



## Primary data in a institutional repository: a case study

LIBER 30th June 2011

This work, developed in the framework of the CLARIN project, has been supported by the Spanish Ministerio de Ciencia y Tecnología (ACI2009-0995), the Departament d'Innovació, Universitats i Empresa of the Generalitat de Catalunya, and the European Union (FP7- INFRASTRUCTURES-2007-1-212230).

# Primary data in a institutional repository: a case study

## Content

1. Introduction
2. The CLARIN project
3. Institutional repositories and their contents
4. The community
5. Conclusions and future development

# 1. Introduction

**3** fields  
campus  
research parks



**Ciutadella campus:**  
social sciences and humanities



**Mar campus:**  
health and life sciences



**Poble Nou campus:**  
communication and  
information technologies



# 1. Introduction

- 8,920 undergraduate students
- 4,198 postgraduate students  
(taking official university master's degrees, UPF-  
endorsed master's degrees and doctoral  
programmes)
- 1,205 teaching and research staff  
(782 equivalent to full-time)
- 687 administrative staff
- 19 bachelor's degree programmes
- 126 master's degree programmes
- 9 doctoral programmes
- 63 recognised research groups
- Budget (2011): 130 million euros

# 1. Introduction

The Institut Universitari de Lingüística Aplicada (IULA) is a research and graduate training center at Universitat Pompeu Fabra



# 1. Introduction



**Linguistic Resources  
Technologies group**

**CLARA**

Common Language Resources and their Applications — a Marie Curie ITN



<http://www.iula.upf.edu/trl/rprojos.htm>

# Primary data in a institutional repository: a case study

## **Content**

1. Introduction
2. The CLARIN project
3. Institutional repositories and their contents
4. The community
5. Conclusions and future development

## 2. The CLARIN project

# CLARIN

Common Language Resources and Technology Infrastructure



About CLARIN

Services

Publications

Activities

Events

Links

Glossary

Contact

Join CLARIN

Internal Web Site

**January 28, 2011**

New Virtual Language  
Observatory launch



For making a research infrastructure feasible we need a light and transparent framework that protects everybody's rights and interests.

**Steven Krauwer**

2010



Activities



Publications



Solutions



About  
CLARIN





## 2. The CLARIN project

The screenshot shows the CLARIN-ES website homepage. At the top right, there are language options for 'Español' and 'Català'. The main header features the 'CLARIN-ES' logo and the tagline 'Common Language Resources and Technology Infrastructure'. A navigation menu includes 'Principal', 'Noticias', 'Eventos', 'Descargas', 'Web Services', 'Preguntas Frecuentes', 'Newsletter', 'Equipo', and 'Contacto'. A quote by Ignacio Bosque is displayed on the left. On the right, logos for the Spanish Government and the Catalan Government are shown. Below these are sections for 'Últimas Noticias' (Latest News) and 'Proyectos relacionados' (Related Projects).

Español Català

# CLARIN-ES

Common Language Resources and Technology Infrastructure

Principal Noticias Eventos Descargas Web Services Preguntas Frecuentes Newsletter Equipo Contacto

*"precisamente porque las máquinas nos proporcionan y nos ordenan... los datos que les pedimos, debemos dedicar a la tarea de reflexionar sobre ellos buena parte del tiempo que antes empleábamos en conseguirlos"*

Ignacio Bosque, director del diccionario REDES.

**GOBIERNO DE ESPAÑA**  
**MINISTERIO DE CIENCIA E INNOVACIÓN**

**Generalitat de Catalunya**  
**Departament d'Innovació, Universitats i Empresa**

### ¿Qué es CLARIN?

CLARIN (*Common Language Resources and Technologies*) es un proyecto de colaboración europea a gran escala para dar acceso genérico a grandes bancos de datos lingüísticos (textos, grabaciones multimedia, diccionarios, etc.), así como a los instrumentos de análisis y explotación de estos datos (segmentadores, etiquetadores, analizadores sintácticos, etc.), mediante la utilización de tecnologías de la información y las comunicaciones.

Para ello implementará en una estructura de red Grid, y gracias a tecnología de servicios web v de web semántica. una única interfaz de

#### Últimas Noticias

- Colaboración con el proyecto "Municipals'11 online"
- Doble número 11-12 de la Newsletter CLARIN.eu
- Publicación del borrador de informe "The Future European Multilingual Information Society" de la Red de Excelencia META-NET

#### Proyectos relacionados

- CLARIN Europa
- Enlaces

# Primary data in a institutional repository: a case study

## **Content**

1. Introduction
2. The CLARIN project
3. Institutional repositories and their contents
4. The community
5. Conclusions and future development

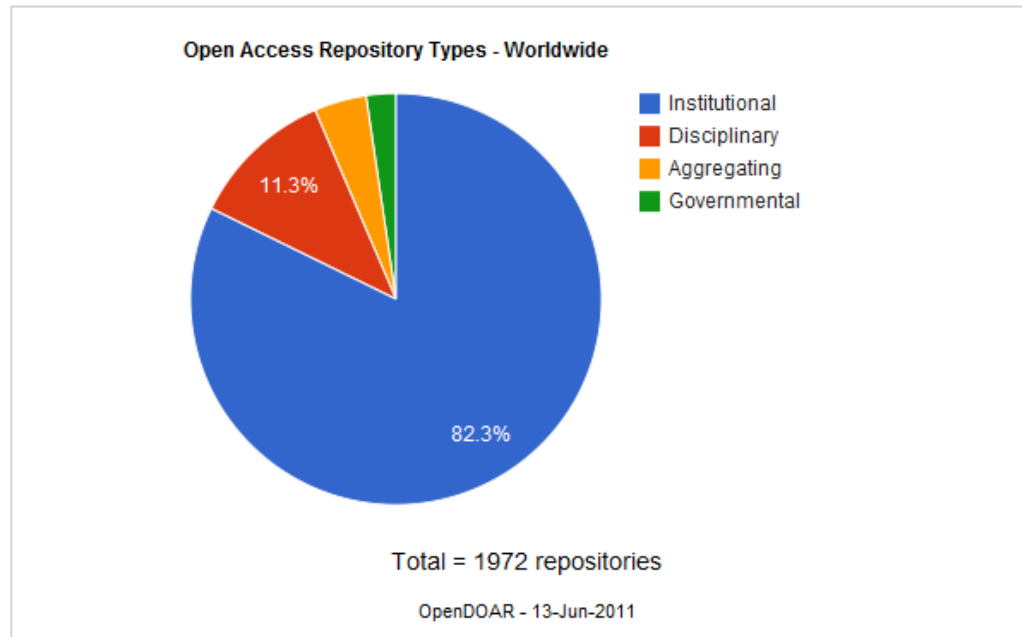
# 3. The institutional repositories and their contents

OpenDOAR

Directory of Open Access Repositories

[Home](#) | [Find](#) | [Suggest](#) | [Tools](#) | [FAQ](#) | [About](#) | [Contact Us](#)

## Open Access Repository Types - Worldwide



Further Explanation of Repository Types

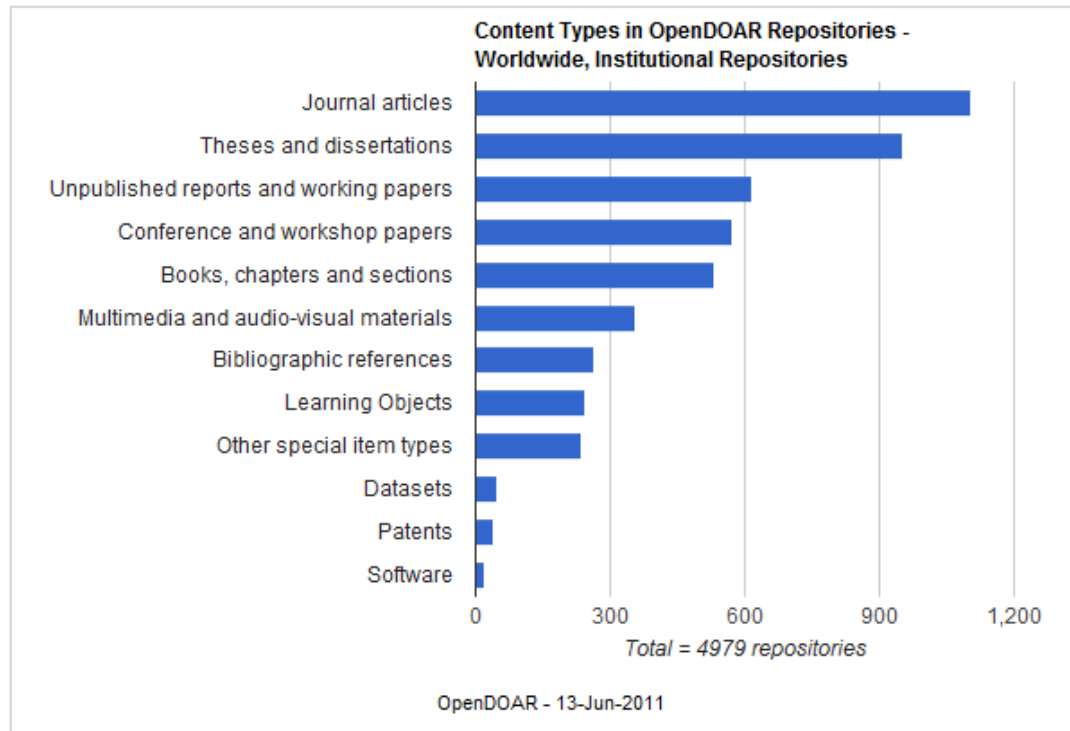
# 3. The institutional repositories and their contents

OpenDOAR

Directory of Open Access Repositories

[Home](#) | [Find](#) | [Suggest](#) | [Tools](#) | [FAQ](#) | [About](#) | [Contact Us](#)

## Content Types in OpenDOAR Repositories - Worldwide, Institutional Repositories



N.b. Most repositories hold several Content Types.

### 3. The institutional repositories and their contents

#### Progression from data to content

- Materials (not textual or written):  
ncipient incorporation into contents of  
institutional repositories.





# Primary data in a institutional repository: a case study

## **Content**

1. Introduction
2. The CLARIN project
3. Institutional repositories and their contents
4. The community
5. Conclusions and future development



**TDX**

**RECERCAT**

**RACO**

**mdx**

**e-Repository**

**Google**

**Alster**  
*...find the pearls*

**Thesis**

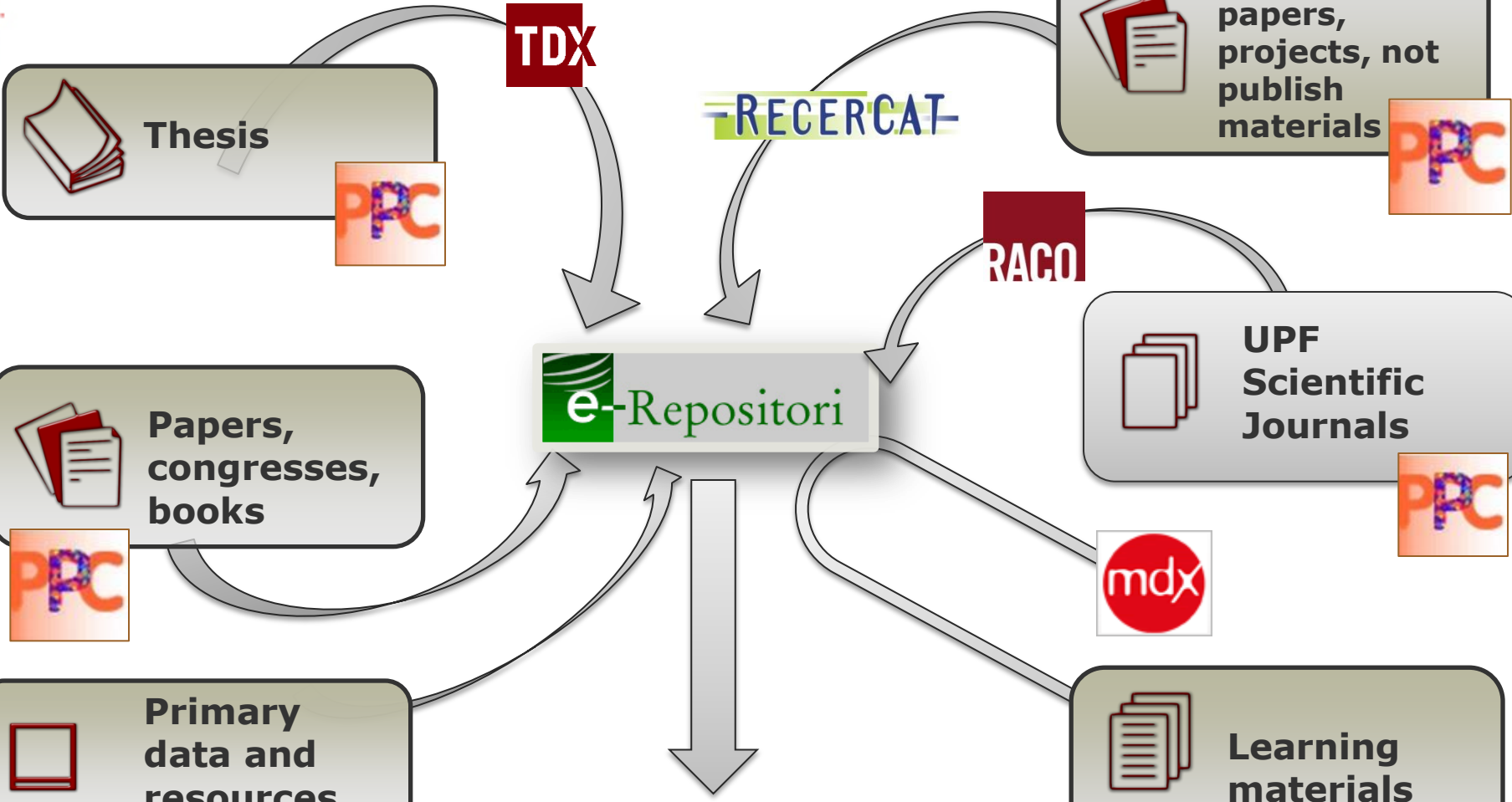
**Working papers, projects, not publish materials**

**Papers, congresses, books**

**UPF Scientific Journals**

**Primary data and resources**

**Learning materials**



[Home](#)**Search** [Advanced Search](#)**Browse by:**

- **All the e-Repository**
  - [Communities & Collections](#)
  - [Titles](#)
  - [Authors](#)
  - [Subjects](#)
  - [By Issue Date](#)

**My Account**

- [Login](#)

[About the e-Repository](#) [Statistics](#)

## Welcome to the UPF Digital Repository

The institutional repository collects, disseminates and preserves, in digital form, the intellectual output that results from the academic and research activity of the UPF and also the scientific magazines and the institutional publications. Its purpose is to increase the impact of research done at the UPF and its intellectual memory.

## Search

Enter some text in the box below to search.

## Communities

Select a community to browse its collections.

- [Docència](#) [37]
- [Recerca: articles, congressos, llibres](#) [33]
- [Recerca: tesis](#) [540]
- [Recerca: working papers, PFC, etc.](#) [1.349]
- [Recursos i dades primàries](#) [42]
- [Revistes científiques](#) [2.771]
- [Continguts institucionals](#) [84]
- [Vida universitària](#) [3]





## 4. The community “Resources and research primary data”

- Two collections
  - “IULA. Tools and resources”
  - “IULA. Primary data from PhD Theses”
- Content:
  - Primary data
  - Resources
  - Linguistic tools

From PhD theses and research projects

Inicio → Recursos i dades primàries

### Buscar

Buscar

- Esta comunidad
- [Búsqueda avanzada](#)

### Listar por:

- **Todo el e-Repository**
  - [Comunidades y colecciones](#)
  - [Títulos](#)
  - [Autores](#)
  - [Materias](#)
  - [Por fecha de publicación](#)
- **Esta comunidad**
  - [Títulos](#)
  - [Autores](#)
  - [Materias](#)
  - [Por fecha de publicación](#)

### Mi cuenta

- [Iniciar](#)

¿Qué es el e-Repository ?

Estadísticas

SHERPA  
RoMEO

DULCINEA

## Recursos i dades primàries

Introduzca el texto:

Listar por:

Dades primàries de recerca i els instruments d'anàlisi i explotació d'aquestes dades.

### Colecciones en esta comunidad

- [IULA. Dades primàries de les tesis doctorals](#) [6]
- [IULA. Eines i recursos](#) [36]

## Envíos recientes

[Corpus d'entrevistes sociolingüístiques a angloparlants immigrants a Espanya en temps aparent \(TA\)](#) [Corpus d'entrevistes sociolingüístiques a angloparlants immigrants a Espanya en temps real \(TR\)](#)

[Cercador NEOROM](#)

[Eines de processament de corpus](#)

[Corpus Técnico del IULA](#)

## Buscar

 Ir

Buscar

- Esta colección  
 [Búsqueda avanzada](#)

## Listar por:

- **Todo el e-Repository**
  - [Comunidades y colecciones](#)
  - [Títulos](#)
  - [Autores](#)
  - [Materias](#)
  - [Por fecha de publicación](#)
- **Esta colección**
  - [Títulos](#)
  - [Autores](#)
  - [Materias](#)
  - [Por fecha de publicación](#)

## Mi cuenta

- [Iniciar](#)

[¿Qué es el e-Repository ?](#) [Estadísticas](#)

## IULA. Eines i recursos

Introduzca el texto:  IrListar por: [Títulos](#) [Autores](#) [Materias](#) [Fecha de publicación](#)

Eines i recursos resultants dels projectes de recerca de l'Institut de Lingüística Aplicada (IULA), descrits d'acord amb l'esquema de metadades proposat en el marc del projecte europeu CLARIN (Common Language Resources and Technologies).

## Envíos recientes

[Corpus d'entrevistes sociolingüístiques a angloparlants immigrants a Espanya en temps aparent \(TA\)](#) [Corpus d'entrevistes sociolingüístiques a angloparlants immigrants a Espanya en temps real \(TR\)](#)

[Cercador NEOROM](#)[Eines de processament de corpus](#)[Corpus Técnico del IULA](#)

## Search

Search

 This Collection[Advanced Search](#)

## Browse by:

- **All the e-Repository**
  - [Communities & Collections](#)
  - [Titles](#)
  - [Authors](#)
  - [Subjects](#)
  - [By Issue Date](#)
- **This Collection**
  - [Titles](#)
  - [Authors](#)
  - [Subjects](#)
  - [By Issue Date](#)

## My Account

- [Login](#)

[About the e-Repository](#)[Statistics](#)

## IULA. Dades primàries de les tesis doctorals

Search within this collection:  Browse by:    

Eines i recursos annexos a les tesis doctorals defensades a l'Institut de Lingüística Aplicada (IULA), descrits d'acord amb l'esquema de metadades proposat en el marc del projecte europeu CLARIN (Common Language Resources and Technologies).

## Recent Submissions

[Aplicació Lexicogràfica per als CConnectors Reformulatius: ALCOR](#)[AquaLexic](#)[Diccionari de sufixos verbalitzadors del català](#)[Interfície de consulta de la base de dades](#)[Ctxverbals](#)



## 4. The community. Results

<b>Tools and Resources</b>	<b>Primary data from PhD thesis</b>
<p>Fundamental characteristics:</p> <ul style="list-style-type: none"><li>• Set of atypical identification data</li><li>• Physical location in native servers</li><li>• Diffused identification of copyright</li></ul>	<p>Very different kind of objects: source code in Perl, HTML interface, Access database, etc.</p>
	<p>Dublin core schema is not well adapted to managed and recover the material</p>
	<p>Importance of the descriptive element to help understand the nature and scope of the digital objects</p>



## 4. The community. Results

### **Tools and Resources**

Fundamental characteristics:

- Set of atypical identification data
- Physical location in native servers
- Diffused identification of copyright



## 4. The community. Results

### **Primary data from PhD thesis**

Very different kind of objects: source code in Perl, HTML interface, Access database, etc.

Dublin core schema is not well adapted to managed and recover the material

Importance of the descriptive element to help understand the nature and scope of the digital objects



# Primary data in a institutional repository: a case study

## **Content**

1. Introduction
2. The CLARIN project
3. Institutional repositories and their contents
4. The community
5. Conclusions and future development



# 5. Conclusions and future development

## 5.1 Benefits

- making the data, resources and research tools visible and accessible

UNIVERSITAT POMPEU FABRA

Institut Universitari de Lingüística Aplicada

PRESENTACIÓ FORMACIÓ RECERCA PUBLICACIONS

Recursos IULA

- Corpus i Eines
- Dialectometria
- Gestió del coneixement
- Lexicografia
- Lingüística forense
- Lingüística i logopèdia
- Uengua oral
- Neologia
- Terminologia
- Variació lingüística

1

UNIVERSITAT POMPEU FABRA REPOSITORI DIGITAL DE LA UPF e-Repositori

Web BibTIC Castellano English Ajuda El meu compte

Inici

Cerca

Cerca avançada

Lista per:

- Tot l'e-Repositori
  - Comunitats i col·leccions
  - Títols
  - Autors
  - Matèries
  - Per data de publicació

El meu compte

- Entra

Què és l'e-Repositori ?

Estadístiques

SHERPA RoMEO DULCINEA

Benvinguts al Repositori Digital de la UPF

El repositori institucional recull, difon i preserva la producció intel·lectual en format digital que resulta de l'activitat acadèmica i investigadora de la UPF, les revistes científiques i les publicacions institucionals. La seva finalitat és contribuir a augmentar l'impacte de la investigació feta a la UPF i preservar la memòria intel·lectual de la UPF.

Cerca

Introduïu algun mot en la casella de sota per fer una cerca.

Comunitats

Seleccioneu una comunitat per visualitzar-ne les col·leccions.

- Docència [37]
- Recerca: articles, congressos, llibres [30]
- Recerca: tesis [540]
- Recerca: working papers, PFC, etc. [1.348]
- Recursos i dades primàries [42]
- Revistes científiques [2.751]
- Continguts institucionals [84]
- Vida universitària [3]

Avis legal | Nota tècnica | Contacte

RSS 1.0 - RSS 2.0

2





# 5. Conclusions and future development

## 5.1 Benefits

- Handle identifier (supports the persistence and long-term accessibility of the data)

### AquaLexic

<b>Title:</b>	AquaLexic
<b>Author:</b>	Alonso, Araceli
<b>Description:</b>	Base de datos en Access con la finalidad de gestionar y realizar el tratamiento de los elementos principales de análisis (lemas, formas, ocurrencias, combinaciones (colocaciones), colocados y usos) de unidades léxicas nominales representativas del área del Medio Ambiente.
<b>Date:</b>	2010
<b>Citation:</b>	<a href="#">Export to Ref Works</a>
<b>To cite or link this document:</b>	<a href="http://hdl.handle.net/10230/6327">http://hdl.handle.net/10230/6327</a>
<b>Rights:</b>	Todos los derechos reservados.

# 5. Conclusions and future development

## 5.1 Benefits

- Can be harvested by other repositories

This search facility is a trial service that uses Google's [Custom Search Engine](#).  
Please send us [feedback](#) on your experiences.

To search for open access *repositories* rather than their content, please use the [Find](#) page.

Resultados 1 - 10 de **aqualexix** (0.20 segundos)

[AquaLexic](#)

dc.title, **AquaLexic**. dc.type, Dades. dc.subject.keyword, Unidades léxicas nominales.  
dc.subject.keyword, Terminología. dc.subject.keyword, Lexicografía ...  
[repositori.upf.edu/handle/10230/6327?show=full](https://repositori.upf.edu/handle/10230/6327?show=full)



[AquaLexic](#)

Title: **AquaLexic**. Author: Alonso, Araceli. Description: Base de datos en Access con la finalidad de gestionar y realizar el tratamiento de los elementos ...  
[repositori.upf.edu/handle/10230/6327](https://repositori.upf.edu/handle/10230/6327)

## 5. Conclusions and future development

### 5.2 Problems to solved

- adaptability to Dublin Core schema
- Considerations about copyright
- raw data formats

## 5. Conclusions and future

### 5.3 Steps ahead

- Review of repository policies
  - the deposit policy of digital materials
  - the regulations of the deposit of University dissertation projects and Doctoral Theses
  
- Implementation of different solutions
  - Exemple: federated thematic catalogues, databases and repositories at the institutional level



# Primary data in institutional repository

Universitat Pompeu Fabra

# Thank you very much !

**Silvia Arano**

[silvia.arano@upf.edu](mailto:silvia.arano@upf.edu)

IULA (Institut Universitari de Lingüística Aplicada)

**Gemma Martínez**

[gemma.martinez@upf.edu](mailto:gemma.martinez@upf.edu)

IULA (Institut Universitari de Lingüística Aplicada)

**Anna Casaldàliga**

[Anna.casaldaliga@upf.edu](mailto:Anna.casaldaliga@upf.edu)

Library Deputy Director

**Marina Losada**

[marina.losada@upf.edu](mailto:marina.losada@upf.edu)

Repository Manager

**Marta Villegas**

[marta.villegas@upf.edu](mailto:marta.villegas@upf.edu)

IULA (Institut Universitari de Lingüística Aplicada)

**Núria Bel**

[nuria.bel@upf.edu](mailto:nuria.bel@upf.edu)

IULA (Institut Universitari de Lingüística Aplicada)



