ZB) | 1.000 EB; 2^{70} bytes | Es preveu que la quantitat total d'informació existent aquest any ocupi uns 1,2 ZB. |
| Yottabyte (YB) | 1.000 YB; 2^{80} bytes | Actualment és massa gran per a poder-se'l imaginar. |

Font: *The Economist*.

Els prefixos són fixats per un grup intergovernamental, l'Oficina Internacional de Pesos i Mesures. Yotta i zetta es van afegir l'any 1991; encara no s'han fixat els termes per a quantitats superiors.

D'aquesta manera, podem determinar que les eines basades en el model *big data* comencen a ser necessàries en magnituds de terabyte o superiors. Investigadors de la Universitat de San Diego de Califòrnia van examinar les dades que rebien les llars americanes i van determinar que el 2008 van ser bombar-

dejades amb 3,6 zettabytes d'informació (o 34 gigabytes per persona/dia), de manera que l'aportació d'aquesta investigació ens indica la veritable complexitat i abast del problema.

Big data és definit per Kusnetzky (2010) com el conjunt d'eines, processos i procediments que permeten a una organització crear, manipular i gestionar conjunts de dades molt grans, i també les instal·lacions d'emmagatzematge per a aquestes dades. D'altra banda, Dans (2011) fa una aproximació a partir del volum i la gestió de les dades natives indicant que *big data* és:

“el tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales.”

L'entrada de la Wikipedia (2011) estableix *big data* com un terme aplicat a conjunts de dades que superen la capacitat del programari habitual per a ser capturades, gestionades i processades en un temps raonable. Les grandàries del *big data* es troben constantment en una tendència creixent i exponencial: així, el 2012 es trobava dimensionat en una grandària d'una dotzena de terabytes fins a diversos petabytes de dades en un únic *data set*.

Per la seva banda, Martínez López (2012) afirma que el fet de tractar amb grandàries d'informació fora de l'habitual en realitat es converteix en un problema. En primer lloc, d'**escalabilitat**: podem continuar creixent?, fins a quant? En segon lloc, de **tractament**: podem tractar aquestes dades?, i quan en tinguem el doble? I en tercer lloc, de **disponibilitat**: és viable mantenir en línia totes aquestes dades? La resposta tècnica a aquests problemes existeix: **l'estructuració de la informació**.

2. Els llenguatges de marcatge en la representació i l'organització del coneixement

Un llenguatge de marcatge és un conjunt de regles que estableixen quin tipus de marques s'han d'utilitzar, de quina manera es distingiran les marques del text dels documents i com s'inseriran aquestes marques (la gramàtica i la sintaxi), i quines són les marques permeses en cadascuna de les parts del text. De manera genèrica, es poden distingir dos tipus de llenguatges de marcatge:

- **Els llenguatges de marcatge procedimentals:** orientats a la presentació dels documents, especifiquen com ha de ser processat el text perquè surti per diversos mitjans (pantalla ordinador, impressora, etc.) Aquests llenguatges no aporten informació de tipus semàntic o estructural; són poc flexibles, atès que qualsevol canvi en la presentació del document implica modificar-ne el marcatge; i solen ser llenguatges específics d'un sistema de processament propietari, la qual cosa redueix la "portabilitat" d'aquests documents.
- **Llenguatges de marcatge descriptius:** orientats a la descripció formal i de contingut dels documents. Aquests llenguatges aporten informació sobre l'estructura del document i en descriuen el contingut informacional, a més, són llenguatges més flexibles, que diferencien entre el contingut real i la representació del document.

Exemple

Alguns exemples d'aquests llenguatges són RTF (*rich text format*) de Microsoft i PDF (*portable document format*) d'Adobe.

Exemple

Alguns exemples són SGML, HTML, XML, etc.

2.1. La família de normes SGML/XML

El llenguatge de marcatge SGML es va començar a gestar el 1969, quan un equip d'investigadors d'IBM van desenvolupar GML (*generalized markup language*). GML es va desenvolupar com una via per a crear la documentació bàsica de l'empresa en un format electrònic transferible i fàcil de gestionar. Es tracta del primer llenguatge no propietari de marcatge de text (independent del sistema en el qual es creen els documents i de la plataforma en la qual circulen) capaç de definir les estructures lògiques de qualsevol tipus de document basant-se en una sèrie de normes.

GML va aconseguir l'estatus de norma ANSI (American National Standards Institute) i amb la denominació d'*SGML (structured generalized markup language)* el 1978. Aquesta norma va ser adoptada com a norma ISO 8879 el 1986 i forma part d'un conjunt més ampli de normes amb la denominació genèrica d'*information processing - text and office systems - standard generalized markup language*. A més d'aquest conjunt de normes genèriques, la literatura científica

recent s'ha començat a referir a la família de normes SGML/XML com el conjunt de normes o especificacions subsidiàries o complementàries que constitueixen un grup normatiu entorn de la ISO 8879 de 1986.

Dos exemples molt clars d'aquesta família i lligats directament al desenvolupament de SGML són, d'una banda, la norma ISO 10744-1992 *Information technology - hypermedia / Time-based Structuring Language*, que proporciona normes específiques per a establir enllaços hipertext en documents SGML; de l'altra, la norma ISO 10179-1996 *Document Style Semantics and Specification Language*, que especifica el sistema de definició d'estils.

Hi ha tres característiques d'SGML que el distingeixen d'altres llenguatges de marcatge:

- D'una banda, SGML posa més èmfasi en el marcatge descriptiu que en el procedimental.
- De l'altra, introdueix el concepte de *tipus de document (document type)* i, per extensió, el concepte de *definició de tipus de document (DTD - document type definition)*, que s'empra per a definir un tipus de document d'acord amb les parts constituents i l'estructura lògica que adopten aquestes parts.
- Finalment, SGML és independent de qualsevol sistema de representació de l'alfabet en què estigui escrit el text.

2.2. Les DTD de SGML

Una DTD (definició de tipus de document) defineix l'estructura d'un tipus de document específic que comprèn diversos aspectes: els elements que poden formar part del tipus de document, el nom dels elements i si són repetibles, l'ordre dels elements, els continguts dels elements, quins tipus de marcatge poden ser omesos, els atributs i els valors per defecte i els noms de les entitats permeses.

Una DTD es pot desenvolupar per a un document en concret o per a molts documents. La creació d'una DTD per a un sol document no és eficient, per tant, té molt més sentit crear DTD que es puguin usar per a molts documents. Des de l'aparició d'SGML s'han creat nombroses DTD per a ús general. Alguns exemples de les DTD d'SGML que han tingut més repercussió són:

1) HTML (*hypertext markup language*)

El W3C ha desenvolupat una DTD d'SGML destinada a la presentació de la informació per mitjà d'Internet. L'HTML anava destinat al gran públic per la senzillesa, per tant, es va fer prevaler l'aparença visual, sense tenir en compte la importància de marcar l'estructura lògica d'un document electrònic. A partir de l'HTML 4.0 s'incorporen els elements META de l'HTML, que es poden usar per a descriure les propietats d'un document i possibiliten la incorporació de les metadades Dublin Core.

2) TEI (*text encoding initiative*)

Sorgeix el 1987 com un projecte de l'àrea de les humanitats, promogut en un congrés de l'Association for Computers and the Humanities (ACH), però la seva publicació efec-

tiva per a la codificació de textos, fonamentalment literaris, és posterior (1994). Es tracta d'una DTD (definició del tipus de document) madura i ben formada d'SGML per a assegurar un format estàndard amb àmplies capacitats de marcatge per a la indexació i l'intercanvi d'informació textual.

L'esquema TEI presenta alguns avantatges en relació amb les particularitats de la metainformació necessària per a la documentació de recursos informatius electrònics a causa del seu control documental. TEI es compon d'un encapçalament seguit del text pròpiament dit. L'encapçalament consta, al seu torn, de quatre elements que proporcionen informació (principalment bibliogràfica) relativa al document com a fitxer electrònic.

L'encapçalament TEI constitueix una de les contribucions principals per a la codificació SGML d'informació bibliogràfica. A més, les directrius TEI inclouen una secció especial (24.3) sobre els elements d'aquest encapçalament i les seves relacions amb els registres MARC.

3) EAD (*encoded archival description*) DTD

La DTD d'EAD defineix una classe de documents, que, en termes generals, consten d'una pàgina de títol opcional, la descripció d'una unitat de material arxivístic i uns apèndixs també opcionals. Aquesta DTD possibilita l'ús de l'XML a partir de la versió 1.0 i ha estat dissenyada per a reflectir la jerarquia natural que presenta l'organització dels fons, en conjunció amb l'ordre intel·lectual que imposen els arxivers amb les seves pràctiques descriptives.

4) MARC (*machine readable catalogue*) DTD

La DTD de MARC tracta el format MARC com un tipus de document específic. La DTD defineix tots els elements que poden aparèixer en un registre MARC i especifica com s'han de codificar i representar en SGML. L'objectiu principal de MARC DTD va ser que els registres es poguessin importar i exportar a SGML automàticament. Per a això s'ha seguit l'estructura d'un registre MARC, incloent-hi les etiquetes, els indicadors i els codis de subcamp.

2.3. El llenguatge XML

L'XML és un projecte del World Wide Web Consortium (W3C) i el desenvolupament està coordinat per l'XML Working Group. Es tracta d'un subconjunt (o simplificació) adaptat d'SGML que té la intenció d'aprofitar-ne al màxim els avantatges, però permetent-ne la implementació a Internet.

El llenguatge XML consta de quatre especificacions (recomanacions de W3C):

1) **DTD (*document type definition*)**: definició de tipus de document. Es tracta d'un arxiu que tanca una definició formal d'un tipus de document i, alhora, especifica l'estructura lògica de cada document. El DTD en XML és opcional. En tasques senzilles no cal construir una DTD; llavors es tracta d'un document "ben format" (*well-formed*). Si es construeix una DTD, llavors serà un document "validat" (*valid*).

2) **XSL (*extensible stylesheet language*)**: llenguatge d'estil per a l'XML. Es tracta d'un llenguatge per a elaborar fulls d'estil. Consta de tres parts: *XSL transformations* (XSLT), que és un llenguatge per a transformar documents XML; *XML path language* (XPath), que és un llenguatge d'expressió usat per XSLT per a accedir o referir parts d'un document XML (XPath s'usa també en l'especificació XML Linking); *XSL formatting objects* és un vocabulari per a especificar la semàntica del format.

3) **XLL (*extensible linking language*)**: llenguatge d'enllaços en XML. Defineix la manera d'actuar entre diferents enllaços. Es considera un subconjunt de HyTime (ISO 10744) i segueix algunes especificacions de TEI. XLL té dos components importants: XLink i XPointer, amb els quals es va molt més allà dels enllaços simples suportats per l'HTML, ja que els enllaços poden ser bidireccionals, múltiples (anells, múltiples finestres, etc.) i agrupats (múltiples orígens).

4) **XUA (*XML user agent*)**: Estandardització de navegadors XML. S'aplica als navegadors perquè siguin capaços de reconèixer totes les especificacions XML.

Les característiques essencials de l'XML són:

- L'XML és una arquitectura oberta i extensible. No necessita versions per a funcionar en futurs navegadors.
- L'XML posseeix més consistència, homogeneïtat i amplitud dels identificadors descriptius del document (RDF), en comparació dels atributs de l'etiqueta <META> de l'HTML.
- L'XML permet agrupar una àmplia varietat de dades i aplicacions, des de pàgines web fins a bases de dades.
- L'XML aconsegueix que l'estructura de la informació resulti més accessible; per tant, els motors de cerca retornaran respostes més adequades i precises.
- L'XML permet el desenvolupament de cerques personalitzables per a robots i agents intel·ligents.
- L'XML desenvolupa àmpliament el concepte d'*hipertext* mitjançant enllaços bidireccionals, enllaços que es poden especificar i gestionar des de fora del document, hiperenllaços múltiples, enllaços agrupats, atributs per a enllaços, etc.
- L'XML facilita l'exportabilitat a altres formats de publicació (paper, web, CD-ROM, etc.).

2.4. Mecanismes per a representar el coneixement

Cada llenguatge de marcatge d'informació utilitza mecanismes diferents per a la representació del coneixement contingut en un document. En aquest apartat tractarem els dos llenguatges de marcatge més estesos per a la publicació de continguts a Internet: HTML i XML. Per a tots dos llenguatges determinarem els mecanismes que podríem denominar de *representació del coneixement documental* o *codificació de la metainformació documental*.

1) Dublin Core (DC)

La iniciativa de metadades Dublin Core és actualment una iniciativa consorciada per al desenvolupament del model de metadades Dublin Core, sorgit originàriament el 1995 en l'àmbit bibliotecari. El seu desenvolupament, d'acord amb l'evolució d'Internet, l'ha convertit en un format altament normalitzat i utilitzat en diferents sectors. El model DC es compon d'un conjunt de quinze elements que descriuen el coneixement contingut en un recurs d'una manera estructurada.

2) *Resource description framework* (RDF)

RDF és, sens dubte, l'estàndard en desenvolupament més important per a la descripció de continguts web. La seva importància resideix, d'una banda, en l'entitat Consorci Web (W3C), que fomenta el projecte, i de l'altra, el fet de ser una aplicació de metadades que utilitza l'XML a fi de proporcionar un marc estàndard per a la interoperabilitat entre diferents models de metadades per a la representació del coneixement (metainformació) contingut en un document (com per exemple el DC, que admet les especificacions RDF a partir de la versió DC5).

Des que RDF es va convertir en una recomanació de W3C el febrer de 1999, s'han desenvolupat un bon nombre d'eines que permeten treballar amb RDF d'una manera més eficient.

RDF ofereix una varietat d'aplicacions que usen l'XML com a sintaxi d'intercanvi, com ara: catàlegs de biblioteca i directoris web (exemples: Dublin Core Metadata Initiative, OCLC Connexion, Open Directory Project, etc.); categorització i gestió de llistes de notícies, programari i continguts (PICS, XML-New, UK Mirror Service), i també col·leccions de música, fotos i esdeveniments (MusicBrainz Metadata Initiative, RDFPic, etc.).

2.5. Serveis d'informació web basats en XML

Les organitzacions que presten serveis d'informació han de dissenyar sistemes eficaços per a gestionar i difondre la informació. Amb l'aplicació de les tecnologies de la informació en el desenvolupament i la implementació de sistemes d'informació i la utilització de la tecnologia web, un percentatge molt alt dels serveis d'informació es fonamenten en informació electrònica. En aquest context la família de normes SGML resulta clau, ja que assegura la consistència (en la producció, el processament, l'emmagatzematge i la distribució) i proporciona una enorme flexibilitat (en la presentació i en el format, per exemple).

Els professionals de la informació estan molt interessats en el potencial de l'XML per a l'organització de la informació. Un dels camps de més activitat consisteix en la substitució de MARC usant tecnologia XML, de fet la Library of Congress i altres organitzacions ja han dut a terme aquest tipus de projectes.

Cada dia són més nombrosos els projectes XML que s'estan emprenent en l'àmbit dels serveis d'informació comercial. Es tracta d'aplicacions relacionades amb dominis molt variats: banca, telecomunicacions, revistes electròniques, etc. Aquesta proliferació d'aplicacions XML no significa que aquest acabi substituint l'SGML. En principi, mantenir una o diverses bases de dades SGML i filtrar-les a XML per a les operacions d'indexació, altres tractaments de les dades i distribució en el web, es conforma com una estratègia a llarg termini amb gran possibilitat d'èxit. De fet, la majoria de les biblioteques digitals mantenen les seves dades en SGML, però distribueixen la informació en XML.

2.6. L'XML en l'organització del coneixement intern

La literatura científica recent se centra en els avantatges de l'XML per al desenvolupament de llocs web i la seva aplicació a les tecnologies de cerca i recuperació d'informació web. En aquesta ocasió cal tenir en compte els usos de l'XML per a la gestió del coneixement organitzacional. L'XML és un llenguatge que presenta simultàniament el contingut per a ser publicat a Internet i el descriu de manera que un altre programari pot entendre i usar aquest coneixement. La clau de l'XML és que, en oposició a l'HTML, proporciona informació sobre el significat de les dades, la qual cosa facilitarà el processament automàtic de la informació continguda en la informació (metainformació) i la consegüent gestió del coneixement.

L'XML es comença a considerar com "la següent revolució en la gestió del coneixement", i les organitzacions comencen a entendre el potencial d'aquesta tecnologia per al desenvolupament d'arquitectures d'informació corporatives.

La tecnologia XML per si mateixa no reporta cap valor a l'organització, sinó que el valor depèn de com s'usi aquesta tecnologia dins de l'organització. La implementació no ha de ser departamental i, idealment, hauria d'incloure *partners* estratègics i altres organitzacions amb què es necessita compartir dades i informació.

Un exemple clar dels beneficis de l'XML en la gestió de la informació corporativa consisteix en la creació de portals corporatius que ofereixen prestacions per a diferents funcions de negoci, com ara ERP, emmagatzematge de dades (*data warehousing*), sistemes de suport a la decisió i gestió del coneixement organitzacional.

Entre les aplicacions estratègiques que l'XML pot tenir dins d'una organització podem destacar:

1) Cadena de subministrament: la gestió eficient del coneixement organitzacional depèn de l'accés a la informació externa. L'XML pot ajudar a millorar el funcionament de la cadena de subministrament i l'extranet de l'organització. Moltes organitzacions necessiten compartir informació i no sempre tenen una via per a fer-ho, ja que tenen infraestructures tecnològiques diferents i no compatibles amb alguns agents de la cadena de subministrament o amb els clients. L'XML pot proporcionar una forma estàndard per a intercanviar informació amb proveïdors, distribuïdors i clients, la qual cosa es converteix en una diferència competitiva per a l'organització capaç de gestionar la cadena de subministrament mitjançant XML.

2) Retorn d'inversions (ROI): els documents XML poden produir diferents *outputs* i, una vegada la informació organitzacional està organitzada en documents XML, es poden elaborar multitud de productes i serveis d'informació via web, i també en diferents plataformes: PDA, telefonia mòbil, televisió digital, etc. La creació d'un corpus de coneixement organitzacional en XML suposa un gran esforç d'inversió, però aviat reporta un estalvi en els costos directes relacionats amb la inversió en tecnologies de la informació, la formació dels empleats i els serveis de suport.

3) Gestió integral del coneixement: l'XML posseeix un conjunt de característiques que el converteixen en un aliat per a la gestió del coneixement organitzacional. Si es combina el potencial de l'XML i les possibilitats de l'RDF, com a eina per a la gestió del coneixement que permet organitzar, interrelacionar, classificar i anotar el coneixement contingut en els documents XML, es pot obtenir un increment considerable del valor afegit de les dades emmagatzemades en els documents XML i comporta una gestió automàtica del coneixement explícit de l'organització.

3. Arquitectura de la informació, navegació i recuperació

Els serveis electrònics generalment adopten una estructura hipertextual i un format web que poden ocasionar una sèrie de problemes quant a la navegació pels continguts i la recuperació d'informació útil. Aquests problemes solen estar relacionats amb la sobrecàrrega cognitiva i la desorientació. Per superar aquests problemes els sistemes de classificació, com a eines per a l'organització de continguts web, contribueixen a facilitar la recuperació d'informació i l'aprenentatge d'estratègies de navegació. Referent a això, la característica principal de l'aprenentatge que es produeix per mitjà de la navegació hipertextual es fonamenta en la possibilitat d'organitzar determinats coneixements segons estructures diferents que permetin a l'usuari aconseguir el seu objectiu.

Tradicionalment, un sistema de classificació es refereix a un llenguatge documental que organitza lògicament una estructura de conceptes o notacions, i que està destinat a permetre la classificació de documents conforme a les respectives temàtiques. El sistema de classificació, com tot llenguatge documental, estableix tres tipus de relacions bàsiques entre els conceptes que l'integren: de sinonímia o identitat, jeràrquiques o associatives. Quant a la tipologia poden ser molt diversos: per la cobertura temàtica poden ser universals o especialitzats; poden tenir baix, mitjà o alt nivell d'especificitat segons el detall i la profunditat del desenvolupament, i es poden classificar com a enumeratiu, prefaceta o facetat, d'acord amb les característiques estructurals. En general, els sistemes de classificació són aplicables a un univers indeterminat d'objectes i per això afavoreixen l'organització lògica d'un àmbit de coneixement.

3.1. L'arquitectura de la informació

En l'àmbit de la informació electrònica i l'arquitectura de continguts web s'entén per *classificar* l'activitat d'agrupar els elements d'informació d'acord a atributs o propietats comunes entre ells. Els elements d'informació en realitat són continguts, és a dir, els "trossos" d'informació o objectes informatius a organitzar, estructurar i classificar, que poden ser: textos, imatges, vídeos, entre d'altres. Per tant, un sistema de classificació consisteix a triar sobre la base de quins atributs s'agrupen els continguts i com s'organitzen aquests atributs.

Seguint Rosenfield i Morville, els sistemes de classificació poden ser exactes o ambigus. En els sistemes de classificació exactes, els continguts estan perfectament definits i diferenciats de la resta.

Un exemple d'aquest tipus de sistemes és la classificació alfabètica, la cronològica i la geogràfica.

En els sistemes de classificació ambigus, els continguts estan organitzats en categories no definides de manera exacta i precisa.

Un exemple d'aquest tipus de sistemes són la classificació temàtica o per categories; la classificació orientada a tasques; la classificació orientada a l'audiència o tipologia d'usuaris, i la classificació metafòrica.

Tots són sistemes vàlids per a organitzar els continguts web d'una organització i s'utilitzen actualment per a facilitar l'accés a la informació corporativa, i als serveis i productes de les organitzacions públiques.

La visió tradicional de sistemes de classificació enllaça amb la perspectiva actual de la disciplina d'arquitectura de la informació. El terme *arquitectura de la informació* el va utilitzar per primera vegada Richard Saul Wurman el 1975, el qual la defineix com l'estudi de l'organització de la informació amb l'objectiu de permetre a l'usuari trobar la seva via de navegació cap al coneixement i la comprensió de la informació. Per tant, l'arquitectura de la informació restringida a l'àmbit web seria l'art i la ciència d'estructurar i classificar llocs web amb la finalitat d'ajudar els usuaris a trobar i manejar la informació. D'aquesta definició es desprèn que l'arquitectura de la informació d'un lloc web comprèn els sistemes d'organització i estructuració dels continguts, els sistemes de retolació o etiquetatge d'aquests continguts, i els sistemes de recuperació d'informació i navegació que proveeixi el lloc web.

Si es té en consideració la visió tradicional dels sistemes de classificació i es combinen amb les noves tendències quant a sistemes de classificació en l'entorn web, les eines tradicionals, com les llistes de termes, les taxonomies i els tesaurus, entre d'altres, es poden utilitzar per a classificar i organitzar el coneixement electrònic, fonamentalment en els sistemes de classificació ambigus. L'arquitectura de la informació inclou molts aspectes diferents sobre la creació i l'organització d'un lloc web, però les eines principals són les tècniques d'organització d'informació desenvolupada en altres disciplines. De fet, la majoria d'aquestes tècniques provenen de la documentació, en particular respecte a vocabularis estructurats.

3.2. Sistemes de classificació del coneixement electrònic

Per a dur a terme les tasques de classificació dels continguts en un lloc web corporatiu, es poden utilitzar diferents eines de classificació temàtica o per matèries, que contribuiran a l'orientació de l'usuari dels serveis web i a la millora en la recuperació d'informació, ja que aquesta ha estat organitzada *a priori* mitjançant llistes de categories, taxonomies, tesaurus o classificacions facetades, d'acord amb les característiques del servei o producte d'informació web. Totes les eines esmentades són vocabularis controlats i estructurats que s'utilitzen per a etiquetar continguts i per a localitzar-los posteriorment mitjançant sistemes de navegació o cerca. No obstant això, la definició específica d'aquestes eines és molt difusa i no hi ha una definició única i consensuada que permeti diferenciar clarament, per exemple, entre *llista de categories* i *taxo-*

nomia, atès que les definicions existents en la literatura científica tendeixen a ser molt heterogènies entre elles. Per tant, s'aporta una definició molt escarida i genèrica que s'adapti al propòsit d'aquest article.

Una llista de categories és una llista controlada de termes que denota una matèria i que s'utilitza per a classificar, indexar i recuperar recursos d'informació. La funció de la llista de categories consisteix a evitar a l'usuari definir termes sense un significat específic, termes massa amplis o excessivament restrictius, o termes molt semblants quant a la forma.

El terme **taxonomia** prové de l'àrea de les ciències naturals, en particular de la biologia, que estudia la classificació dels éssers vius en estructures jeràrquiques. Es tracta d'un terme molt utilitzat en arquitectura de la informació i, segons en quin context s'utilitzi, se li podria assignar un significat o un altre. De manera molt esquemàtica, una taxonomia és una llista estructurada i jeràrquica d'elements o grups d'elements que presenta una forma arbòria. Una taxonomia és l'organització jeràrquica del conjunt de categories (paraules clau) amb les quals classificar les unitats de contingut.

Per exemple, en el cas d'un servei d'administració electrònica basat en esdeveniments de la vida, s'identifiquen dins la taxonomia "estudis", les categories: "estudis primaris", "estudis secundaris", "estudis universitaris" i "altres estudis".

Per tant, la funció de la taxonomia en l'organització dels continguts web resideix en la possibilitat d'agrupar termes jeràrquicament relacionats amb una taxonomia que faciliti trobar el terme correcte per a la cerca o descripció d'un contingut específic.

Els **tesaurus** es poden considerar una taxonomia amb extres, ja que permeten representar la realitat mitjançant termes no solament organitzats de manera jeràrquica, sinó que permeten un altre tipus de relacions entre ells, com són la relació d'equivalència i d'associació. Els tesaurus estan normalitzats mitjançant la norma ISO 2788, en la qual es defineix *tesaurus* de la manera següent:

"un vocabulari controlat i dinàmic, compost per termes que tenen entre ells relacions semàntiques i genèriques i que s'aplica a un domini particular del coneixement."

A més de l'esmentada funció de control del vocabulari, el tesaurus constitueix una eina que representa, d'una banda, l'estructura dels coneixements transmesos pels creadors de la informació (arquitectes de la informació), i, de l'altra, l'estructura cognitiva dels usuaris que formulen l'expressió de cerca.

La **classificació facetada** parteix de la premissa que una unitat de contingut es pot descriure mitjançant diverses dimensions o facetes, cadascuna de les quals conté la seva pròpia relació de possibles valors o categories. La classificació per facetes va ser proposada el 1933 per Ranganathan, i és també coneguda com a *classificació colonada*. Les classificacions facetades s'utilitzen per a organitzar conjunts d'objectes (llocs web, productes i documents, entre d'altres), amb prou homogeneïtat perquè puguin ser descrits per un nombre determi-

nat d'atributs o propietats (facetes i categories) i els seus valors (pertinença a categories). En principi, la classificació per facetes sembla molt diferent d'un tesaurus, però, de fet, una classificació facetada es pot considerar una via adequada per a construir un tesaurus o bé per a ser utilitzada amb propòsits de classificació de continguts. En l'àmbit dels serveis web existeix un llenguatge amb sintaxi XML per a definir, distribuir i intercanviar metadades en forma de taxonomies facetades. Aquest llenguatge ha estat desenvolupat per Peter van Dijck i es denomina *XFML* (*exchangeable faceted metadata language*).

Les **ontologies** són els instruments o models d'estructures amb una traducció tecnològica que es poden usar per a la presentació i ús en web de catàlegs, bases de dades, llistes d'encapçalament, glossaris, tesaurus, per exemple; i impliquen una especificació, conceptualització i modelització del coneixement que es recull en cadascun dels llenguatges documentals. Com s'observa per la definició aportada, es pot establir un cert paral·lelisme entre *ontologia* i *tesaurus*; fonamentalment, en la delimitació dels termes que han de formar part d'una construcció de representació del coneixement, i de les relacions que s'estableixen entre els termes.

3.3. Navegació i recuperació de la informació web

Hi ha dues formes bàsiques d'arquitectura de la informació en què s'utilitza control del vocabulari per a la classificació de continguts amb l'objecte d'organitzar el coneixement electrònic: els esquemes de navegació i els sistemes de recuperació. Els esquemes de navegació no poden ser ambigus, i s'utilitzen per a orientar els usuaris sobre la tipologia dels recursos d'informació disponibles en un lloc web. En els sistemes de recuperació d'informació, l'usuari utilitza termes o frases per a expressar la seva necessitat d'informació, en general, amb un formulari de consulta que pot incloure un o més camps. La qüestió és que totes dues formes d'accés a la informació són necessàries, ja que són complementàries, i totes dues necessiten cert control del vocabulari.

La informació s'ha d'organitzar d'una manera coherent, de manera que els usuaris puguin accedir a la informació desitjada i important per a ells d'una manera ràpida i eficaç. L'organització de la informació és essencial perquè el sistema d'informació sigui útil.

L'arquitectura de la informació d'un lloc web comprèn els sistemes d'organització i estructuració dels continguts, els sistemes de retolació o etiquetatge d'aquests continguts, i els sistemes de recuperació d'informació i navegació que proveeixi el lloc web.

Un bon sistema d'informació requereix alguna cosa més que la matèria primera. I aquesta necessitat d'organització és cada vegada més clara a mesura que l'usuari disposa de menys temps per a poder dedicar a moure's per les fonts. L'organització de la informació és essencial perquè el sistema d'informació sigui útil.

Exemple

L'organització de la informació en un sistema, per exemple en una pàgina web, és responsabilitat de l'arquitecte d'informació.

4. Tecnologies, programes i formats

4.1. Els servidors

Un servidor és una aplicació informàtica o programa que fa algunes tasques en benefici d'altres aplicacions anomenades *clients*. Alguns serveis habituals són els serveis d'arxius, que permeten als usuaris emmagatzemar i accedir als arxius d'un ordinador i els serveis d'aplicacions, que duen a terme tasques en benefici directe de l'usuari final. Aquest és el significat original del terme. És possible que un ordinador compleixi simultàniament les funcions de client i de servidor.

L'arquitectura **client-servidor**, anomenada *model client-servidor*, és una forma de dividir i especialitzar programes i equips de còmput a fi que la tasca que cadascun d'ells realitza s'efectui de la manera més eficient possible i permeti simplificar les actualitzacions i el manteniment del sistema.

En aquesta arquitectura la capacitat de procés està repartida entre el servidor i els clients.

En la funcionalitat d'un programa distribuït es poden distingir tres capes o nivells:

- 1) Manejador de bases de dades (nivell d'emmagatzematge)
- 2) Processador d'aplicacions o regles del negoci (nivell lògic)
- 3) Interfície de l'usuari (nivell de presentació)

Un **servidor web** és un programa que implementa el protocol HTTP (*hypertext transfer protocol*). Aquest protocol està dissenyat per a transferir continguts web o pàgines HTML (*hypertext markup language*): textos complexos amb enllaços, figures, formularis, botons i objectes incrustats com animacions o reproductors de sons.

No obstant això, el fet que l'HTTP i l'HTML estiguin íntimament lligats no ha de provocar que es confonguin els dos termes. *HTML* és un format d'arxiu i *HTTP* és un protocol.

Cal destacar el fet que la paraula *servidor* identifica tant el programa com la màquina en la qual aquest programa s'executa. Hi ha, per tant, certa ambigüïtat en el terme, encara que no serà difícil diferenciar a quin dels dos sentits ens referim en cada cas.

Un servidor web s'encarrega de mantenir-se a l'espera de peticions HTTP dutes a terme per un client HTTP que se sol conèixer com a *navegador web*. El navegador fa una petició al servidor i aquest li respon amb el contingut que el client sol·licita.

Sobre el servei web clàssic es poden desenvolupar diverses aplicacions web. Aquestes són fragments de codi que s'executen quan es fan certes peticions o respostes HTTP. Cal distingir entre:

- **Aplicacions en el costat del client:** el client web és l'encarregat d'executar-les en la màquina de l'usuari. Són les aplicacions tipus Java o JavaScript: el servidor proporciona el codi de les aplicacions al client i aquest, mitjançant el navegador, les executa. Cal, per tant, que el client disposi d'un navegador amb capacitat per a executar aplicacions (també anomenades *scripts*). Normalment, els navegadors permeten executar aplicacions escrites en llenguatge JavaScript i Java, encara que s'hi poden afegir més llenguatges mitjançant l'ús de connectors o *plugins*.
- **Aplicacions en el costat del servidor:** el servidor web executa l'aplicació; aquesta aplicació, una vegada executada, genera cert codi HTML; el servidor pren aquest codi recentment creat i l'envia al client per mitjà del protocol HTTP. Les aplicacions de servidor solen ser l'opció per la qual s'opta la majoria de les vegades per a fer aplicacions web. La raó és que, com que l'aplicació s'executa en el servidor i no en la màquina del client, aquest no necessita cap capacitat addicional, com sí que passa en el cas de voler executar aplicacions JavaScript o Java. Així doncs, qualsevol client dotat d'un navegador web bàsic pot utilitzar aquest tipus d'aplicacions.

Exemple

Alguns servidors web importants són: Apache i IIS.
Algunes tecnologies relacionades amb les aplicacions web són: PHP, ASP, JSP, CGI i .NET, entre d'altres.

4.2. Cloud computing: Internet/intranet

Les tecnologies de la informació entorn d'Internet evolucionen amb gran rapidesa, fins i tot apareixen i desapareixen sense percebre's els canvis que han implicat en la construcció de nous paradigmes a la Xarxa de xarxes. Les intranets no estan blindades a aquests canvis i durant aquesta primera dècada del segle XXI han sorgit un conjunt de tecnologies que incideixen directament en el seu model de disseny.

La font de coneixement cooperatiu Wikipedia (2011) descriu una intranet de la manera següent:

“una xarxa d'ordinadors privats que utilitza tecnologia Internet per a compartir dins d'una organització part dels seus sistemes d'informació i sistemes operacionals.”

La concepció d'una intranet evoluciona a l'empara de l'avenç de les tecnologies de la informació i comunicació a Internet (Pérez i Solana, 2006). Aquestes tecnologies són la base del model intranet, com es desprèn de la seva pròpia definició.

A partir d'aquesta succinta descripció, part d'aquest treball es marca com a objectiu conèixer el punt en què es troben les intranets, tant en l'àmbit conceptual, metodològic com tecnològic, i oferir una nova visió amb la incorporació de noves tendències al particular univers d'intranet.

4.2.1. Visió retrospectiva d'intranet documental

En un primer estadi es mostren els resultats d'una anàlisi retrospectiva sobre un conjunt de treballs que aclareixen els principals canvis esdevinguts, al llarg de més de quinze anys, del concepte d'*intranet documental*. Per a això, es proposa fer tres parades cronològiques, destacant aquells treballs que ofereixen un instant concret de l'estat de les intranets i que, al seu torn, permetin establir un fil conductor de les diferents perspectives, sobre la tecnologia comentada, en l'àrea de documentació.

Aquesta mirada retrospectiva s'inicia l'any 1998 (Abadal, 1998), quan l'extingida Socadi va fer una sessió presencial i documentada per la revista *EPI*. En aquesta reunió un conjunt d'experts van opinar sobre un concepte emergent en aquell temps, *intranet*, i en un context documental. Aquests professionals i investigadors van dibuixar un estat de l'art i van exposar què s'entenia per *intranet* en un moment tecnològic i temporal concret. Lluís Codina indicava com a característica més destacable de la intranet “la gran capacitat d'integració”. Al seu torn, Codina referenciava com a bondats de la intranet els elements següents:

“des d'un sol programa (el navegador) es poden consultar informacions molt diferents (documents de totes classes, bases de dades, etc.); s'eliminen les diferències entre sistemes locals i remots; i apareixen noves possibilitats de relació entre els sistemes de recuperació de la informació i els hipertextuals, i també entre informació i comunicació.”

D'altra banda, Josep Àngel Borràs (1998) va definir que la característica bàsica de la intranet era:

“el seu model d'arquitectura de xarxa que usa protocols i estàndards oberts, transparent amb Internet i que disposa d'un client universal (el navegador).”

Endinsant-se i perfilant alguns serveis, va argumentar que la intranet era:

“l'objecte sobre el qual s'apliquen són tots els documents que una organització va produint (textos, correu electrònic, fulls de càlcul, etc.) i també totes les seves aplicacions (bases de dades, etc.).”

Finalment, cal destacar els comentaris d'Ernest Abadal (1998), que emfatitzava que els avantatges de posar en marxa una intranet residien en el cost, ja que es podia iniciar amb pocs mitjans. Al seu torn, va aprofundir en elements organitzacionals i va sostenir que les intranets ajudaven institucions amb organigrames en xarxa o piràmides de comandament planes i podrien funcionar molt bé.

L'any 2002, amb el títol "Intranets: sonrisas y lágrimas", Adela d'Alòs-Moner publicava un article en què exposava que els sistemes d'informació desenvolupats seguint models d'intranet estaven en plena explosió, resultat de factors com els següents:

- Aprofita la infraestructura existent de xarxes i ordinadors.
- Té un cost d'implantació baix.
- Es basa en una administració centralitzada.
- Es posa en funcionament ràpidament.

L'última parada, en aquesta revisió ràpida, es fa en l'article "Definición del modelo estratégico de desarrollo de una intranet corporativa: metodología basada en un caso práctico" (Bustelo-Ruesta i altres, 2008), treball en què s'exposa per primera vegada l'existència de cicles en el disseny i el desenvolupament d'una intranet. Estipulaven el següent:

- Una primera etapa en què la intranet destaca per ser una eina orientadora i facilitadora d'informació per a empleats, com un sistema per a la millora de comunicació vertical de l'organització i d'ajuda per a la publicació d'informació i documentació corporativa.
- En una segona etapa, les intranets se centren a millorar l'eficàcia i l'eficiència del coneixement intern, s'introdueixen elements per a augmentar la direccionalitat en la comunicació, entre persones i unitats organitzatives i apareixen les primeres eines per a descentralitzar la intranet.
- I l'últim cicle és una etapa centrada en la reflexió i redisseny des d'una perspectiva estratègica de la intranet en l'organització, el focus de la qual se centra en la gestió global i integrada de la informació de l'organització.

Aquesta expeditiva revisió, sense aprofundir en més detalls, ni metodològics ni tecnològics, ens fa plantejar dues qüestions sobre el seu ús.

- Primera qüestió: com es pot fer pública la informació treballada de manera interna?, quina tecnologia permet que sigui fàcil sense trencar el concepte *intranet*? Per tal de resoldre aquestes preguntes, aquest treball planteja una

solució híbrida, centrant-se en un exemple d'ús d'una tecnologia concreta basada en **SharePoint**.

- Segona qüestió: el concepte d'*intranet*, a més de ser un espai privat, es basa en tecnologia d'Internet. Què passa quan aquesta tecnologia evoluciona?, quines són aquestes noves tecnologies que poden incidir en les intranets? Per a aquests plantejaments, l'article se centrarà a estudiar els punts en comú entre intranets i *cloud computing* i presentarà un concepte "nou" d'intranet basat en l'anomenat *núvol privat*.

4.2.2. SharePoint, una solució híbrida

Segons un estudi realitzat per Forrester Research (Kisker, 2010) en el quart trimestre de 2009, el 24% de grans empreses i pimes d'Amèrica del Nord i Europa es manifesten atretes pel SaaS o programari com a servei alternatiu als desplegaments en les seves instal·lacions, especialment pel que fa al concepte d'*intranet*. Els avantatges relacionats amb l'estalvi en servidors i costos administratius, alliberament dels recursos humans TI i la possibilitat de dedicar-los a qüestions més crítiques formen una sèrie de valors a favor de la informàtica en núvol, com és el cas que ens ocupa.

Si bé és cert que a primera vista dissenyar una estratègia per a "pujar" serveis clàssics al núvol, com el correu electrònic, no representa una gran complexitat, en el moment en què ens plantegem aquesta acció per a serveis complexos i de més envergadura com una intranet, el procés és molt més complex, ja que hem de tenir en compte una sèrie de paràmetres que requereixen una anàlisi exhaustiva i que afecten elements com el disseny de la gestió de documents, el desenvolupament d'aplicacions pròpies, les seqüències de col·laboració o les polítiques de publicació interna i externa de continguts de la intranet corporativa, entre d'altres.

Partint d'un dels models de creació i gestió d'intranets corporatives més desplegat en organitzacions com és SharePoint, tal com es desprèn en els últims informes de Jakob Nielsen "10 Best Intranets" dels anys 2009, 2010 i 2011, podem afirmar que fins a l'última versió no oferia prou garanties en l'àmbit de seguretat, conformitat, privacitat i compatibilitat d'aplicacions per al seu model de núvol si es comparen amb els que sí que poden oferir en l'àmbit intern dins de la infraestructura de l'organització.

SharePoint (actualment en l'última versió 2010) és una plataforma de col·laboració organitzacional que es tradueix en el que podríem definir com un gestor de contingut empresarial que és capaç d'integrar-se amb les diferents eines de Microsoft (programari servidor, client de correu, paquet ofimàtic o

suite, etc.), la qual cosa li confereix un gran potencial per a gestionar una intranet, especialment en les organitzacions en què ja s'utilitzen productes de Microsoft.

D'aquesta manera, una organització pot utilitzar SharePoint per a facilitar la col·laboració, configurar els privilegis d'administració del contingut, implementar processos basats en aplicacions a mida i donar accés a la informació imprescindible per a complir els objectius i processos de l'organització. És lògicament molt més que un gestor de contingut CMS pensat per a entorns organitzatius per a crear i gestionar una intranet o una extranet, ja que engloba més d'un potent gestor documental, un sistema de polítiques de publicació (fluxos de treball, plantilles d'edició, col·laboració grupal, escenaris de portals, etc.) per a gestionar la nostra organització tant internament com externament.

És precisament aquest punt en el qual, de totes les característiques de la solució SharePoint, volem destacar el model que dóna resposta a la primera qüestió plantejada prenent com a exemple el nostre cas concret dins de l'Oficina Oberta d'Innovació de la UOC.

Per a això ens focalitzem en dos elements clau dins de la solució, com són l'administració del contingut per a la generació de portals i els mecanismes de col·laboració per a la gestió de continguts.

1) Administració de contingut per a crear un portal web

Un dels avantatges de treballar en un entorn intranet és el fet de controlar de manera eficaç i eficient els fluxos de gestió de la documentació, les polítiques d'accés, modificació, eliminació i publicació, però al seu torn molta d'aquesta informació ha de formar part de terceres parts com pugui ser la publicació web, ja sigui per a un accés puntual en una pàgina web aïllada o en un portal vertical. En aquest sentit, SharePoint 2010 ens aporta moltes característiques útils per a dissenyar, distribuir i administrar aquest tipus de continguts per mitjà de models interns de publicació automàtica com portals de la intranet, llocs web de presència corporativa a Internet i portals departamentals aïllats sense dependència jeràrquica amb la resta de l'estructura de la intranet. Gràcies a això eliminem la duplicitat existent en molts casos quan ens enfrontem a la publicació de continguts que han de tenir una presència fora de la nostra intranet.

En treballar de manera totalment integrada, es poden establir processos paral·lels però amb finalitats diferents sota la mateixa estructura, definint quines "illes" documentals són les que tindran una visibilitat pública fora de la nostra intranet. La capacitat per a crear i publicar contingut web de manera automàtica, basant-nos en plantilles predefinides segons les necessitats de cada entorn, ens permet reduir el cost final principal (econòmic i temporal), i també les despeses indirectes resultants de l'administració de diversos llocs.

Vegem de manera resumida quins són els agents principals que ens han permès agilitar les polítiques de creació i publicació de contingut sense haver de crear costosos desenvolupaments web paral·lels:

a) Pàgines web basades en plantilles: utilitzant el sistema d'administració de contingut, podem crear, editar i configurar diferents tipologies de pàgina web per a cada necessitat. N'hi ha prou amb associar un patró de disseny mare a cada tipologia d'element documental i fer els canvis necessaris per a cadascuna quan volem assignar a un contingut una sortida web. D'aquesta manera, si hem de treballar com a creadors de contingut, podem crear, editar i publicar les plantilles de lloc (per a la visió intranet, vegeu la figura 3) per a satisfer les necessitats de publicació de contingut específic que pugui tenir la nostra organització.

Les plantilles del lloc del grup "Publicació web" permeten que els administradors dissenyin, distribueixin i gestionin els portals de la intranet de l'organització, els diferents llocs web corporatius amb presència a Internet, i també els portals departamentals o sectorials (vegeu la figura 4). Cadascuna d'aquestes plantilles disposen de característiques de publicació, inclosa una barra d'eines de modificació de pàgines, un editor de contingut enriquit i funcions de desprotecció que poden ser modificades *a posteriori* per l'administrador de grup. Per a una edició avançada dels estils de cada pàgina s'integra la solució Office SharePoint Designer 2010, que permet el control total sobre els diferents fulls d'estil aplicats a cada plantilla o a cada portal.

b) Lloc de notícies: es tracta d'una plantilla de lloc que ens permet crear una secció específica per a difondre notícies i proporcionar vincles a notícies de manera ràpida i fàcil. Permet autogestionar un arxiu per a les notícies quan deixen de tenir vigència i un format de sortida estàndard per als lectors de notícies RSS i un mòdul específic anomenat "Aquesta setmana en imatges" que ens permet generar un model visual basat en fotografies.

c) Lloc de publicació amb flux de treball: mitjançant aquest sistema podem crear de manera programada un lloc gestionat per mitjà de fluxos de treball d'aprovació. Gràcies a aquest element, s'estableix un patró automàtic de pre-producció que permet revisar un contingut per part d'un validador abans que passi a ser publicat tant de manera interna com pública.

d) Portal de col·laboració: si volem crear una jerarquia de llocs bàsica per a un portal d'intranet departamental, podem optar per aquesta solució, que normalment s'utilitza per a allotjar llocs de diferents grups. Des d'una pàgina principal s'accedeix a un lloc de notícies, un directori de llocs i un centre de recerca amb fitxes. Aquest tipus de plantilla ens pot ajudar a crear un portal departamental en el qual els empleats puguin col·laborar i publicar documents,



Figura 3. Imatge de gestió de fitxa de projecte des de la intranet



Figura 4. Model de fitxa des del portal sectorial públic

i també pàgines web sense necessitat de coneixements tècnics en edició web, i es pot fer des d'un simple document en Word mitjançant un procés automàtic de conversió per a la seva publicació web.

e) **Portal de publicació:** aquesta plantilla ens permet crear i gestionar una jerarquia de llocs bàsica per a un lloc orientat a Internet o un portal d'intranet de gran grandària. Aquest lloc inclou per defecte una pàgina principal, un sublloc de comunicats de premsa d'exemple, un centre de recerca i una pàgina d'inici de sessió, tots modificables en funció de cada necessitat i cada organització.

2) La col·laboració en la gestió de continguts

Aprofitant les capacitats que SharePoint aporta per a la creació de mòduls específics basats en les diferents necessitats, s'ha desenvolupat un mòdul específic per a la gestió de projectes que permet a cada gestor treballar dins de la intranet de grup amb el seu equip de treball i els membres de l'Oficina d'Innovació. Aquest mòdul proporciona les eines necessàries als equips de treball per a la gestió de la documentació i la informació de manera senzilla mitjançant la creació automàtica d'una àrea compartida de projectes gestionada amb un formulari senzill en el qual es donen d'alta les dades i integrants de cada projecte.

Cada àrea de projectes està dissenyada per a incloure la informació bàsica del projecte: equip de desenvolupament, descripció, objectius, estratificació de tasques mitjançant un diagrama de Gantt amb la definició de les fases del projecte, registre de riscos associats al projecte amb la possibilitat d'incloure els plans de contingència per a mitigar-los, i també un registre dels problemes associats al projecte en què poder anotar les responsabilitats i la seva solució en una línia de temps.

Així mateix, aquest mòdul ha estat especialment dissenyat per a aprofitar les capacitats de treball col·laboratiu que aporta SharePoint i que permeten:

- Compartir documents, contactes, tasques i calendaris de manera sincronitzada amb Outlook amb la intenció de fer el seguiment dels projectes de l'equip.
- Contribuir col·laborativament a l'intercanvi d'idees per mitjà d'un wiki, o per a col·laborar en un disseny d'equip, crear una enciclopèdia de coneixements o, simplement, recopilar informació rutinària del projecte de manera senzilla i àgil gràcies al fet que no és necessari tenir coneixements tècnics ni utilitzar cap programari específic.
- Compartir entrades en blogs de creació fàcil i alt nivell de personalització, on els comentaris a les entrades es poden activar (per defecte estan desactivats), amb la qual cosa es permet a diferents membres d'una comunitat intercanviar punts de vista, compartir suggeriments o notícies concretes.

- Fer el seguiment de les tasques del projecte, modificar les fites i supervisar els possibles desfasaments del projecte des de la visualització compartida del diagrama de Gantt.
- Establir reunions per mitjà de la gestió integrada dels calendaris compartits, i fins i tot planificar reunions de veu i vídeo a distància gràcies a la integració amb Office Communicator, que permet simultàniament diverses maneres de comunicació com missatgeria instantània, videoconferència, telefonia per VoIP, i l'ús compartit d'aplicacions remotes i transferència d'arxius entre els membres de l'equip.

Finalment, per poder treballar amb un únic sistema de gestió de la documentació sorgida durant el projecte, inclou una biblioteca de documents en la qual es pot emmagatzemar la documentació classificada per carpetes, amb la integració del control de versions dels documents, i un control d'edició que impedeix modificar documents protegits pels usuaris. També s'ha habilitat l'opció de treballar sense connexió en els arxius d'una biblioteca i en els elements de calendaris, contactes, tasques i discussions. D'aquesta manera, quan un usuari es torni a connectar, podrà actualitzar els arxius del servidor i tots els elements de llista s'actualitzaran automàticament.

4.2.3. Tecnologia en el núvol

No es pretén ser exhaustiu en l'anàlisi del concepte d'*informàtica en núvol*, o en la seva definició, criteris estratègics o requisits tecnològics. L'objectiu bàsic és compartir les característiques bàsiques per al seu estudi i entendre la seva capacitat camaleònica.

Informàtica en el núvol, tecnologia en núvol o *cloud computing* (en aquest treball s'empraran indistintament els diferents descriptors) es defineix a la Wikipedia de la manera següent:

“tipus de computació on tot el que pot oferir un sistema informàtic s'ofereix com a servei, de manera que els usuaris puguin accedir als serveis disponibles en el núvol d'Internet sense coneixements (o, almenys sense ser experts) en la gestió dels recursos que usen.”

La definició que ofereix la Wikipedia no és gaire aclaridora, però igual que succeeix amb el terme *intranet*, el concepte *cloud computing* es pot reduir a una mínima expressió com el programari que no s'executa en un entorn local. Aquesta descripció es pot considerar massa simplista, des del punt de vista d'aquest estudi, i en un intent d'aclarir el concepte, es defineix com a usuaris que accedeixen per mitjà de diferents dispositius (ordinadors, mòbils, tauletes...) a un conjunt d'informació i serveis directament allotjats en servidors que utilitzen la infraestructura d'Internet.

Per a apropar, reforçar i entendre més el concepte d'intranet i *cloud computing* es pot observar la figura 5 que il·lustra les diferències i semblances entre tots dos conceptes.

Figura 5



En una anàlisi ràpida dels termes *intranet* i *informàtica en el núvol*, es pot establir que tant els significats com les il·lustracions apunten que tots dos conceptes semblen el mateix, i la seva diferència es troba en la idea de l'accés privat que inclou el terme *intranet*, però hem de recordar que un dels models del *cloud computing* és privat.

A partir del que hem exposat, podem afirmar que *intranet* i *cloud computing* són conceptes formalment molt propers, ja que tots dos contextos utilitzen la mateixa arquitectura de xarxa basada en Internet, i posen a la disposició dels usuaris la informació i serveis de manera remota i deslocalitzada. Aquestes similituds es deuen al fet que l'evolució de les intranets es pot forjar aplicant models de tecnologia en el núvol. En definitiva, des de fa uns anys s'utilitza la tecnologia en el núvol per a crear o migrar intranets, el tema és com fer-ho de la millor manera possible.

D'aquest raonament sorgeix una pregunta evident: com pot existir la possibilitat que una intranet i el *cloud computing* puguin ser el mateix? o, segons altres punts de vista i analitzant l'arquitectura tecnològica, puguin ser alguna cosa totalment diferent. Voas i Zhang (2009) van escriure un editorial el títol del qual pot ser aclaridor per a resoldre aquesta ambigüitat, "Cloud Computing: New Wine or Just a New Bottle?" ("*Cloud computing*: vi nou o només una ampolla nova?"), treball en què convidaven, a partir de la lectura de tres articles, a respondre si la informàtica en el núvol era un nou paradigma, si era la tecnologia coneguda en un nou embolcall o realment és una tecnologia totalment nova. Aquesta ambigüitat permet a aquest treball obrir una via on es pot establir que el *cloud computing* és un nou paradigma tecnològic per a un embolcall antic com són les intranets (Erdogmus, 2009; Gartner, 2008; Leavitt, 2009).

4.2.4. Avantatges i inconvenients principals del *cloud computing*

La tecnologia en el núvol no requereix inversió en maquinari, ja que el servei està allotjat a Internet i s'hi accedeix remotament. El cost ve donat pel seu consum, per exemple, temps d'ús o funcionalitat, encara que hi ha serveis gratuïts, i es mantenen econòmicament amb altres fonts de finançament, com la publicitat i les donacions.

Un altre avantatge enfront dels sistemes d'informació tradicionals és la no-submissió a evolucions i actualitzacions del programari i maquinari, que amb l'aplicació de tecnologies en el núvol desapareix, és automàtica, sense cost i transparent a l'usuari, ja que la contractació o l'ús del servei assumeix aquestes millores. A més, aquests serveis ofereixen un avantatge entorn del cost de manteniment, ja que s'inclou en el paquet del proveïdor del servei, tant si aquest és local, regional o global.

L'ús d'informàtica en el núvol millora la racionalització d'ús, quan s'incorpora un ordinador/servidor a la intranet, aquest es troba condicionat per l'ús. La majoria de vegades estan infrautilitzats.

Per exemple, quants ordinadors funcionen les vint-i-quatre hores?, quants ordinadors aprofiten el 100% del potencial que tenen?

Amb la tecnologia en el núvol es paga el que s'usa, s'obté el que es necessita. Aquest tipus de servei s'anomena *pagament pel que uses (pay as you go)*.

Una millora possible del *cloud computing* es troba en l'accés únic per mitjà d'Internet, encara que si hi hagués cap problema amb el proveïdor de dades no hi ha serveis i, d'altra banda, si el proveïdor de servei té problemes amb la infraestructura tampoc s'oferirà accés.

L'entrada als serveis de tecnologies en el núvol és universal, s'hi pot accedir des de qualsevol punt geogràfic, igual que una intranet virtualitzada, però un inconvenient del *cloud computing* és la programació de manteniment, ja que són alienes a les necessitats del client, i no són programades segons els criteris dels usuaris, de manera que és possible trobar-nos amb alguna fallada en l'accés a un servei en un moment crític, encara que avui dia amb els sistemes redundats més que un temor real passa a ser un advertiment.

A més d'esmentar algunes bondats, no sense certs riscos, de les diferents característiques del *cloud computing*, també cal destacar algunes ombres importants sobre aquest entorn tecnològic. Els arguments més crítics que s'empren contra la tecnologia en el núvol són maduresa i seguretat. Actualment el *cloud computing* és un model emergent amb tecnologia en plena ebullició que impedeix en certa manera afrontar decisions tecnològiques sobre el canvi en aquest model i, d'altra banda, es comencen a detectar els primers atacs a serveis i sis-

temes en el núvol, en què la pregunta més incisiva i preocupant alhora és qui cobreix aquesta seguretat o qui es responsabilitza de les fallades o forats amb robatori de dades.

Altres punts negres es troben en els processos de gestió documental, en què resulta difícil determinar la relació de dades personals/corporatives amb serveis generals externs a Internet. En el camp de la informació i la documentació sorgeixen una sèrie d'interrogants: com es pot regular l'accés i la protecció de dades?, com els fluxos d'informació poden "saltar" entre aplicacions i ser propietari d'aquests vincles?, de qui és el coneixement generat si el servei deixa de funcionar o volem donar-nos de baixa?, i què succeeix amb l'escassa regulació sobre els drets de les dades o els serveis evolucionats amb API de programari en *cloud*?, etc. Una infinitat de dubtes que poden quedar ocults pels cants de sirena de la informàtica en el núvol.

4.2.5. Entenent el *cloud computing*

Sobre el *cloud computing* podem afirmar que es tracta d'una de les grans evolucions tecnològiques dels últims anys en les TIC, i pot arribar a tenir un impacte semblant al renaixement d'Internet; de fet, Gartner considera aquesta tecnologia com la més important (Grand Hype) en la seva llista de les 10 tecnologies estratègiques (vegeu la figura 6). Per a reforçar aquesta idea, en un informe recent d'IDC (Broderick, 2010) s'assenyala el mercat del *cloud computing* com un dels més actius, prolixos i amb més creixement que pot arribar a superar els 6,4 bilions de dòlars de facturació el 2014.

Per tal de poder albirar aquest nou paradigma i veure el que pot oferir el *cloud computing*, és fonamental establir alguns conceptes i definicions que ens poden servir per a aclarir el –de vegades– confús panorama del "núvol". Moltes vegades es barregen els conceptes de *núvol privat* i *núvol públic* amb els d'*infraestructura com a servei* (IaaS) i *plataforma com a servei* (PaaS). A continuació, desgranarem una sèrie de fonaments sobre què s'entén per cadascun, de manera que ens permeti poder-los delimitar sense barrejar-los ni errar en el seu model, estructura i aplicabilitat. Així, podem dividir el núvol de manera genèrica en tres categories principals en funció dels serveis que s'ofereixen actualment:

- **IaaS (*infrastructure as a service*):** quan parlem d'IaaS ens referim a una infraestructura de computació com a servei, s'obté una solució basada en virtualització i en la qual es paga per consum de recursos com a espai en disc utilitzat, temps de CPU, espai en base de dades, transferència de dades, etc. Per a això un proveïdor de servei posa el seu centre de dades o *data-center* al servei dels clients perquè puguin pujar màquines virtuals o aplicacions i s'executin en aquestes màquines virtuals. Dins d'aquest model podem obtenir una escalabilitat automàtica o semiautomàtica, de manera que puguem contractar més recursos a mesura que els anem necessitant.

- **PaaS (*platform as a service*):** aquest model engloba, a més dels serveis d'infraestructura, els sistemes operatius, i ofereix tot el necessari per a suportar el cicle complet de construcció i posada en funcionament d'aplicacions i serveis web. Es tracta d'un entorn en el qual no solament tenim la possibilitat d'executar accions sobre una plataforma, sinó que a més disposem d'una plataforma en si mateixa on es poden construir nous desenvolupaments. Bàsicament, és posar en un *datacenter* en el núvol màquines virtuals o aplicacions, on a més d'executar-se permet interactuar amb altres serveis tant en el núvol com en mode local, més conegut com a *on-premise*.
- **SaaS (*software as a service*):** és un producte final, una solució completa que un proveïdor posa en el núvol, com un CRM, un servei de correu electrònic, o qualsevol altre servei que requereixi connectivitat contínua, en què l'usuari final simplement accedeix a un programari ja instal·lat en el núvol. Existeix una variant (Velte, 2010) que és el **S+S (*software + service*)**, i que podem considerar un element més de SaaS, ja que es tracta de sumar el Programari com a servei amb un client instal·lat localment, que a més permet treballar fora de línia.

Pel que fa als models d'integració i virtualització de serveis en el núvol, ens trobem amb tres escenaris possibles que formen part d'una capa diferenciada de les esmentades anteriorment (vegeu la figura 6): el model de núvol públic, el model de núvol privat i el model de núvol híbrid.

El **model de núvol públic** es defineix com una solució a la mida de les necessitats tecnològiques i de servei, en què es pot aprofitar tot el potencial del núvol per a suportar grans projectes amb total flexibilitat i la màxima rapidesa a l'hora de disposar i gestionar els diferents recursos, i pagar només per l'ús que se'n faci. En l'àmbit de seguretat i disponibilitat, un dels elements clau del servei, s'ha de tenir la garantia que els elements de xarxa i de maquinari estan redundats i balancejats, a més d'aïllats de manera que només les persones autoritzades puguin accedir a la informació de la seva organització.

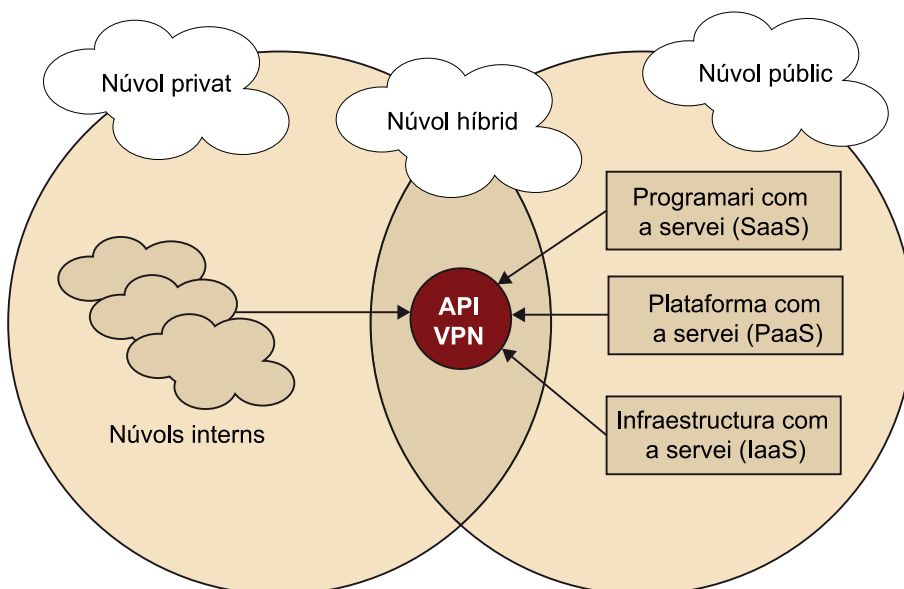
El **model de núvol privat** planteja una plataforma altament escalable en la qual es pot decidir el nivell d'exclusivitat que es requereix, tant en l'àmbit de servidors com de connexions (Rhoton, 2009). És important tenir present els diferents nivells d'aïllament que es puguin necessitar. Es pot optar per un aïllament de xarxa pel qual podem accedir de manera segura a la infraestructura per mitjà d'una xarxa privada virtual (VPN), amb la qual cosa es garanteix que tota la informació que es transfereixi viatgi de manera xifrada. Aquesta possibilitat ens permet establir una connexió de xarxa aïllada amb els nostres servidors amb la mateixa confiabilitat que s'estableix en el model *on-premise*, però aprofitant les possibilitats de connexió del núvol. Al seu torn, es pot recórrer a l'aïllament dels recursos, de manera que tota la infraestructura

de maquinari (servidors, tallafores o *firewall*, balancejadors de càrrega, sistemes d'emmagatzematge físic o lògic) estarà aïllada amb dedicació exclusiva, la qual cosa confereix un gran nivell de seguretat i integritat a la infraestructura.

Des del punt de vista de la seguretat, punt crític en tot sistema TI, cal destacar que tots els elements de xarxa (des del *datacenter* fins a les comunicacions, l'arquitectura de virtualització, etc.) estan redundats per a garantir la màxima disponibilitat del servei. Així, i lligat amb la disponibilitat de servei, o el que es coneix com a *acord de nivell de servei* (SLA en les sigles en anglès), els proveïdors han de fer pública mitjançant contracte la seva política que fa referència a això, i indemnitzar de diferents maneres el client en cas d'interrupció del servei. Un altre element que s'ha de tenir en compte és conèixer per endavant on queden allotjades les dades des del punt de vista de la ubicació física dels *datacenters* del proveïdor, així podrem garantir que el servei proporciona els requisits necessaris per al compliment de la LOPD (Llei orgànica de protecció de dades) en el cas espanyol.

El **model de núvol híbrid**, per la seva banda, permet conjuminar els avantatges del núvol públic i del núvol privat atès que té en compte la integració del *cloud* públic amb altres servidors de la infraestructura que no estiguin en el núvol. Aclarit això, podem afirmar que l'única diferència entre un núvol públic i un núvol privat consisteix en si els recursos estan posats a la disposició de tercers o si, per contra, es troben només disponibles dins d'una organització.

Figura 6. Model d'integració entre "núvols"



Font: elaboració pròpia

4.2.6. Intranets en el núvol

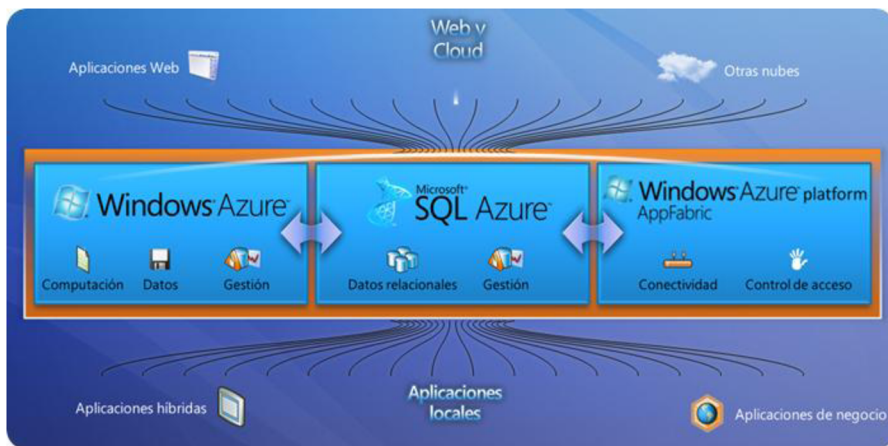
Partint d'una definició estàndard d'*intranet*, entesa com una xarxa d'ordinadors privats que utilitza la tecnologia d'Internet per a compartir sistemes d'informació i sistemes operacionals, podem establir una gran afinitat

amb el concepte del núvol privat tal com l'hem definit anteriorment, en què es produeix una transformació tecnològica que ajuda de manera clara a centrar-se en la funcionalitat i no en el funcionament ni en els esquemes de la infraestructura de maquinari o els de xarxa. El model d'intranet conceptualitzat per a ser desplegat totalment en el núvol ens porta a una clara reducció de costos, però el més significatiu recau en el valor que aporta la transformació en la manera d'utilitzar les TIC en l'organització, l'escalabilitat i la disponibilitat de serveis sota demanda i ús real (Armbrust, 2010).

Per tal de beneficiar-se al màxim d'aquest potencial, l'entorn ideal d'implementació consisteix en l'establiment d'una solució global, entesa com un entorn complet, en la qual es pugui oferir un sistema d'allotjament i connectivitat (IaaS), una plataforma de desenvolupament (PaaS) que permeti centrar-se en les aplicacions i no en la infraestructura ni en la gestió del programari, i un conjunt de solucions (SaaS) personalitzades per a executar-se en els *datacenters*. Windows Azure Platform implica tots aquests elements des d'un entorn propi (allotjat en els *datacenters* de Microsoft) que busca no limitar el concepte tecnològic que el PaaS va néixer per al núvol públic, i així oferir aquests beneficis en entorns tant públics com privats gràcies a la conceptualització d'una plataforma completa.

Aquesta plataforma (vegeu la figura 7) s'estructura en un sistema operatiu en el núvol com a servei per a poder desenvolupar, administrar i executar les aplicacions, i ofereix capacitat de computació, emmagatzematge i connectivitat (Windows Azure), un sistema d'emmagatzematge binari (Windows Azure Storage), un gestor relacional (SQL Azure), una solució per a connectar serveis *cloud* i components interns per mitjà d'aplicacions (plataforma AppFabric), i un *marketplace* o repositori central d'informació per a dades i serveis *cloud* dins de la plataforma.

Figura 7. Escenaris d'aplicació amb Microsoft Windows Azure Platform



Font: MSDN Blog, Luis Panzano <http://bit.ly/h6z7j2>.

Tenint en compte totes les potencialitats que Windows Azure Platform proporciona, és fàcil pensar a abandonar els models d'implementació actuals *on-premise* o en els *datacenters* públics per a migrar al núvol completament amb

la finalitat d'estalviar costos i gestió en TI. Malgrat que hem fet èmfasi en més d'una ocasió en l'efecte de reducció de costos, no hem de pensar únicament en aquest punt de vista, i sí en les implicacions que té tot un procés de migració de serveis, ja que moltes, amb independència de ser crítics o no, requeriran un nou desenvolupament específic. Lògicament, vist el plantejament de Windows Azure Platform, la pregunta que ens podem plantejar és en quin moment podríem utilitzar un model com aquest. Per tal de donar una resposta ràpida i concisa, i sense menysprear altres funcions i aplicacions possibles, a continuació esmentem una sèrie de serveis que poden ser susceptibles de pertànyer i residir en un model de núvol amb Windows Azure Platform (Li, 2009):

- Aplicacions que requereixin una escalabilitat dinàmica, ajustant a cada moment el servei a les necessitats reals de la web, com pot ser l'operativa d'una aplicació social d'alt impacte.
- Aplicacions que requereixin una alta fiabilitat i consistència, tant de serveis, de dades com gestió d'identitats digitals.
- Aplicacions que tinguin una càrrega variable de computació en què les puntes de servei s'estructurin en pics màxims puntuals i factiblement controlables.
- Aplicacions amb un temps de durada curt o no previsible *a priori* que es puguin posar en marxa sense afectar altres processos.
- Aplicacions de processament paral·lel que generalment se solen utilitzar de manera puntual i que possibilitin la connexió amb les seves aplicacions de negoci *on-premise*.
- Aplicacions sobre les quals no es té una gran certesa d'ús o que necessiten ser escalades ràpidament davant una demanda massiva d'ús.
- Aplicacions que es puguin beneficiar de les bondats de l'emmagatzematge extern, o en forma de còpia de seguretat o transaccional no constant.

4.3. RIA (*rich internet applications*)

Hi ha una relació entre tecnologia web i pluralitat tecnològica? Sense posar-nos a pensar, la resposta seria rotundament "sí". Segons la RAE *pluralitat* ens indica "multitud, nombre gran d'algunes coses, o el major nombre d'elles": aquesta accepció es pot assignar directament al concepte *web* que tots coneixem; d'altra banda, un altre significat és la qualitat de ser més d'un, i en aquest sentit el concepte de *pàgina web* en podria quedar exclòs, ja que crear continguts per a diferents navegadors (Internet Explorer, Mozilla, Safari, etc.) o plataformes (Linux, Mac, Windows, etc.) requereix una codificació o recodifica-

ció específica perquè no es produeixin problemes de visualització i renderització, la qual cosa ens condueix a la readaptació dels continguts per a cada tecnologia en la qual es vulgui utilitzar.

Encara que sembli una mica incipient, les aplicacions RIA responen al problema plantejat, és a dir, les tecnologies RIA aplicades als continguts permeten que es visualitzin en qualsevol navegador i sistema operatiu sense modificar res del codi on es troba embolicat. Per tant, un contingut desenvolupat amb tecnologia RIA el visualitzen de la mateixa manera diferents agents d'usuari i amb diferents sistemes operatius.

Una altra de les possibilitats que ens planteja és l'evolució de càrrega de dades. Actualment els continguts en format web estan fent contínuament peticions de connexió als servidors per a proveir nova informació, i aquest trànsit provoca una càrrega excessiva del sistema i pot arribar, algunes vegades, a tornar-lo pràcticament inoperatiu. La tecnologia RIA gestiona de manera independent les càrregues de continguts en una primera connexió per a executar-les i solament genera trànsit nou per a l'accés a dades externes, tant si són fitxers o com bases de dades.

Aquest canvi de paradigma fa que les aplicacions RIA siguin adaptades cada vegada més sovint, però més enllà d'aquest fet (important tecnològicament parlant), la seva veritable potencialitat es troba en la capacitat interactiva i multimèdia.

Intentarem aportar un exemple pràctic. Per editar un fitxer Flash tradicional, primer hem de planificar totes les seqüències, incloent-hi els paràmetres on s'executaran els processos d'interacció al costat del recurs multimèdia que es presentarà, perquè, una vegada establert, es pugui codificar i empaquetar. Qualsevol canvi o error es pot solucionar únicament obrint de nou la pel·lícula Flash, editant aquest canvi i tornant a empaquetar el fitxer. La tecnologia RIA permet que cada interacció, com les realitzades en elements multimèdia, sigui part d'un tot que es pot alterar sense haver d'obrir i empaquetar de nou fent que cada peça sigui alterada sense afectar el conjunt. Al seu torn, permet reproduir contingut multimèdia independentment del reproductor instal·lat en el propi sistema operatiu.

A manera de conclusió, hem d'indicar que la tecnologia RIA s'actualitza independentment i permet fer millores en els sistemes de visualització, càrrega de dades, processos d'interacció amb tecnologies prèviament creades, i garanteix la interoperabilitat entre plataformes i formats de manera que ens permet generar continguts multiplataforma i multiformat, una característica excel·lent que hem de tenir en compte a l'hora de gestionar la difusió de continguts i la seva reutilitat.

4.4. Els programes de programari

Un programa informàtic (programari) és la unió d'una seqüència d'instruccions que un ordinador pot interpretar i executar i una o diverses estructures de dades que emmagatzemen la informació independent de les instruccions que aquesta seqüència d'instruccions maneja.

El programari es classifica en dues categories:

1) **Programari de base o de sistema.** Programari el propòsit del qual és facilitar l'execució d'un altre programari. Entren en aquesta categoria:

- Sistemes operatius
- Compiladors
- Sistemes gestors de bases de dades
- Etc.

2) **Programari d'aplicació.** Programari que automatitza un sistema d'informació, és a dir, amb rellevància per a una fi concreta. Entren en aquesta categoria:

- Processadors de text
- Fulls de càlcul
- Etc.

Les aplicacions són els programes amb els quals l'usuari final interactua, és a dir, són els programes que permeten la interacció entre l'usuari i la computadora. Aquesta comunicació es porta a terme quan l'usuari tria entre les diferents opcions o fa activitats que li ofereix el programa.

Els programes d'aplicació es divideixen en molts tipus, entre els quals:

- Processadors de text (Microsoft Word, Corel WordPerfect, OpenOffice Writer, etc.)
- De disseny (CorelDraw, Photo-Paint, Painter, PhotoShop, Photo Editor, Paint, etc.)
- D'edició (Corel Ventura, QuarkXPress, PageMaker, InDesign, FrameMaker, etc.)

- De càlcul (SPSS, StatGraphics, etc.)
- Compiladors (Visual Basic, C++, Pascal, Visual FoxPro, etc.)
- Full de càlcul (Corel Quattro Pro, Lotus, Microsoft Excel)
- Maneig de base de dades (MySQL, Microsoft Access, Dbase, etc.)
- Comunicació de dades (Safari, Firefox, Internet Explorer, Kazaa, Messengers –Yahoo, MSN, ICQ, AOL–, etc.)
- Multimèdia (Mplayer, reproductor de Windows, Winamp, Real One o Real Player, Quick Time, etc.)
- De presentació (Microsoft PowerPoint, Persuasion, OpenOffice Impress, Corel Presentations, etc.)

Tots i cadascun d'aquests programes tenen una funció específica: estalviar temps i diners a l'usuari en permetre-li fer coses útils amb la computadora, alguns amb millors prestacions, altres amb millor disseny, més amigables o fàcils d'usar.

4.5. El programari lliure

Els conceptes essencials que apareixen relacionats en aquest camp de coneixement són *sistemes oberts*, *programari lliure*, *programari gratuït*, *programari de font oberta*, *freeware*, *shareware*, *formats oberts de documents* i altres termes.

Sistema obert es defineix com els sistemes i components que poden ser especificats i adquirits de fonts diferents en un mercat competitiu. Les especificacions dels sistemes oberts les han de controlar organitzacions internacionals de normalització o, almenys, un especificador tan independent com sigui possible amb un ampli grau d'acceptació en el mercat (normes d'ipso). És a dir, es refereix a sistemes normalitzats en l'àmbit nacional o internacional que no suposin una dependència de marca i que constitueixen l'origen de la filosofia del programari lliure.

Programari lliure és el programari, producte o desenvolupament a mida que es distribueix mitjançant una llicència segons la qual l'autor cedeix una sèrie de llibertats bàsiques a l'usuari en el marc d'un acord de concessió. Aquestes llibertats dels usuaris del programari, recollides en la filosofia de la Fundació per al Programari Lliure (Free Software Foundation), són: la llibertat d'usar el programa amb qualsevol propòsit; la llibertat d'estudiar com funciona el programa i adaptar-lo a les seves necessitats; la llibertat de distribuir còpies, i la llibertat de millorar el programa i fer públiques les millores als altres, de manera que tota la comunitat se'n beneficiï. Per tant, el programari lliure és un programa o seqüència d'instruccions usat per un dispositiu de processament

digital de dades per a dur a terme una tasca específica o resoldre un problema determinat, el propietari del qual renuncia a la possibilitat d'obtenir utilitats per les llicències, patents o qualsevol forma que adopti el seu dret de propietat. Una altra característica és que el codi font del programari està disponible, per la qual cosa es pot modificar el programari sense cap límit, i sense pagar a qui el va inventar o el va llançar al mercat.

Programari propietari (també tancat o no lliure) és el programari que és impossible d'utilitzar en un altre maquinari o terminal, modificar o transferir sense pagar drets a l'inventor o creador.

Programari gratuït (*free software* o *freeware*), concepte molt diferent de *programari lliure*, encara que no totalment oposat. Es tracta del programari que ha estat cedit per part dels autors sense cap cost. Tradicionalment, atès que el terme original en anglès *free software* es pot traduir en català com a *programari lliure* i també com a *programari gratuït*, s'han considerat semblants, encara que denoten filosofies i formes de treball diferents. Igualment diferent és el concepte de *programari compartit* (*shareware*), que consisteix en la possibilitat de baixar el programari i utilitzar-lo durant un cert temps, però no es pot accedir a la font i, generalment, no es pot usar de manera continuada sense pagar una certa quantitat.

El 1997 sorgeix la iniciativa OSI (Open Source Initiative / Iniciativa de Fonts Obertes) i s'introdueix el concepte de **programa de font oberta** (*open source*) que, a diferència del programari lliure, no es basa en una llicència pròpiament dita, sinó en un conjunt de requisits perquè una llicència sigui considerada de font oberta. D'altra banda, la seva definició és pràcticament la mateixa que la de programari lliure.

Hi ha una gran quantitat de programari, cada vegada més, disponible amb llicències de programari lliure. Els observadors (i adeptes) sovint interpreten aquest fenomen com el moviment del programari lliure. Alguns projectes notables de programari lliure inclouen els kernels dels sistemes operatius GNU-Linux i BSD, els compiladors GCC, el depurador GDB i les biblioteques de C, el servidor de noms BIND, el servidor de transport de correu Sendmail, el servidor web Apache, els sistemes de bases de dades relacionals MySQL i PostgreSQL, els llenguatges de programació Perl, Python, Tcl i PHP, el sistema X Window, els entorns d'escriptori GNOME i KDE, el paquet ofimàtic OpenOffice.org, el navegador Mozilla, el servidor de fitxers Samba i l'editor de gràfics GIMP.

4.6. Els formats de documents

En l'àmbit de producció de continguts i el treball en el context d'una organització s'han de seleccionar un conjunt comú d'estàndards de format de fitxer: gràfic, text, dades, àudio i vídeo que facilitin l'accés i la circulació de la infor-

mació, i la seva posterior recuperació i conservació. A continuació es presenten els formats més apropiats i de més durabilitat i que es consideren els adequats per a la seva utilització en projectes web.

Els formats de text:

- **TXT:** format simple que en permet la lectura a qualsevol.
- **PDF:** permet visualitzar documents reproduint totes les característiques de l'original en fitxers de menys grandària, independents de l'aplicació i de les plataformes, l'especificació és pública i també està estès per a la distribució, la difusió formal, l'accés i la visualització de documents.
- **RTF:** format que constitueix un mínim comú entre processadors de text diferents.
- **SGML:** norma internacional ISO 8879, del món editorial, que emmagatzema el text i la seva estructura, però no té atributs de presentació; actualment s'està reemplaçant per l'XML i l'HTML.
- **XML:** dialecte de l'SGML adequat per a definir documents independents de la plataforma i processar-los de manera automàtica, ja que distingeix entre estructura, contingut i presentació, i ofereix més possibilitats que l'HTML.
- **HTML:** versió simplificada de l'SGML que s'utilitza en els servidors web, molt útil per a la difusió d'informació.
- **SXW:** format dels documents de text manejats pel programari lliure openoffice.org.
- **Encapsulated PostScript:** utilitzat per a enviar i imprimir documents juntament amb la seva presentació, de manera que s'asseguri que la sortida impresa és correcta amb independència del dispositiu utilitzat.
- **Especificació CSV:** per a l'intercanvi de taules, delimitades per comes.

Els formats de dades estructurades:

- **XML:** adaptació de l'SGML adequat per a definir documents independents de la plataforma i processar-los de manera automàtica, ja que distingeix entre estructura, contingut i presentació, i ofereix més possibilitats que l'HTML.
- **Bases de dades:** bases de dades relacionals conformes amb les normes internacionals sobre SQL, ANSI X3.135-1992/ISO 9075:1992.

- **MIME:** per a missatges de correu electrònic i intercanvi electrònic de dades i fitxers adjunts.

Els formats comprimits:

Especificació ZIP 2.0 per a l'intercanvi de dades comprimides.

4.7. Els documents multimèdia

Multimèdia és la qualitat d'un sistema o document que utilitza més d'un mitjà de comunicació al mateix temps en la presentació de la informació com a text, imatge, animació, vídeo i so.

Encara que aquest concepte és tan antic com la comunicació humana, ja que en expressar-nos en una xerrada normal parlem (so), escrivim (text), observem el nostre interlocutor (vídeo) i accionem amb gestos i moviments de les mans (animació), amb prou feines ara, amb l'auge de les aplicacions multimèdia per a ordinador, aquest vocable ha entrat a formar part del llenguatge habitual.

Quan un programa, un document o una presentació combina adequadament els mitjans, es millora notablement l'atenció, la compressió i l'aprenentatge, ja que s'aproparà una mica més a la forma habitual que els éssers humans ens comuniquem, quan emprem diversos sentits per a comprendre un mateix objecte o concepte.

La utilització de tècniques multimèdia va permetre el desenvolupament de l'hipertext, una manera d'enllaçar temes amb paraules en els textos, de manera que es pugui accedir a temes d'interès específic en un o diversos documents sense haver-los de llegir completament, simplement fent clic amb el ratolí en les paraules remarcades (subratllades o d'un color diferent) que estiguin relacionades amb el que se cerca. El programa porta immediatament a la pantalla altres documents que contenen el text relacionat amb aquesta paraula. Fins i tot, es poden posar marques de posició (*bookmarks*). Així es controla l'ordre de lectura i l'aparició de les dades en la pantalla, d'una manera més semblant a la nostra forma de relacionar pensaments, en què el cervell va responent per lliure associació d'idees, i no seguint un fil únic i lineal.

Però la vinculació interactiva no es va limitar a textos solament. També es pot interactuar amb sons, animacions i serveis d'Internet relacionats amb el tema que s'està tractant, la qual cosa ha donat origen a un nou concepte, *hipermèdia*, resultat de la fusió dels conceptes *hipertext* i *multimèdia*. Podem entendre sistemes d'hipermedis com a organització d'informació textual, gràfica i sonora per mitjà de vincles que creen associacions entre informació relacionada dins del sistema.

1) Tipus d'informació multimèdia

- **Text:** sense formatar, formatat, lineal i hipertext.
- **Gràfics:** utilitzats per a representar esquemes, plànols, dibuixos lineals, etc.
- **Imatges:** són documents formats per píxels. Es poden generar per còpia de l'entorn (escanejat, fotografia digital) i tendeixen a ser fitxers molt voluminosos.
- **Vídeo:** presentació d'un nombre d'imatges per segon, que creen en l'observador la sensació de moviment. Poden ser sintetitzades o captades.
- **So:** pot ser parla, música o altres sons.

Un **text** és una composició de signes codificats en un sistema d'escriptura (com un alfabet) que forma una unitat de sentit. La seva grandària pot ser variable, des d'una obra literària com *El Quixot* a una consulta en SQL.

Una **imatge** és una representació visual d'un objecte mitjançant tècniques diferents de disseny, pintura, fotografia, vídeo o altres. Les imatges en l'ordinador, bàsicament, es poden emmagatzemar i processar de dues maneres: per mapa de bits o mitjançant vectorització.

2) Tipus de formats d'imatges

- **BMP:** utilitzat per Microsoft com a format estàndard d'imatge, és en el seu concepte un mapa de bits.
- **PCX:** és un format utilitzat en programaris com Paintbrush, és un format molt utilitzat per a imatges de 256 colors.
- **TIFF:** aquest format es va idear com a format d'intercanvi, perquè el pogués llegir qualsevol sistema, l'utilitza PageMaker.
- **GIF:** aquest format va ser creat per Compuserver perquè fos visualitzat independent de l'ordinador, és el format més utilitzat a Internet a 256 colors.
- **TARGA (TGA):** és un format creat per Truevision d'alta qualitat. Des del primer moment treballa a 24 bits.
- **JPEG:** aquest format el va desenvolupar la Joint Photographic Experts Group per als treballs de fotografia digital. Aquest format depèn, fonamentalment, del nivell de codificació i el nivell de qualitat. L'important del sistema és que l'usuari pot actuar sobre aquests paràmetres definint així la qualitat que desitja a cada moment.

- **PSD** d'Adobe Photoshop: aquest format és el que generalment s'exigeix en la majoria d'impremtes, que solen treballar amb aquest programari.
- **SWF**: és un contenidor de Flash, i com a tal pot tenir imatges jpg, gif, png, etc., vídeos avi, mov, etc. o imatges vectorials pròpies d'aquest programari que tenen la capacitat de moviment.
- **PCD**: photo CD, clàssic dels CD d'imatges de Kodak.

El **vídeo** és un enregistrament realitzat amb un gravador de vídeo o algun altre dispositiu que capturi imatges en moviment.

3) Tipus de formats de vídeo

- **AVI i AVI 2.0**. El format AVI (Audio Video Interleave) té un funcionament molt simple, ja que emmagatzema la informació per capes, guarda una capa de vídeo seguida per una d'àudio. Els seus còdecs estan desenvolupats com a controladors per a ACM (Audio Compression Manager) i VCM (Video Compression Manager), i també poden ser usats per algunes altres arquitectures, incloses DirectShow i Windows Media.
- **Microsoft Windows Media Video** és una de les últimes propostes de Microsoft que funciona amb el Windows Media Player de la versió 6.2 o superiors. Ha tingut un gran impuls amb Windows XP, ja que està integrat en aquest sistema operatiu. També té una opció per a reproducció en temps real o *streaming* que està inclosa en el Windows 2000 Server. Les extensions d'aquest tipus de continguts són les .asf i .wmv per al vídeo i .wma per a l'àudio.
- **Real Video** ha estat molt utilitzat per a *streaming* d'àudio en diversos mitjans. També té una proposta per a vídeo anomenada Real Video. Requereix el seu propi reproductor o *player*, que és el Real Player (posteriorment s'ha llançat Real One) i per a fer reproducció en temps real o *streaming* requereix el Real Server. En el lloc de Real també hi ha informació per a convertir arxius .avi a aquest format. Real sempre té una versió simple i limitada dels seus productes i una versió professional que es pot comprar.
- **Apple Quicktime**. Apple té una interessant opció nativa dels sistemes Mac. Els seus arxius .mov requereixen un reproductor o *player* especial que és el Quicktime Player per a visualitzar-los. Aquest reproductor té una versió senzilla gratuïta i una versió professional que, entre d'altres, permet realitzar vídeos en aquest format i editar-ne algunes qualitats.

El **so** és la percepció de l'oïda. El so digital és la captació, la transmissió i la reproducció del so per mitjans informàtics.

4) Tipus de format d'àudio digital

- **WAV:** els arxius de forma d'ona (o simplement *wave*) són els formats de so més comuns en les plataformes Windows. Els arxius WAV també es poden reproduir en Mac i en altres sistemes amb programari reproductor.
- **MPEG (MP3):** el format Motion Pictures Experts Group (MPEG) és un format estàndard amb una capacitat de compressió important. Els arxius MPEG de nivell 3 o MP3 són utilitzats freqüentment per a distribució de música pel web.
- **RealAudio (.rm):** Real Audio és la tecnologia que predomina actualment en el web. Necessita un reproductor propietari, però les versions bàsiques del reproductor estan disponibles gratuïtament.
- **MIDI:** el format Musical Instrument Digital Interface no és un format d'àudio digital. Representa notes i altra informació perquè es pugui sintetitzar la música.
- **MOD:** es considera una espècie de barreja entre el format MIDI i el format WAV, ja que, d'una banda, emmagatzema el so en forma d'instruccions per a la targeta de so, però, de l'altra, pot emmagatzemar també sons d'instruments musicals digitalitzats, i poden ser interpretats per qualsevol targeta de so de 8 bits. No és un format estàndard de Windows, per la qual cosa el seu ús és més indicat per a sistemes Mac, Amiga o Linux.
- **Real Audio:** és un format de qualitat mitjana, encara que permet fitxers molt comprimits, que guarda amb extensió .rmp o .ra. Per a reproduir-los cal tenir instal·lat el connector o *plugin* Real Audio.

4.8. El format PDF

El PDF (format de document portàtil, *portable document format*) és un llenguatge de descripció de pàgina. L'HTML descriu el contingut del document de manera abstracta, en principi sense especificar-ne l'aparença. Un llenguatge de descripció de pàgina està orientat a especificar com s'ha d'imprimir o mostrar en pantalla com a pàgina impresa. Un dels avantatges del PDF és que permet una reproducció fidel del format original. No obstant això, aquest avantatge fa el document inaccessible per a molts usuaris.

Ideat per Adobe, el format de documents portàtils és l'eina pública utilitzada en empreses amb estàndards mundials per a una distribució i un intercanvi segurs i fiables de documents electrònics. Governos, empreses i formadors de tot el món han adoptat el format PDF d'Adobe® per a agilitar l'intercanvi de documents, augmentar la productivitat i reduir la dependència del paper.

Algunes prestacions del format PDF són:

- **Manteniment de l'aspecte i la integritat dels documents originals:** els arxius PDF d'Adobe tenen el mateix aspecte que els documents originals, al mateix temps que mantenen les fonts, les imatges, els gràfics i el disseny de qualsevol arxiu original, independentment de l'aplicació o plataforma emprades per a crear-lo.
- **Possibilitat de compartir documents amb qualsevol persona:** tothom pot compartir, veure i imprimir documents PDF d'Adobe en qualsevol sistema mitjançant el programari gratuït Adobe Reader®, independentment del sistema operatiu, l'aplicació original o les fonts.
- **Fàcil d'usar:** els arxius PDF d'Adobe són compactes i fàcils d'intercanviar. La creació d'aquests arxius és tan simple com fer clic en un botó des d'aplicacions com Microsoft Word, Excel i PowerPoint.
- **Intercanvi de documents més segur:** amb el programari Adobe Acrobat 7.0 es poden protegir els documents PDF amb contrasenya i així es pot evitar la visualització i la modificació no autoritzades.

4.9. Els formats ODF i OpenXML

El maig de 2006 s'aprova la norma ISO/IEC DIS 26300, referida a l'estàndard obert OpenDocument Format, ODF (format obert de documents) proposat per OASIS. Aquest estàndard obert és un format de fitxer o conjunt d'especificacions basat en XML per a documents en suport electrònic: textos, fulls de càlcul, presentacions i, fins i tot, gràfics.

Microsoft no es queda enrere i també vol convertir en norma ISO el seu nou format OpenXML, per així poder competir amb OpenDocument.

Actualment, ODF té suport dels paquets OpenOffice.org 2.0 i StarOffice 8. ODF també s'incorporarà al client Workplace d'IBM. Mentrestant, OpenXML serà el format estàndard de la propera versió del paquet ofimàtic de Microsoft, Office 2007. Com era previsible, Microsoft continuarà sense oferir suport per a OpenDocument en les aplicacions ofimàtiques: OpenOffice és gratuït i és capaç d'obrir més varietat de formats de documents que Microsoft Office.

OpenDocument ha estat dissenyat perquè reutilitzi els estàndards XML oberts existents quan aquests estiguin disponibles. Així, OpenDocument utilitza Dublin Core d'XML per a les metadades, MathML per a les fórmules matemàtiques, SVG per als gràfics vectorials, SMIL per a la multimèdia, etc.

5. Visualització d'informació: Casos

Abans d'endinsar-nos a estudiar formes i models de visualització, hem de saber que hi ha vuit coses que s'han de tenir en compte en produir visualitzacions de dades (adaptat de Data Visualization Techniques per SEOMozBlog.)

La visualització de dades tracta de presentar la informació en algun tipus de forma gràfica. L'ull humà té una capacitat de detecció d'elements molt potent, i així, en transformar una taula en un gràfic, per exemple, no hi afegim informació nova, però resulta més fàcil interpretar aquestes dades o veure tendències a partir d'elles. Les visualitzacions es poden utilitzar com a valuoses eines d'anàlisi que permeten identificar patrons en les dades que, d'una altra manera, serien molt difícils de trobar o veure.

A continuació, presentem alguns consells sobre com es pot obtenir el màxim rendiment en la visualització de dades.

1) Recopilar dades: és important utilitzar eines de desenvolupament complet per a generar remescles o *mashups* per mitjà de diferents fonts de dades i obtenir així una visió única.

2) Utilitzar un gran disseny: avui dia disposem de múltiples aplicacions que tenen dissenys predefinitos i podem utilitzar per als nostres models de representació, i alguns fins i tot permeten un cert grau de personalització. El que sí que hem de tenir clar és que si volem generar un model gràfic basat en dades, hem de triar un format que sigui com més senzill i funcional millor, sempre que les dades ens ho permetin.

3) Interactivitat: quan es vulgui treballar amb dades en línia, tenim l'oportunitat de proporcionar als usuaris la interactivitat. De vegades, infografies estàtiques són prou bones per a mostrar informació, però generalment aconseguirem més interès, ús i comprensió si proporcionem un sistema de control en què els usuaris puguin jugar sobre la base de les seves necessitats i interessos.

4) El recurs del record: les dades en cru són pràcticament impossibles de recordar, i el record sobre un model de dades és un eix clau per a analitzar i comprendre el modelatge de les dades com a font. Per a això, una de les millors maneres d'exposar-les a una comprensió directa és la comparació, ja que les comparacions interessants i imatges fortes milloren la compartibilitat de qualsevol model de dades. Si recordem un model per comparació esclarem millor el missatge que si únicament mostrem els simples fets.

5) **Conèixer els objectius:** hem de saber en primer lloc quins són els nostres objectius, i després trobar la manera d'arribar-hi. Cal assegurar-se que el model de visualització és vist pel mercat objectiu, o companys de treball, clients o alts directius, i fer un esforç per a orientar els nínxols en aquest mercat.

6) **Permetre que es pugui compartir:** és fonamental proporcionar el codi d'inserció per a compartir el model (amb l'enllaç). Si les dades que es mostren ofereixen un tractament específic a dades vinculades i disgregades moltes vegades, és probable que augmenti la probabilitat que les persones incrustin el seu contingut o el referencin, de manera que arribarà a més usuaris.

7) **Evitar l'excés d'informació:** hi ha una tendència a mostrar-ho absolutament tot, totes les dades, tota la gamma, cada columna i fila, però el millor model és el de les visualitzacions que reflecteixen claredat i on s'exclou tot allò que resulta innecessari o aporta poc valor.

8) **Autosuficiència:** tots els gràfics, quadres i infografies han de ser autosuficients, és a dir, tota la informació continguda, la interpretació de les dades i els models utilitzats han de ser compresos de manera directa i intuïtiva. Per a això, és fonamental definir un títol clar i concís, una llegenda, l'origen de les dades, etiquetes clares, ús de colors concordes a la necessitat de cada model, etc.

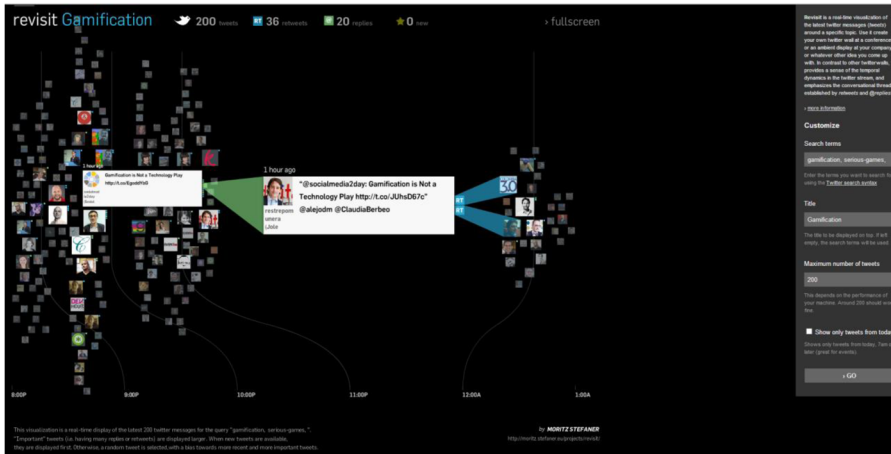
Els models de visualització d'informació més comuns són els gràfics i els mapes. És evident que un gràfic és molt més poderós visualment que un conjunt de dades tabulars, ja que d'un sol cop d'ull podem aprendre el valor de les dades que s'hi inclouen. Poder interpretar les dades clarament és el gran valor que podem aplicar al tractament de les dades de manera gràfica, i permetre que es puguin interpretar ràpidament és una de les tasques més complexes que se'ns plantegen.

A continuació presentem una sèrie de casos i d'exemples de models de visualització de dades de diversos tipus que permetran extreure conclusions sobre la no-existència d'un únic punt de vista sobre la modelització visual de les dades.

5.1. Visualització d'informació

1) **Newsmap** (<http://newsmap.jp/>).

Figura 9



a) Descripció

Aplicació que permet la visualització en temps real de missatges de Twitter sobre una temàtica específica ajustant-los a la dinàmica temporal en el flux de Twitter, posa l'accent en els temes de conversa establerts a través dels *retweets* (repiulades) i les respostes a usuaris (*@replies*).

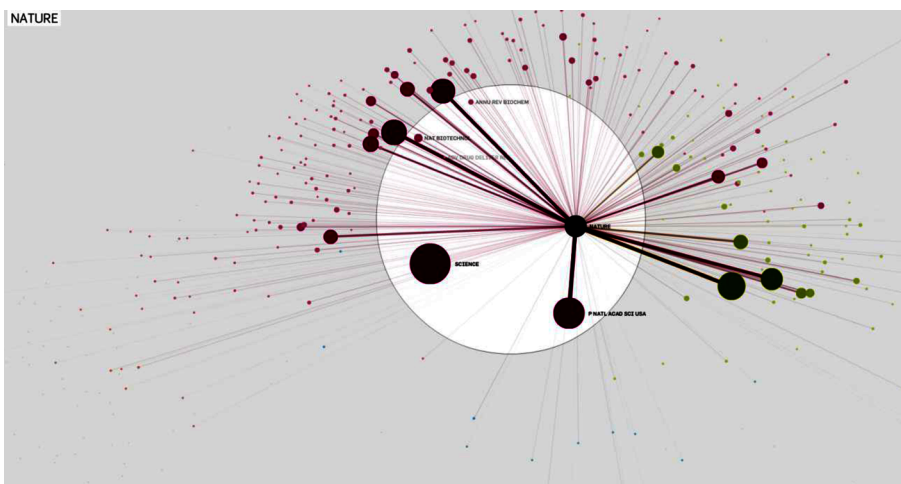
Quan arriben noves piulades o *tweets*, es mostren en la part central de la pantalla i destacades. Els *tweets* que estableixen connexions amb d'altres perquè s'envien o es contesten, s'indiquen amb connexions de colors de manera que es pot seguir el fil d'una conversa i el seu context.

b) Data set

Missatges globals de l'aplicació Twitter (només permet consultes sobre els darrers vuit dies). Utilitza l'API propi de Twitter per a generar terceres aplicacions híbrides.

3) Well-formed Eigenfactor (<http://well-formed.eigenfactor.org/map.html>).

Figura 10



a) Descripció

Eina per a la visualització del flux d'informació sobre ciència basada en un model de mapa relacional. Aquesta visualització situa revistes científiques que són citades amb freqüència pels autors, estableix índexs de freqüència i proximitat entre autors, citacions i títols de revistes científiques que permeten avaluar la importància de cada revista.

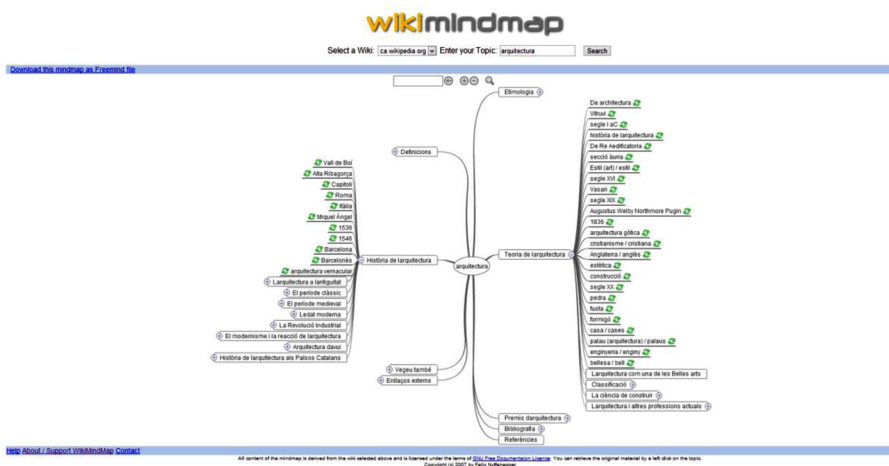
Arrossegant la lent d'augment blanca permet ampliar la vista de nodes i ofereix una visualització més propera. En fer clic en un dels nodes posarà en relleu totes les seves connexions, de manera que les grandàries dels nodes relacionats representen la quantitat relativa de flux de citació (d'entrada i sortida) pel que fa a la revista seleccionada. Si no se selecciona cap títol, la vista per defecte s'escala sobre la base de l'Eigenfactor® Score que té cada revista.

b) Data set

S'utilitza un subconjunt de les dades de citacions de citacions de *Thomson Reuters' Journal Citation Reports* 1997-2005. Les dades agregades completes, en l'àmbit de revista, comprenen aproximadament 60.000.000 de citacions de més de 7.000 revistes en l'última dècada. Per a generar un subconjunt interessant, se seleccionen revistes ordenades pel seu factor Article Influence® el 2005, però sense que contingui més de vint-i-cinc revistes d'un sol camp o temàtica. Per a fer que el subconjunt sigui coherent, s'assegura que de les revistes seleccionades s'inclouen tots els anys i que cobreixen les deu revistes amb més puntuació Eigenfactor® Score. Per a agrupar les xarxes, s'utilitza el mètode de la teoria de la informació presentada en els mapes de flux d'informació per a revelar l'estructura d'una comunitat en xarxes complexes [PNAS 105, 1118 (2008)].

4) Wikimindmap (<http://www.wikimindmap.org/>).

Figura 11



a) Descripció

Eina per a navegar fàcilment i de manera eficient pel contingut de la Wikipedia que s'inspira en la tècnica dels mapes mentals. Les pàgines de la Wikipedia són complexes internament per la gran quantitat d'enllaços i relacions que es generen, però la visualització lineal per mitjà de la pàgina web no permet englobar de manera ràpida i directa les relacions entre conceptes d'una mateixa categoria o temàtica.

Una manera eficient d'entendre les estructures entre conceptes i permetre una cerca d'informació relacionada és per mitjà dels mapes mentals, i és en aquest punt en el qual aquesta eina pretén ajudar els usuaris a obtenir una bona visió general comprensible, estructurada i fàcil del tema o concepte que es busca.

b) Data set

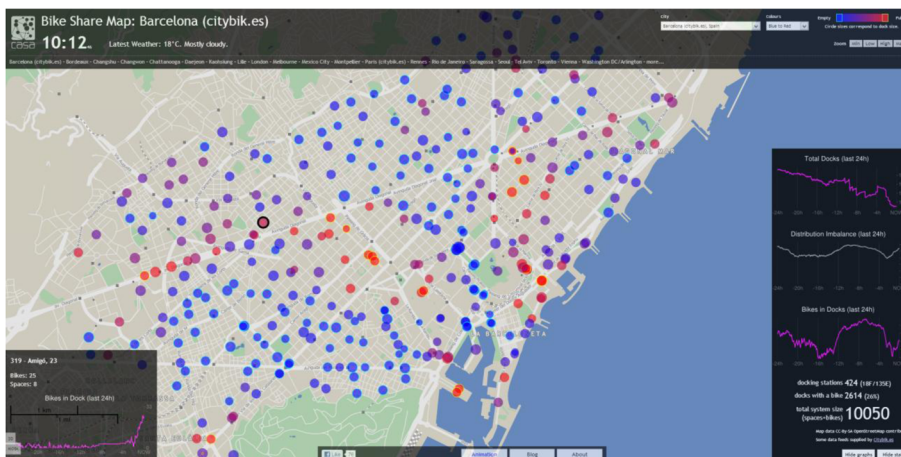
Entrades de la Wikipedia en l'àmbit global en temps real sobre els seus més de vint milions d'entrades. Permet la segmentació per traduccions d'idiomes com l'alemany, francès, català, espanyol, italià, danès, hebreu, hindú, polonès, holandès, portuguès i suec.

Utilitza l'API de Wikipedia per a la generació dels mapes mentals de conceptes.

5) Bike Share Map-Barcelona

(<http://bikes.oobrien.com/barcelona/>).

Figura 12



a) Descripció

Aplicació que ens permet representar de manera geolocalitzada en un mapa i en temps real tots els punts de lloguer de bicicletes del sistema públic posat en marxa a diverses ciutats del món. Podem visualitzar informació sobre la dis-

tribució d'aquests punts, el nivell d'ús de les bicicletes en un moment concret, la progressió en el temps de l'ús de cada punt de recollida i la disponibilitat o no de bicicletes en cada punt.

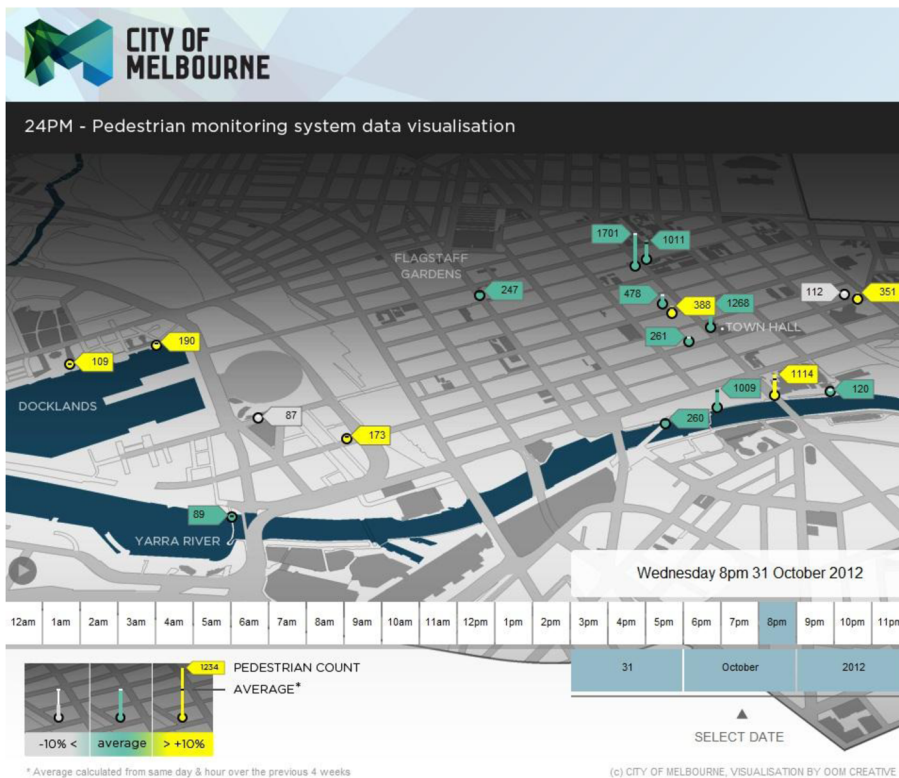
b) Data set

El sistema es nodreix dels registres recollits pels terminals de cada punt d'accés a les bicicletes mitjançant la transacció que fa cada usuari amb la seva targeta personalitzada. Aquestes dades emmagatzemades en els servidors centrals poden mostrar dades sobre la traçabilitat de cada bicicleta, des del seu lloc de partida en una estació fins que és retornada en un altre punt diferent de l'original.

6) 24PM Pedestrian Monitoring System

(<http://www.pedestrian.melbourne.vic.gov.au>).

Figura 13



a) Descripció

Aplicació que genera un sistema de vigilància de vianants les vint-i-quatre hores del dia a la ciutat de Melbourne. Mesura l'activitat dels vianants en el centre de la ciutat, districtes i Docklands cada dia. Aquest sistema permet comprendre millor com la gent es mou pels diferents punts de la ciutat en diferents moments del dia, així es pot gestionar com funcionen i planificar les necessitats futures sobre la base de l'afluència de vianants.

Els objectius del projecte són informar sobre l'ús urbà de diferents punts de la ciutat per a poder planificar i gestionar actuacions sobre aquests punts, identificar oportunitats per a millorar el trànsit i el transport, mesurar l'impacte dels esdeveniments i les campanyes de màrqueting específiques sobre l'activitat per als vianants, supervisar l'activitat comercial a la ciutat i ajudar a la presa de decisions per a la comunitat empresarial sobre les estratègies de màrqueting.

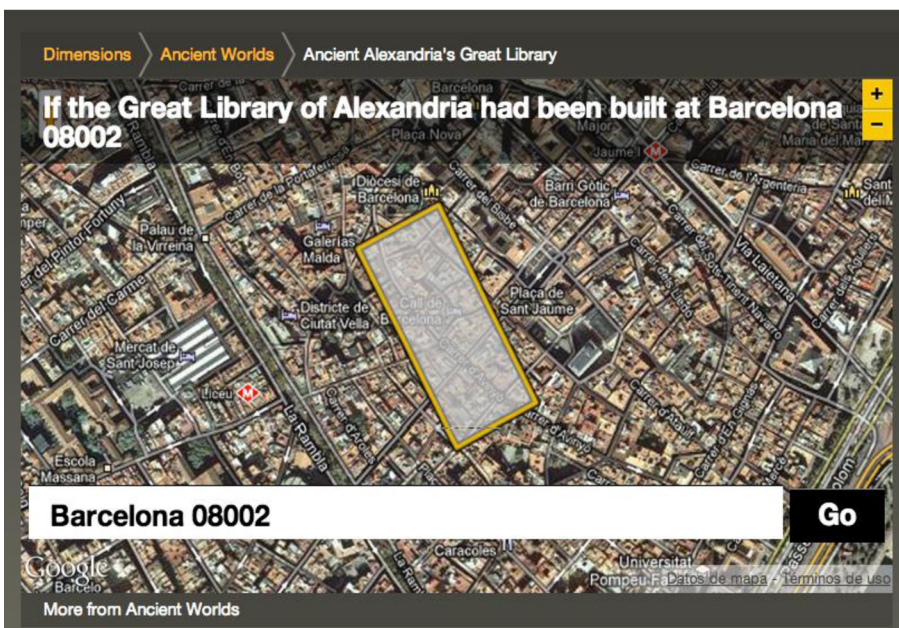
b) Data set

Les dades es prenen en temps real des dels divuit sensors de comptatge per als vianants de què consta el sistema. Les dades recollides amb dispositius sense fils s'envien a una base de dades centralitzada en un servidor on s'emmagatzemen i es processen i es transmeten al portal web per a visualitzar-les.

7) BBC Dimensions - How big really?

(http://howbigreally.com/dimension/ancient_worlds/).

Figura 14



a) Descripció

Aquesta eina busca fer patent la dimensió de diferents elements de manera comparada amb la nostra realitat actual presentant-los en forma de mapa digital. La possibilitat de poder comparar radis d'acció d'un fenomen esdevingut en un lloc concret i extrapolar-lo a una nova ubicació actual, o situar, com el cas que es presenta en el gràfic, una edificació antiga en una població moderna, ens ajuda a comprendre el missatge que ens transmet de manera directa i visual, la dimensió de les coses.

Si ens diguessin com era de gran la biblioteca d'Alexandria, quants metres quadrats ocupava i quants documents va arribar a allotjar, per mitjà de dades pures numèriques, és molt probable que no aconseguiríem fer-nos una idea de l'impacte real. Si situem la seva planta en un districte actual de la ciutat de Barcelona podem veure realment quina extensió ocupava.

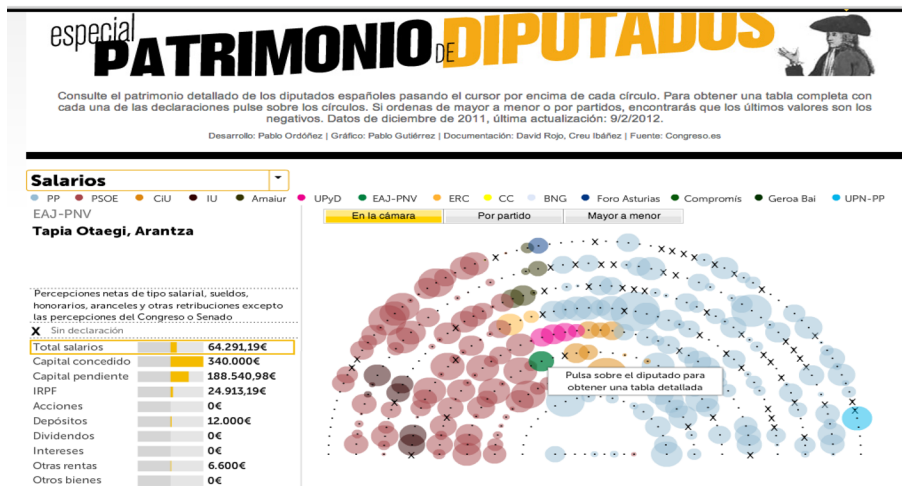
b) Data set

Dades quantitatives obtingudes de diferents fonts de mesurament planimètric. La representació visual ens permet geolocalitzar diferents elements en una àrea o zona geogràfica per a poder aprehendre la seva dimensió.

8) Patrimoni dels Diputats del Congrés d'Espanya

(<http://especiales.lainformacion.com/espana/patrimonio-diputados/>).

Figura 15



a) Descripció

Aquesta eina permet consultar el patrimoni detallat de les declaracions de béns i renda i d'activitats dels diputats de la IX legislatura. A partir de la representació gràfica dels diferents membres situats en l'hemicicle, es poden consultar les xifres de salari, capital, inversions i altres rendes que formen el patrimoni individual. La manera de representar aquestes dades, que fins ara solament es mostraven en interminables llistes, facilita la comprensió de les dades de manera molt directa i intuïtiva.

b) Data set

Dades aportades per Congreso.es sobre les valoracions dels béns declarats pels diputats des que el 19 de juliol de 2011 es va aprovar la reforma de l'article 160.2 de la Llei orgànica del règim electoral general perquè se'n coneguessin al detall els béns.

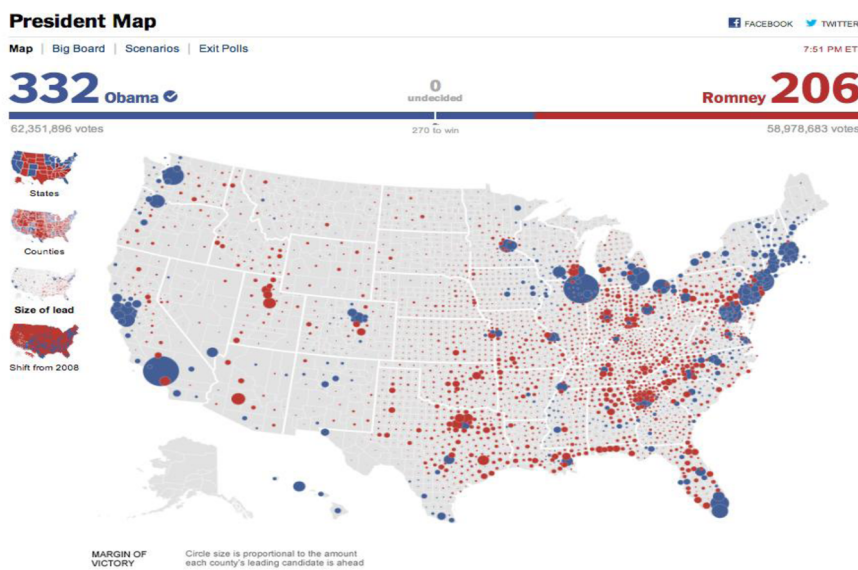
5.2. Exemples de models de visualització

A continuació presentem deu models de representació visual de la informació diferents que ens permetran analitzar la viabilitat de cada model sobre la base de la font de dades representada, de manera que es podrà determinar com és d'intuïtiu cada model i si la manera de visualitzar la informació és la més adequada a la finalitat concreta.

1) President Map - Elections 2012

(<http://elections.nytimes.com/2012/results/president>)

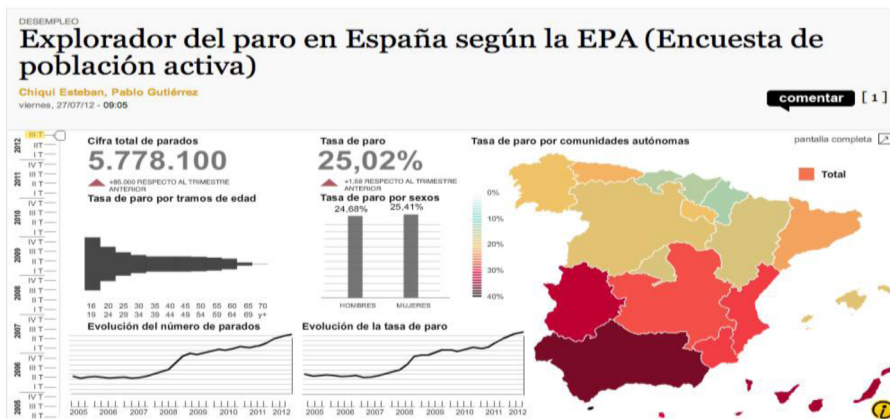
Figura 16



2) Explorador dels índexs d'atur a Espanya segons l'Enquesta de població activa

(<http://goo.gl/ryxgy>)

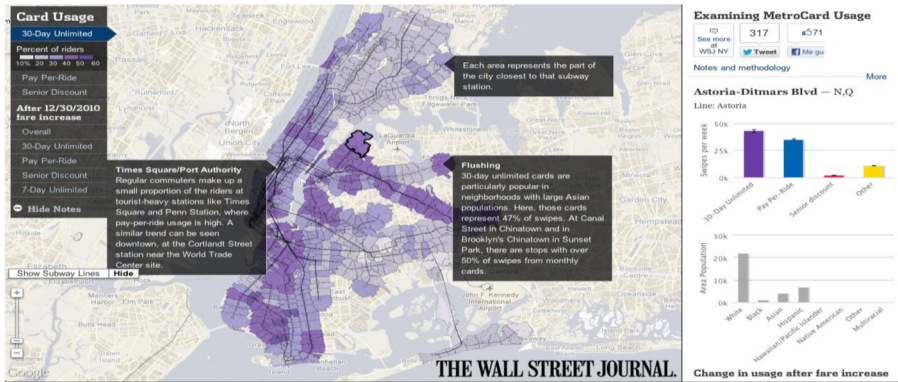
Figura 17



3) Examining MetroCard usage - New York

(<http://graphicsweb.wsj.com/documents/mtafares1108>).

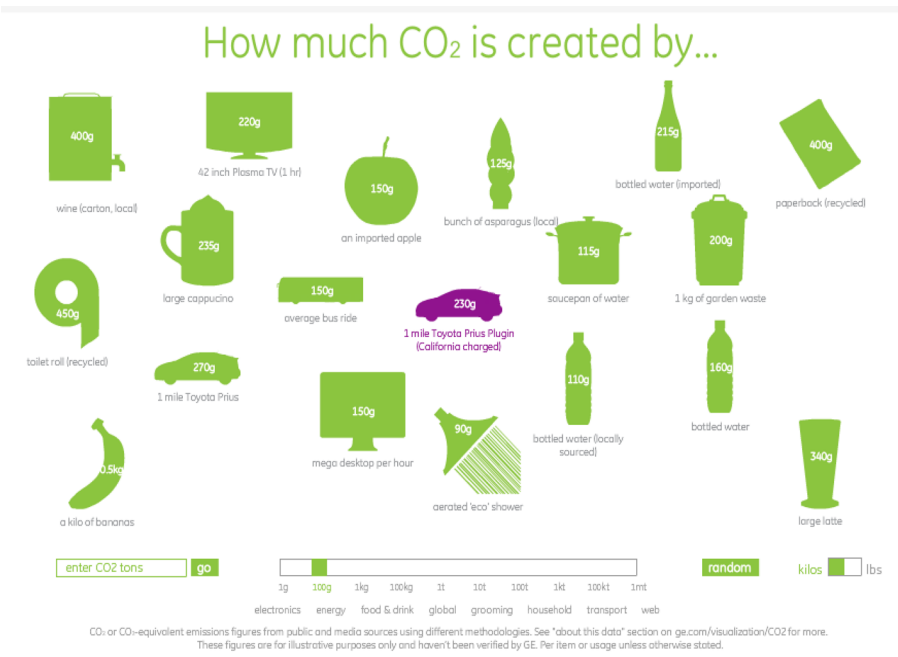
Figura 18



4) How much CO2 is created by...

(<http://visualization.geblogs.com/visualization/co2/>).

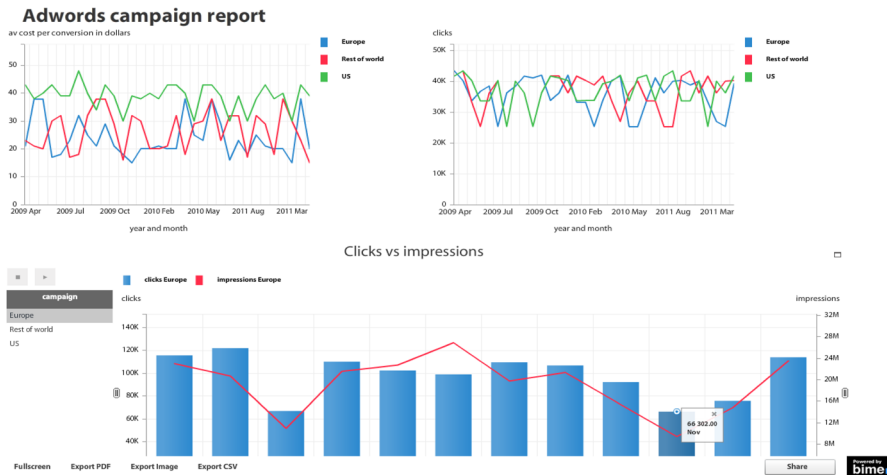
Figura 19



5) Colours in Culture

(<http://www.informationisbeautiful.net/visualizations/colours-in-cultures/>).

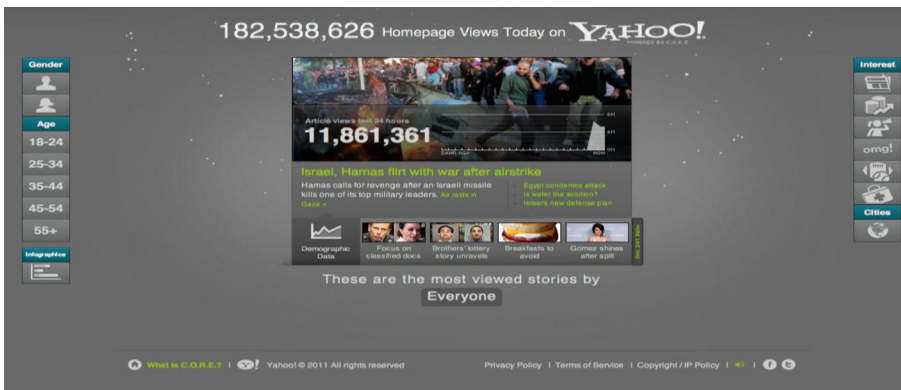
Figura 24



10) The most viewed stories by Yahoo!

(<http://beta.visualize.yahoo.com/core/>).

Figura 25



5.3. Recursos web complementaris

- Axismaps (<http://www.axismaps.com/index.php>).
- Visual Complexity (<http://www.visualcomplexity.com/vc/>).
- Bime Analytics (<http://bimeanalytics.com/showcase/>).
- Visualizing.org (<http://www.visualizing.org/explore>).
- Information is Beautiful (<http://www.informationisbeautiful.net/visualizations/>).

6. Cas model SocialNet: gestió del coneixement amb dades en obert

Des del punt de vista documental, Peset, Ferrer-Sapena i Subirats-Coll (2011) destaquen que amb l'entrada del món de les dades en obert es generen possibilitats d'innovació sense límits, però necessiten nous projectes que valoritzin aquestes dades i també altres formes de visualització que encara s'han d'explorar. Les autores alerten que el sector de la documentació no es defineix, no obstant això, perquè desenvolupa productes i serveis; fins i tot destaquen que el professional se sent més còmode amb protocols estandarditzats per a gestionar els seus fons més que per destacar un vessant més innovador.

A continuació presentem un concepte que vol conjuminar els elements esmentats anteriorment en un projecte específic, el prototip SocialNet, que busca generar un nou producte per a valoritzar les dades, oferir noves formes de visualització i mitigar la falta de desenvolupaments propis del camp de la informació i la documentació respecte a aquest nou univers explicat fins ara.

SocialNet es concep com un projecte per a generar de manera automàtica xarxes de coneixement a partir del binomi de relació simple "autors-documentos". La regla del desenvolupament és senzilla quant a la definició, el sistema a partir de subministrar-li un autor d'una organització ha de rastrejar Internet buscant documents, assignar-li l'autoria d'aquests documents creats, detectar si hi ha coautors, extreure aquests coautors i assignar-los nous treballs, fent que la xarxa del sistema es vagi retroalimentant i creixi gradualment a mesura que la generació de noves obres, amb o sense coautors, s'afegeixin al sistema.

Per a executar la regla exposada, es van detectar tres camps d'acció fonamentals, la normalització de l'estàndard, la gestió de les dades en obert i la visualització i l'anàlisi de la informació. Amb aquestes premisses, el sistema es va generar a partir d'un projecte que abordava de manera consistent el disseny i la creació de tres peces independents que unides entre elles permetessin oferir els resultats esperats. Per a això es va necessitar definir, dissenyar i conceptualitzar una sèrie d'elements clau com els següents:

- 1) **Un recollidor de dades:** cerca a Internet documents, autors i matèries.
- 2) **Un repositori:** espai per a conservar i editar els documents oposats propis i externs de la institució.
- 3) **Un sistema de visualització:** on establir les relacions visuals entre autors i documents.

Per a la normalització es van comparar diferents estàndards específics com LOM, SCORM i Dublin Core (DC) Qualificat, i es va decidir treballar sobre aquest últim model per a representar els recursos en el projecte (<http://goo.gl/0im5b>). Per a això es va necessitar crear un sistema de notació i etiquetatge en XML DC que marqués els diferents camps de cada registre recollits pel recollidor de dades, la qual cosa ens va confirmar la falta, en molts casos, d'estructuració estàndard de la informació o dada que es troba de lliure accés a Internet.

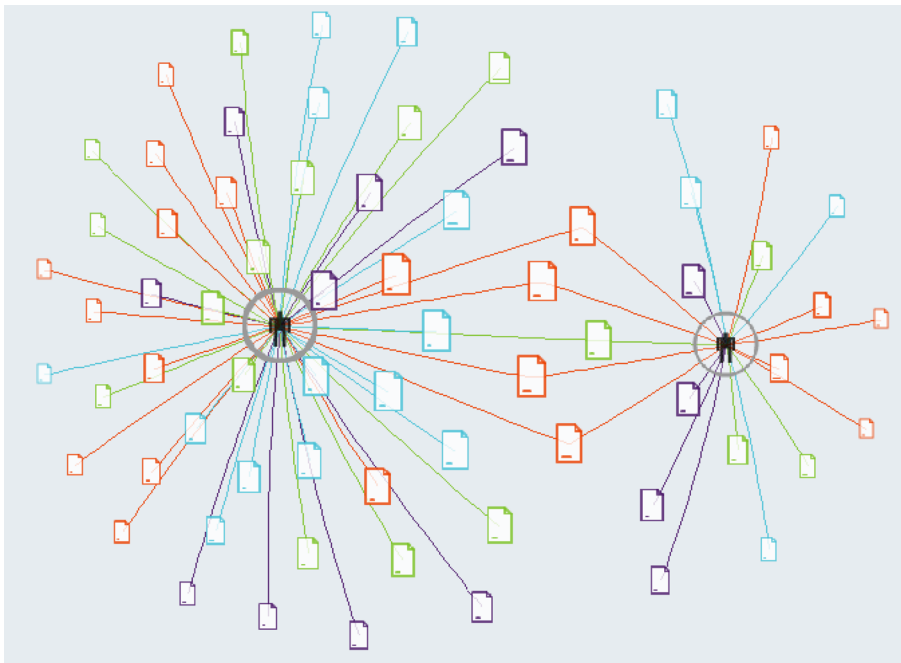
6.1. L'aplicació SocialNet per dins

Aquesta eina, que permet explorar la xarxa d'autors i documents indexats a la Universitat Oberta de Catalunya, es regeix per un conjunt d'eines interactives que la fan viable, com ara, un visor de grafs, un cercador, un panell estadístic i una fitxa d'autor. Vegem-los una mica més detalladament.

1) Visor de grafs

Aquest element, que possibilita la representació visual de dades, constitueix el component principal de l'aplicació. L'eina genera un espai hiperbòlic que s'autoordena i dimensiona perquè els elements representats no se superposin en l'espai visual de càrrega, i en el qual apareixen visualitzats dos tipus de nodes: autor, document i les relacions existents entre ells.

Figura 26. Graf de relació autor-document i autor-autor a partir de documents



Els nodes de tipus Autor apareixen a dins en un cercle amb un grossor de línia variable. El gruix del cercle és directament proporcional a la quantitat de documents amb què l'autor està relacionat.

Els nodes de tipus Document estan representats de manera que la intensitat de vora dels quals és proporcional a la quantitat d'autors amb què estan relacionats. Aquests nodes tenen assignat un color que en facilita la diferenciació visual respecte a d'altres documents en pantalla.

En obrir un node, desplega el conjunt de connexions amb què està relacionat. Al seu torn, el node que apareix en pantalla automàticament dibuixa les relacions amb altres nodes a la xarxa. El color de la relació està determinat pel node tipus document de la relació.

En passar el ratolí sobre un node, apareixen icones addicionals que permeten amagar el node al visor o, en el cas dels autors, accedir a la fitxa d'informació detallada. També apareix una finestra flotant que ofereix informació mínima sobre cada node, de manera que l'ajudi a reconèixer.

2) Cercador

El cercador permet trobar ràpidament qualsevol autor mitjançant una llista alfabètica. Cada element de la llista conté el nom de l'autor i una barra horitzontal d'escala que representa el nombre de documents amb els quals està relacionat.

Els elements de la llista es poden filtrar per nom, usant el camp d'entrada de text, o per categories. Les categories d'un autor estan determinades pel conjunt de matèries de tots els documents amb els quals està relacionat. Tant el cercador de text, com el filtre per categories, funcionen en temps real sense necessitat de confirmar la selecció.

Figura 27. Panell de cerca lliure, per autor i matèries

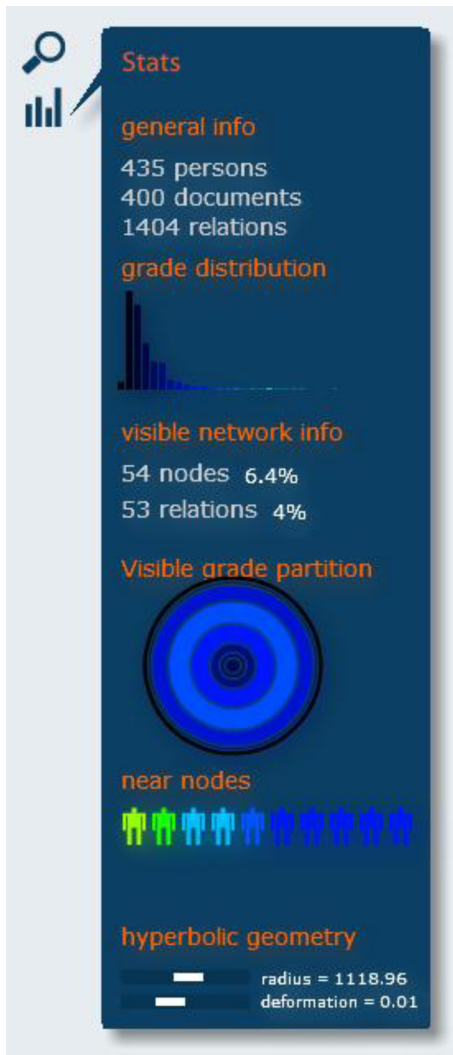


3) Panell estadístic

El panell d'estadístiques ofereix informació general sobre la xarxa generada, i també detalls de la subxarxa visible. Al seu torn, permet ajustar els paràmetres bàsics del sistema geomètric de representació. Consta de les seccions següents:

- a) **General info:** nombre d'autors, documents i relacions.
- b) **Grade Distribution:** gràfica de barres que representa la distribució d'autors segons el grau de les relacions.
- c) **Visible Network info:** nombre i proporció respecte del total de nodes i relacions actualment en pantalla.
- d) **Visible grade partition:** gràfica radial que representa el grau de partició de la subxarxa visible.
- e) **Near nodes:** llista de nodes més propers a la subxarxa visible.
- f) **Hyperbolic geometry:** controls de radi i deformació de la geometria de l'espai.

Figura 28



4) Fitxa d'autor

La fitxa d'autor ofereix informació detallada sobre els nodes de tipus Autor conforme als camps següents:

- Nom complet i organització
- Fotografia (deshabilitat)
- Descripció/CV
- Categories temàtiques associades
- Enllaços vinculats a xarxes socials de l'autor

La secció "Syndication" ofereix la llista de documents amb els quals l'autor està relacionat. Es pot accedir a l'URL original de cada document fent clic sobre el títol.

Figura 29. Model de fitxa d'autor i descripció



Lara Navarra, Pablo
Universitat Oberta de Catalunya

foros virtuales · aprendizaje colaborativo · entornos semipresenciales · E-learning · gestión de información · gestión de contenidos · tecnologías de la información y capacitación tecnológica · learning objects · gestión de contenidos · estrategias organizativas · e-learning · gestión de contenidos · TIC · e-learning · normalización · e-learning · bibliotecas digitales · gestión de contenidos · interoperabilidad de contenidos · administración electrónica · Administración pública · Universitat Oberta de Catalunya · UOC · e-learning · campus virtual · universidad · modelo educativo · innovación · UOC · Open University of Catalonia · e-learning · virtual campus · university · higher education · innovation · educational model · Gestión de la información · Administración pública ·

Description
Syndication

Description
Director of Technology Transfer Center in Bdigital-UOC And Director of Corporative R+D in Universitat Oberta de Catalunya

He holds a PhD from the University Pompeu Fabra. He completed the Master of Information Society and Knowledge for the Open University of Catalunya (UOC). It has a Diploma in library and information science and a BA in documentation by the University of Granada.

His teaching career is linked, as a Director of Innovation & Strategic Projects, Lecturer at the Faculty of Information Sciences and Communication of the University, and as academic director of the Accessible Technology Master of Information Society. This program is an educational project with the UOC Technosite belonging to the Foundation ONCE.

[f](#) [in](#) [t](#) [v](#)

Així mateix, mitjançant el panell d'autenticació (*log in*) és possible accedir a les fitxes d'autor en mode edició. Aquest accés permet editar les dades dels autors que no estiguin inferides per la xarxa, com descripció o enllaços a les xarxes socials on estigui present l'autor, i es poden aplicar opcions d'edició de format bàsiques.

Figura 30. Model d'edició de la fitxa d'autor



Maniega-Legarda, David
Universitat Oberta de Catalunya

FG Local government · HP · e-recursos · BD · Information society · E-learning · gestión de información · gestión de contenidos · interoperabilidad de contenidos · administración electrónica · Administración pública · Universitat Oberta de Catalunya · UOC · e-learning · campus virtual · universidad · modelo educativo · innovación · UOC · Open University of Catalonia · e-learning · virtual campus · university · higher education · innovation · educational model · Gestión de la información · Administración pública ·

Director de la Oficina de Innovación de la UOC
Writes about innovation, social media, emerging technologies and strategy.

Web technologies - Section chief Editor
Libraries Cooperation Service - Generalitat de Catalunya
September 2006 – October 2007

Webmaster - Project manager Editor

<http://www.facebook.com/dmaniega>

<http://es.linkedin.com/in/davidmaniega>

<http://twitter.com/dmaniega>

<http://delicious.com/dmaniega>

B **I** **U** **URL**

Bibliografia

Berners-Lee, Tim. (2009, febrer). "The next Web of open, linked data". *TED*. (https://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html)

Berners-Lee, Tim. Linked data (<http://www.w3.org/DesignIssues/LinkedData.html>)

Boletín Oficial de las Cortes Generales (http://www.senado.es/legis8/publicaciones/pdf/congreso/bocg/A_136-09.PDF).

Carrión, J. "Nuevos modelos en Internet para gestionar el talento". *Conjugación eficiente de las competencias estratégicas, técnicas y conductuales, con el entorno y la organización, para el logro de un desempeño superior y el conocimiento*. (<http://www.expansion.com/2001/02/08/empresas/808405.html>).

Castells, M. (1996). *The rise of the network society*. Oxford: Blackwell.

Chong, C. W.; Holden, T.; Wilhelmij, P.; Schmidt, R. (2000). "Where does knowledge management add value?". *Journal of Intellectual Capital* (vol. 1, núm. 4, pàg. 366-380).

Choo, C. W. (1998). *The knowing organization: how organizations use information to construct meaning, create knowledge, and make decisions*. Nova York: Oxford University Press.

Dans, E. "Big Data: una pequeña introducción". (<http://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html>).

Davenport, T.; Prusak, L. (1998). *Working knowledge: how organizations manage what they know*. Boston: Harvard Business School Press.

Domínguez, Eva. (2010). "Open data, ¿se abre la caja de Pandora?". *El cuarto bit* (19 nov.).

Esteban Navarro, Miguel Ángel; Navarro Bonilla, Diego (2003, juliol-agost). "Gestión del conocimiento y servicios de inteligencia: la dimensión estratégica de la información". *El Profesional de la Información* (vol. 12, núm. 5, pàg. 269-281).

Gorman, G. "DARPA and the Big Data, the future of autonomous Systems" (<http://huntall.com/darpa-big-data-future-autonomous-systems>).

Govern de les Illes Balears (<http://www.caib.es/caibdatafront/definicio.jsp?lang=es>).

KPMG. "Knowledge Management Research Report" (http://www.providersedge.com/docs/km_articles/kpmg_km_research_report_2000.pdf).

Kusnetzky, D. "What is 'Big Data'?" (<http://www.zdnet.com/blog/virtualization/what-is-big-data/1708>).

Martínez, H. "The next big thing". (<http://blog.upcnet.es/2012/01/27/the-next-big-thing>).

Open Data RISP. "Decálogo Open Data". (<http://red.gnoss.com/comunidad/OpenData/recurso/Decalogo-Open-Data/e350c5b3-78ec-470e-b5aa-a0af9bb0a594>).

Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma. (2011, març-abril). "Open data y linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información* (vol. 20, núm. 2, pàg. 165-173). (<http://dx.doi.org/10.3145/epi.2011.mar.06>).

Saz del, Miguel Ángel (2001, abril). "Gestión del conocimiento: pros y contras". *El Profesional de la Información* (vol. 10, núm. 4, pàg. 14-28).

Wikipedia. Big Data. (http://es.wikipedia.org/wiki/Big_data).

Abadal, E. i altres (1998). "Intranets documentales". *El Profesional de la Información*.

Alòs-Moner, Adela d' (2002, gener-febrer). "Intranets: sonrisas y lágrimas". *El Profesional de la Información* (vol. 11, núm. 1, pàg. 4-8).

Armbrust, M.; Fox, A.; Griffith, R. i altres (2010, abril). "A View of Cloud Computing". *Communications of the ACM* (vol. 53, núm. 4, pàg. 50-58).

Broderick, K; Bailey, M; Eastwood, M. (2010, abril). *Worldwide Enterprise Server Cloud Computing 2010-2014 Forecast*. IDC.

Bustelo-Ruesta, C.; García-Morales Huidobro, E.; García de Paso Gómez, E. (2006, setembre-octubre). "Definición del modelo estratégico de desarrollo de una intranet corporativa: metodología basada en un caso práctico". *El Profesional de la Información* (vol. 15, núm. 5, pàg. 352-362).

Erdogmus, H. (2009). "Cloud Computing: Does Nirvana Hide behind the Nebula?". *IEEE Software* (vol. 26, núm. 2, pàg. 4-6).

Gartner. (2008). "Cloud Computing Will Be As Influential As E-business". (<http://www.gartner.com/it/page.jsp?id=707508>).

Kisker, H.; Ried, S; Shey, H. (2010). *The State Of Enterprise Software And Emerging Trends: 2010*. Forrester Research.

Leavitt, N. (2009). "Is Cloud Computing Really Ready for Prime Time?". *Computer* (vol. 42, núm. 1, pàg. 15-20).

Li, H. (2009). *Introduction to Windows Azure: an introduction to cloud computing using Microsoft Windows Azure*. Berkeley / Nova York: Apress / Springer-Verlag New York, cop.

Nielsen, J. (2011). "10 Best Intranets of 2011". (http://www.useit.com/alert-box/intranet_design.html).

Pérez González, D.; Solana, P. (2006, setembre-octubre). "Intranets medición y valoración de sus beneficios en las organizaciones". *El Profesional de la Información* (vol. 15, núm. 5, pàg. 331-341).

Rhoton, J. (2009). *Cloud computing explained: handbook for enterprise implementation*. Londres: Recursive Ltd.

Velte, A.; Velte, T.J.; Elsenpeter, R.C. (2010). *Cloud Computing: A Practical Approach*. Nova York: McGraw-Hill.

Voas, J.; Zhang, J. (2009, març-abril). "Cloud Computing: New Wine or Just a New Bottle?". *IT Professional* (vol. 11, núm. 2, pàg. 15-17) doi:10.1109/MITP.2009.23.

Corkern, Carla (2001). "XML Application integration and knowledge management". *XML Europe 2001. How XML powers industry applications*. 21-25 maig, 2001, Berlín. Disponible a: (http://www.gca.org/attend/2001_conferences/europe_2001/knowledgemanagement.htm).

Díaz Ortuño, Pedro Manuel (2003). "Problemática y tendencias en la arquitectura de metadatos web". *Anales de Documentación* (núm. 6, pàg. 35-58).

Eden, Brad. (2002, setembre-octubre). "Metadata standards". *Library Technology Reports*.

García Jiménez, Antonio. (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de Documentación* (núm. 7, pàg. 79-95).

Gardner, John Robert (2001). "Information architecture planning with XML". *Library Hi Tech* (vol. 19, núm. 2, pàg. 231-241).

Martín Galan, Bonifacio; Rodríguez Mateos, David. (2000). "Estructuración de la información mediante XML: un nuevo reto para la gestión documental". *VII Jornadas Españolas de Documentación: la gestión del conocimiento, retos y soluciones de los profesionales de la información* (octubre de 2000, pàg. 113-123). Bilbao.

Martinez Usero, José Angel (2006). "El uso de metadatos para mejorar la interoperabilidad del conocimiento en los servicios de administración electrónica". *El Profesional de la Información* (vol. 15, núm. 2, pàg. 114-126). (<http://www.ucm.es/eprints/5638/>).

Martínez Usero, José Ángel; Beltrán Orenes, María Pilar (2005). "Ontologies in the context of knowledge organization and interoperability in e-government services". A: *Conference on Digital Divide, Global Development and the Information Society*. Tunísia. (<http://www.ucm.es/eprints/5631/>).

Mendoza, Cornelius (2000). "Another use of XML: internal knowledge management". *Serverworld Magazine* (núm. 6). (<http://www.serverworldmagazine.com/compaqent/2000/06/anotheru.shtml>).

Montero, Yusef Hassan; Nuñez Peña, Ana (2005). *Diseño de arquitecturas de información: descripción y clasificación*. (http://www.nosolousabilidad.com/articulos/descripcion_y_clasificacion.htm).

Peis, Eduardo; Moya de, Felix (2000, juny). "SGML y servicios de información". *El Profesional de la Información* (vol. 9, núm. 6, pàg. 4-17).

Resource Description Framework (RDF). W3C Semantic Web Activity. (<http://www.w3c.org/RDF>).

Rosenfeld, Louis; Morville, Peter (2002). *Information architecture for the World Wide Web: designing large-scale web sites*. O'Reill.

San Segundo, Rosa; Beltrán, Pilar. "Aplicación de ontologías en la organización de información en Internet". A: *CISCI 2003. Memorias. 2da. Conferencia iberoamericana en sistemas, cibernética e informática*. Orlando (Florida).

Serradell López, Enric; Juan Pérez, Ángel A. (2003). *La gestión del conocimiento en la nueva economía*. (<http://www.uoc.edu/dt/20133/index.html>).

