

Les metadades als dipòsits digitals de la Universitat Politècnica de Catalunya

Barcelona, setembre de 2008

Ester Badia

Ester.Badia@upc.edu

Toni Prieto

Antonio.Juan.Prieto@upc.edu

Jordi Prats

Jordi.Prats@upc.edu

Universitat Politècnica de Catalunya
Servei de Biblioteques i Documentació

1. Presentació
2. **UPCommons: portal d'accés obert al coneixement de la UPC**
3. Metadades
4. Serveis
5. Conclusions

1. Presentació (dades 2007)

Biblioteques	13
Visitants	2.887.237
Préstecs	276.201
Volums	535.545
Revistes-e	7.960
Bibliotècnica - sessions	5.747.036
UPCommons - sessions	2.229.788
Catàleg - sessions	309.261
Infraestructura TIC	696 PC/ portàtils i 7 servidors
Punts de lectura	3.125
Bibliotecaris	94
Tècnics de suport	44
Administratius	9
Personal TIC	4

- 1. Presentació**
- 2. UPCommons: portal d'accés obert al coneixement de la UPC**
- 3. Metadades**
- 4. Serveis**
- 5. Conclusions**

2. UPCommons: portal d'accés obert al coneixement de la UPC

Objectius principals

- Oferir a la comunitat universitària de la UPC una eina electrònica per publicar els seus treballs acadèmics, en accés obert i en text complet a Internet.
- Garantir l'accés perpetu als treballs arxivats mitjançant el seu emmagatzematge en servidors segurs mantinguts i gestionats pel Servei de Biblioteques i Documentació.
- Millorar la visibilitat de la universitat i dels seus autors a la xarxa.

2. UPCommons: portal d'accés obert al coneixement de la UPC

2001: TDX

2002: Arxiu Gràfic de l'ETSAB

2003: Videoteca Digital de la UPC

2004: Treballs acadèmics UPC

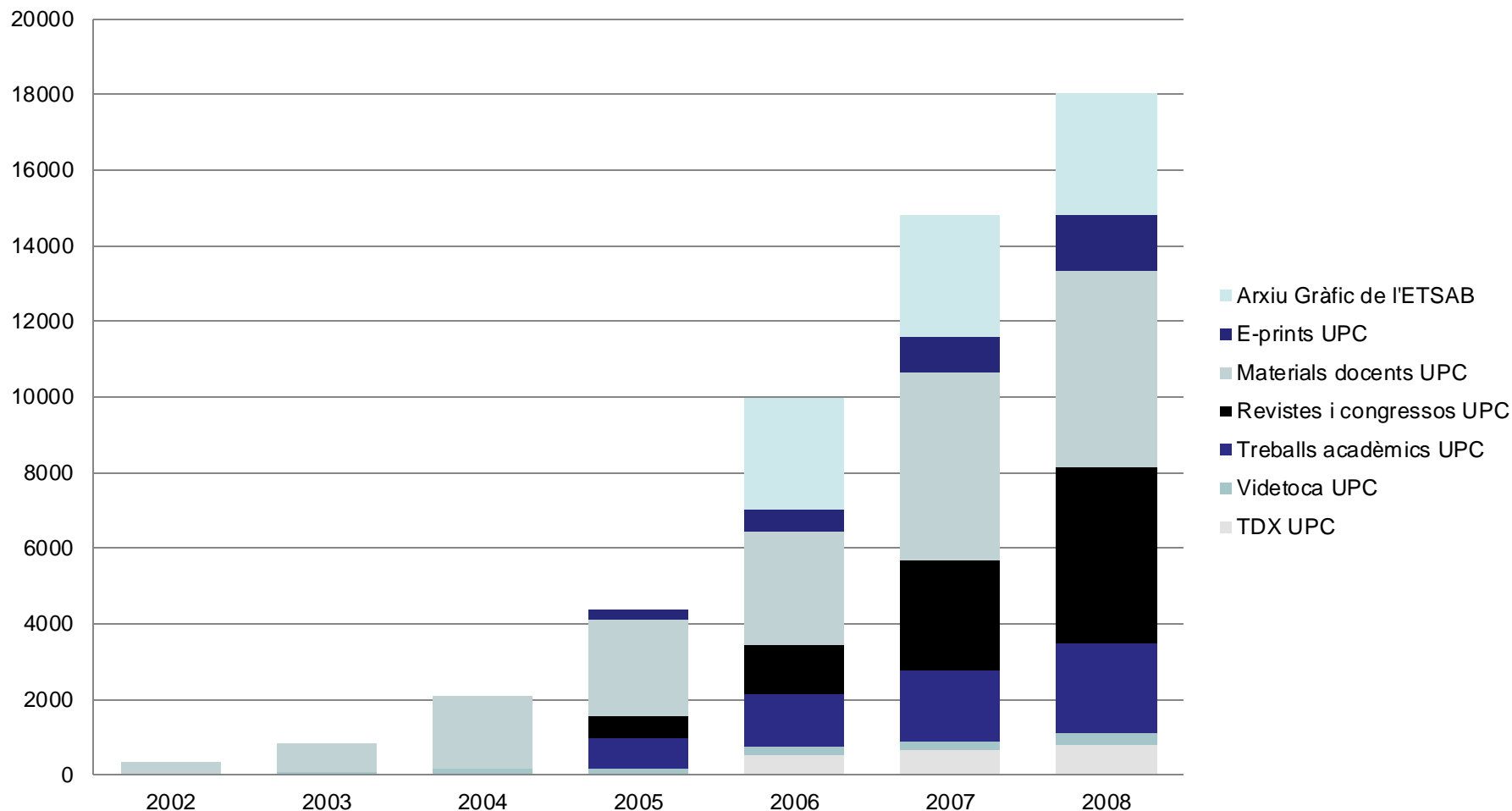
2005: Revistes i congressos UPC

2006: E-prints UPC

2007: Materials docents / OCW

2. UPCommons: portal d'accés obert al coneixement de la UPC

Evolució anual d'UPCommons quan a nombre de documents



- 1. Presentació**
- 2. UPCommons: portal d'accés obert al coneixement de la UPC**
- 3. Metadades**
- 4. Serveis**
- 5. Conclusions**

3. Metadades. Model de metadades UPCommons

Model de metadades UPCommons?

- Es tracta d'un projecte evolutiu, que s'ha generat per la suma dels repositoris que s'han creat a la UPC o en els que la universitat participa.
- No es planteja en el seu moment un model de metadades per a tots els continguts que hi poden haver potencialment en els diferents productes, sinó que s'aprofita l'experiència adquirida en el desenvolupament de nous serveis.
- La Unitat de Recursos per a l'Aprenentatge ha estat sempre l'encarregada de dissenyar i evolucionar els models de metadades per a cada un dels repositoris.
- El resultat es un model coherent que ens fa pensar que disposem d'un model de metadades UPCommons.

3. Metadades. Model de metadades UPCommons

Basat en Dublin Core

- Altres models (p. ex. IMS Resources Meta-data Information Model) presenten metadades massa específiques i complexes que dificulten la descripció dels objectes.
- Les metadades i qualificadors DC recullen les nostres necessitats descriptives d'objectes emprats tant a la docència/aprenentatge com a la recerca.
- Possibilitat d'utilitzar qualificadors propis.
- Per defecte DSpace està configurat amb l'esquema de metadades DC.
- Dublin Core Metadata Terms és extensament emprat a dipòsits de tot el món: dipòsits del CBUC, etc.
- Fàcil conversió de les metadades MARC21 a Dublin Core.

3. Metadades

Fonts consultades

- Dublin Core Metadata Initiative <<http://dublincore.org/>>
 - DCMI Education Working Group.
 - Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata
- Sistemes d'informació de la pròpia universitat.
- CBUC.
 - Pautes i recomanacions del CBUC per a l'ús de metadades Dublin Core en recursos web.
 - Model de metadades Recercat.
- IMS Resources Meta-data Information Model
<http://www.imsglobal.org/metadata/imsmdv1p2p1/imsmd_infov1p2p1.html>
>
- Altres dipòsits de recerca: DIGITAL.CSIC <<http://digital.csic.es/>>.

3. Metadades

Elements DC a UPCommons

- Elements DC emprats:
 - Conjunt dels 15 elements del Simple DC Metadata Element Set.
 - + element “Audience” proposat pel DCMI Education Working Group com a ampliació dels 15 elements DC <<http://dublincore.org/documents/2000/10/05/education-namespace/>>
- Qualificadors DC emprats:
 - Qualificadors aprovats pel DCMI Usage Committee.
 - + altres qualificadors propis amb què poder recuperar els objectes segons les nostres necessitats docents i de recerca (p. ex.: “Subject.CCUC”, Description.Peerreviewed”, ...)

44 metadades:

- 16 comuns a tots els repositoris.
- 31 obligatoris en algun dels repositoris.

3. Metadades

Origen

- Només 6 elements capturats automàticament.
- 13 elements procedents de l'autoarxiu (E-prints UPC i OCW).
- La resta incorporats manualment a les biblioteques.

3. Metadades

Qualitat

- Un **bibliotecari** revisa les metadades mínimes del “submitter” i introdueix altres metadades completes per millorar la descripció i recuperació dels objectes:
 1. Comprova la correcció del fitxer/s dipositat/s
 2. Completa la descripció amb altres metadades (matèries, tipus de document, ...)
 3. Valida les entrades d'autor i matèries segons el CCUC (LEMAC/LENOTI), la Library of Congress Subject Headings i l'arbre de matèries de Bibliotècnica
 4. Revisa i corregeix bimensualment les possibles entrades d'autor duplicades
- Ús, quan es possible, de llistes controlades i taxonomies.
- Disponibilitat d'especificacions per a cada tipus de document.

3. Metadades

Actualment

- Adaptació a recomanacions DRIVER dels dipòsits de recerca:
 - Reconversió dels valors de la metadada DC “Type” emprats fins al moment a UPCommons en els valors recomanats per les Directrius DRIVER (<http://www.driver-support.eu/en/guidelines.html>)[1], la qual cosa ha estat possible totalment als dipòsits de recerca E-prints UPC, Revistes i Congressos UPC i Videoteca UPC.
- Estudi per instaurar l'autoarxiu al dipòsit “Treballs acadèmics UPC” (actualment els alumnes, a diferència del PDI i PAS, són els únics que no poden autoarxivar els seus treballs).
- Semi-automatització d'alguns dels elements obligatoris.

- 1. Presentació**
- 2. UPCommons: portal d'accés obert al coneixement de la UPC**
- 3. Metadades**
- 4. Serveis**
- 5. Conclusions**

4. Serveis

1. Integració amb altres sistemes
2. Importació de continguts
3. Preservació

4. Serveis. Integració amb altres sistemes

1.- Sistemes d'informació de la UPC

- Documentació generada per l'activitat docents i de recerca de la UPC.
- Metadades ja utilitzades anteriorment en els processos de gestió de la universitat.
- Bidireccional: *UP Commons* necessita accedir a informació normalitzada i de qualitat pel seu funcionament i a al vegada pot aportar metadades sobre publicacions de la universitat.

4. Serveis. Integració amb altres sistemes

2.- Projectes de col·laboració

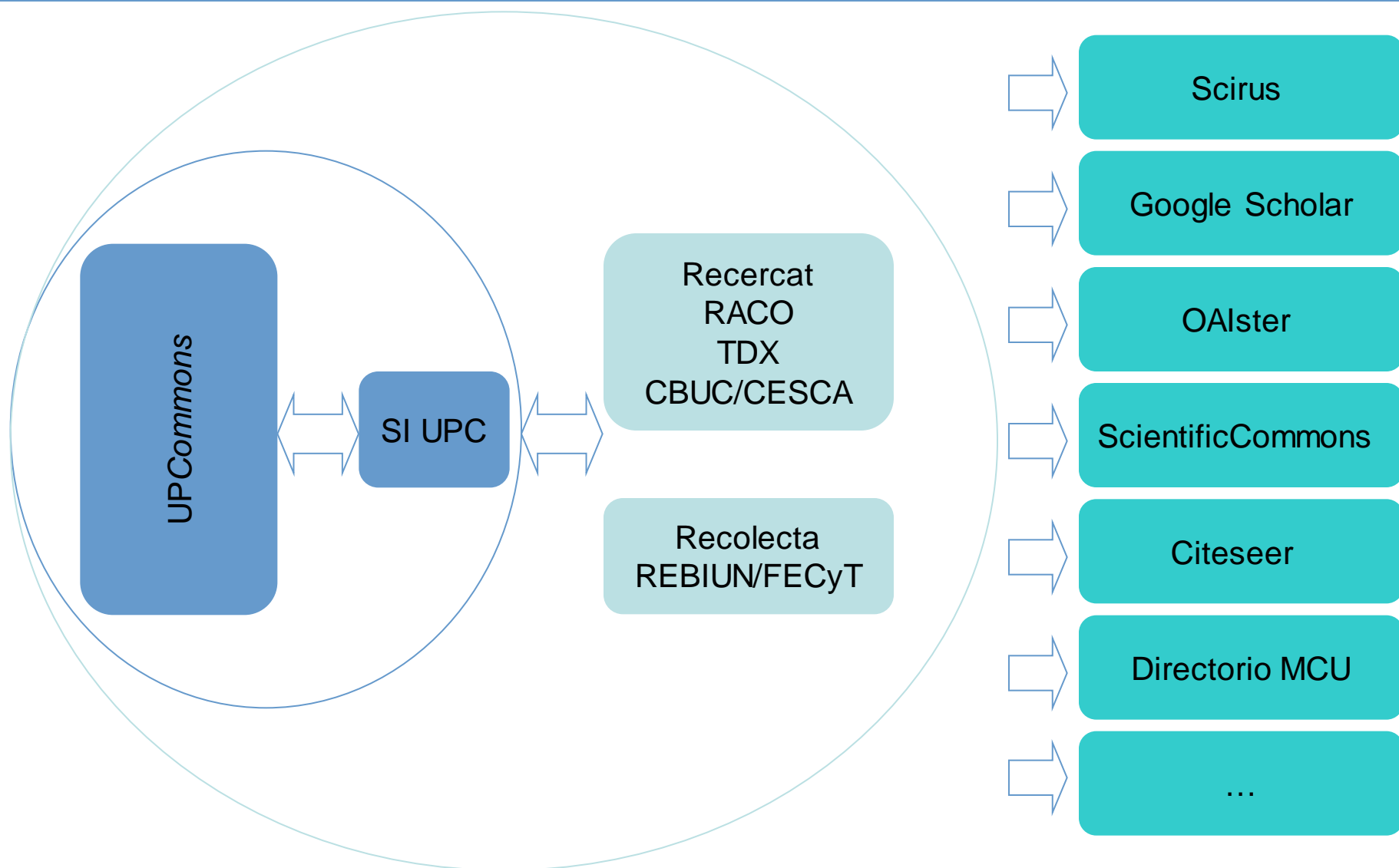
- Permeten compartir experiències i treballar conjuntament en el disseny i desenvolupament de productes.
- El desenvolupament d'estàndards oberts permet trobar fórmules de col·laboració flexibles i adaptades a les necessitats de cada una de les institucions participants.

4. Serveis. Integració amb altres sistemes

3.- Visibilitat en el món

- Augmentar la presència i la visibilitat de la universitat i dels seus autors a Internet es un dels principals objectius del projecte.
- Sistemes casi desatesos: poc contacte amb els proveïdors de serveis.

4. Serveis. Integració amb altres sistemes



4. Serveis. Integració amb altres sistemes

- Us rellevant del protocol OAI-MHP...
- ... acompanyat d'altres solucions: serveis SOAP, SRU, importacions de dades, metadades a la capçalera de documents HTML, etc...
- Valor de la qualitat de les metadades per sobre de la quantitat.
- Estratègic pel desenvolupament d'UPCommons:
 - Estalvi de costos en la creació de metadades.
 - Simplificació dels processos de publicació.
 - Millora de la visibilitat dels continguts publicats.

4. Serveis. Migració

Migració de registres i continguts

- Basat en solucions pròpies desenvolupades pels sistemes que s'estan utilitzant (format d'importació de DSpace, etc...).
- Molt utilitzat, sobretot en processos de digitalització de continguts o migració de documents d'altres plataformes.
- Necessitat de treballar amb estàndards per tal de poder obtenir un model de migració de continguts apta per a qualsevol repositori.

4. Serveis. Migració

Migració de registres i continguts

Transformació de metadades => projecte de creació del repositori de col·leccions històriques..

- Hem realitzat conversions de MARC21 a Dublin Core qualificat per migrar registres DSpace:
 - Extracció de registres en MARC21.
 - Conversió dels registres MARCXML amb Yaz Toolkit.
 - Conversió a DC qualificat a través d'un fulla XSL disponible a la web de MARCXML que s'ha modificat per adaptar-la al format de metadades d'UPCommons.

4. Serveis. Migració

Migració de registres i continguts

Importació de paquets METS (Metadata Encoding & Transmission Standard)

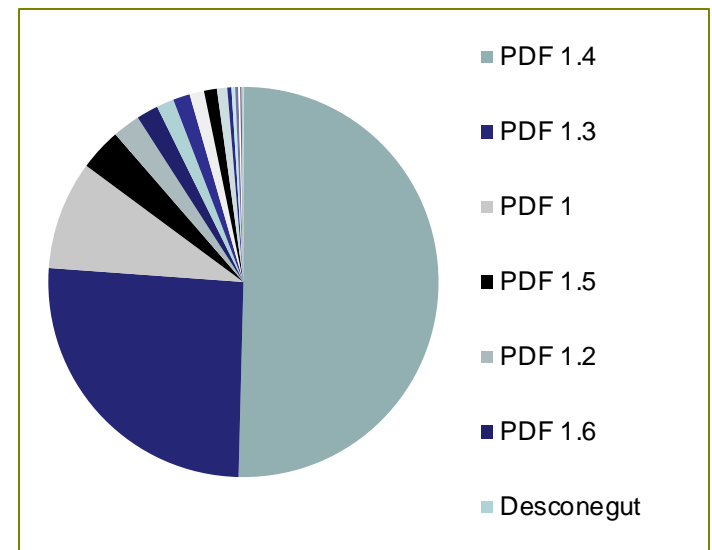
S'està treballant per realitzar importacions de registres des d'altres unitats empaquetats en METS:

- Actualment en fase de desenvolupament per incloure el projectes finals de Màster de la Facultat d'Informàtica de Barcelona al Dipòsit de Treballs acadèmics.
- Fent servir l'esquema MODS (Metadata Objects Description Schema) per les metadades descriptives.
- S'ha partit de la conversió de MODS a Dublin Core qualificat disponible inicialment amb DSpace que s'ha ampliat amb noves correspondències per metadades que es van definir posteriorment (departament, nivell d'estudis, escola, ...).

4. Serveis. Preservació

Preservació

- No es disposa encara d'una estratègia de preservació a llarg termini.
- Possible estratègia basada en la transformació de formats.
 - ODF (ISO 26300).
 - PDF (ISO 19005).
- 93% dels arxius en format PDF (font ROAR)
- No es disposa encara d'una definició de la informació que necessitaríem per realitzar el procés.



- 1. Presentació**
- 2. UPCommons: portal d'accés obert al coneixement de la UPC**
- 3. Metadades**
- 4. Serveis**
- 5. Conclusions**

5. Conclusions

- Alt cost en la generació de metadades.
 - Volum cada cop més important d'informació.
 - Possible crisi del sistema d'autoarxiu.
 - Alternatives: reutilització de metadades, automatització en la generació, ajustar el model de metadades a les necessitats dels projecte, ...
- Però rellevància la necessitat de disposar de matedades verificades i de qualitat.
 - Pel desenvolupament de serveis de valor afegit.
 - Per facilitar la integració dels repositoris en els SI de la pròpia institució.
 - Per garantir la preservació dels documents a llarg termini.
 - Per identificar no publicats.
 - ...

5. Conclusions

- Importància en la disponibilitat d'estàndards oberts de transmissió de dades, que faciliten el ràpid desenvolupament de sistemes que els suporten.
- Canvis en la definició dels rols de les persones que intervenen en la gestió dels repositoris:
 - Bibliotecaris temàtics => catalogadors
- Valoració a la pròpia institució del Servei de Biblioteques i Documentació quan a experts en metadades, tant quan al disseny de models com en la garantia de qualitat.

Gràcies per la vostra atenció

Ester Badia

Ester.Badia@upc.edu

Toni Prieto

Antonio.Juan.Prieto@upc.edu

Jordi Prats

Jordi.Prats@upc.edu

Universitat Politècnica de Catalunya
Servei de Biblioteques i Documentació