

# From CNR annual report to an Institutional Repository: which successful strategies?

Rosa Di Cesare, Roberta Ruggieri,  
Loredana Cerbara, Daniela Luzi  
Consiglio Nazionale delle Ricerche,  
Istituto di Ricerche sulla Popolazione  
e le Politiche Sociali

mailto {d.cesare, d.luzi, r.ruggieri,l.cerbara} @irpps.cnr.it

# Background of the study

In 2008 a CNR OA supporter group promoted a survey to acquire a precise picture of OA CNR practices

Main result of the survey was:

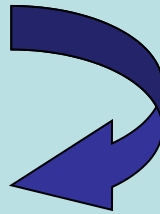
- Identification of the current electronic acquisition system of CNR scientific production as a building block for a future IR

CNR OA supporter group required CNR top management to sign the Berlin Declaration and start the development of a CNR's IR

Currently the Top management is developing an OAI-PMH compliant IR on the basis of the previous electronic acquisition system

# Aim

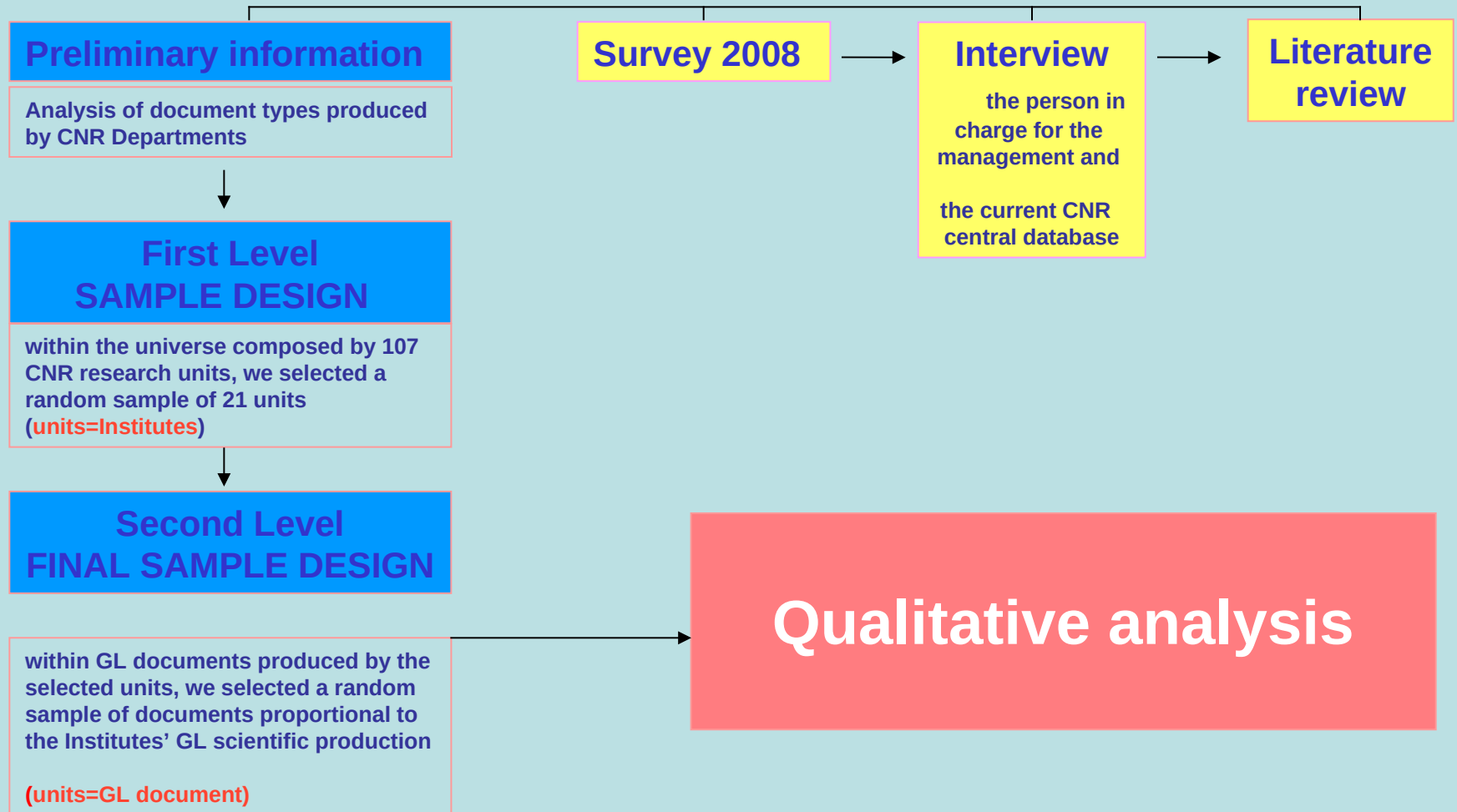
Analysis of the information content of the actual database  
(document types, quality of bibliographic descriptions)



To provide a contribution in:

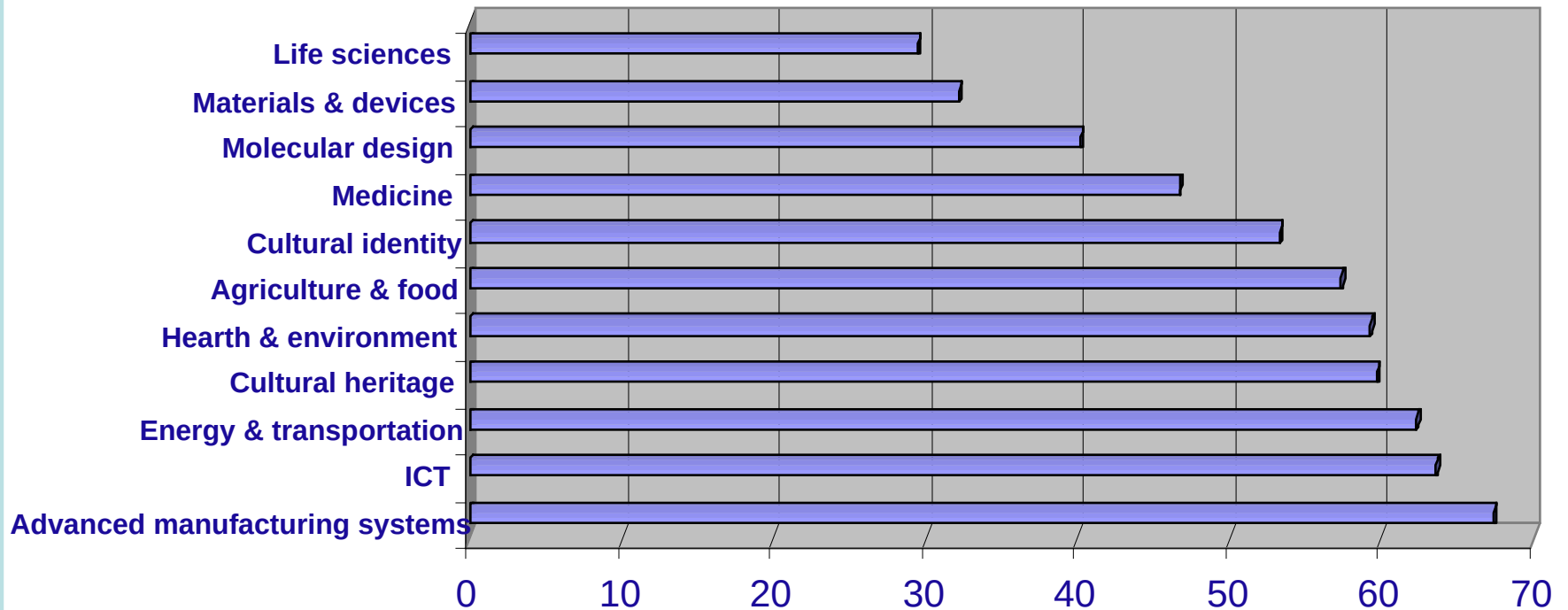
- Identification of metadata able to describe GL documents
- Development of guidelines focusing on the identification of GL bibliographic collections

# Survey design



# Profile of the universe

Percentage of GL documents by Department (2003-2007)



source: <http://www.cnr.it/istituti/Retescientifica.html/>

# Bibliographic collections by Departments (2003-2007)

<i>Department</i>	<i>Total number of publications</i>	<i>Journal articles</i>	<i>Books and book chapters</i>	<i>Conf papers</i>	<i>Oral presentations</i>	<i>Reports</i>	<i>In-house publications</i>	<i>Other</i>
Hearth & nvironment	16757	32,2	8,5	16,6	27,9	10,7	1,4	2,7
Energy and transportation	3529	35,6	2,0	27,7	18,1	13,6	0,3	2,7
Agriculture & food	6963	36,0	6,6	17,2	28,2	6,8	0,9	4,3
Medicine	8207	50,0	3,2	6,1	36,4	2,2	0,3	1,7
Life sciences	1980	67,7	2,8	1,4	25,9	0,9	1,2	0,2
Molecular design	10514	56,9	2,8	9,2	24,8	3,1	0,3	2,7
Materials & devices	15292	64,9	2,9	12,3	13,9	3,0	1,0	2,1
Advanced manufacturing systems	5019	29,2	3,3	19,8	15,0	19,5	1,2	11,9
ICT	7023	30,9	5,4	29,4	9,8	17,5	3,2	3,8
Cultural identity	9262	22,6	24,0	13,5	20,4	10,9	5,3	3,3
Cultural heritage	2154	21,8	18,4	30,8	15,0	7,8	1,5	4,6
<b>Total</b>	<b>86700</b>	<b>42,4</b>	<b>7,1</b>	<b>15,6</b>	<b>21,7</b>	<b>8,4</b>	<b>1,6</b>	<b>3,3</b>

source: <http://www.cnr.it/istituti/Retescientifica.html/>

# Sample: Institutes and documents

<i>Department</i>	<i>Research units</i>	<i>Sampled of research units</i>	<i>Total number of documents produced</i>	<i>Total number of GL documents</i>	<i>GL Sampled documents</i>	<i>GL examined (pre-text)</i>
Hearth & environment	13	3	3295	1985	114	53
Materials & devices	12	3	6550	1064	60	34
Molecular design	14	3	2072	1059	59	43
Cultural identity	15	3	1945	952	51	57
Medicine	12	2	1784	983	54	24
ICT	7	1	1010	558	31	19
Agriculture & food	10	2	2375	1342	73	38
Advanced manufacturing systems	7	1	631	415	24	14
Energy and transportation	6	1	431	262	14	8
Cultural heritage	5	1	464	281	15	15
Life sciences	6	1	74	36	2	6
<b>Total</b>	<b>107</b>	<b>21</b>	<b>20631</b>	<b>8937</b>	<b>497</b>	<b>311</b>

## GL documents examined:

Data was collected directly from primary source;

Data analysed: publications produced in the period 2003-2007;

Data was gathered in September 2009

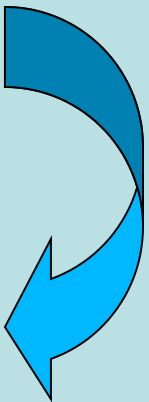
We selected following GL document types:

- conference papers
- oral presentations
- reports
- In-house publications

## Qualitative analysis considers:

- metadata used to describe each GL document type
- use of note field and analysis of its content

- Verify whether the actual metadata are suitable to describe GL document types
- Verify whether the note field is used to increase the information consistency of the bibliographic record





# Conference papers

## Metadata

### Mandatory fields:

- Conference types:
  - International or national,
  - ISI indexed
  - Invited / Invited & ISI
  - Refereed
- Author & institution information:
  - author affiliation,
  - author's research project,
  - author ID

### Optional fields:

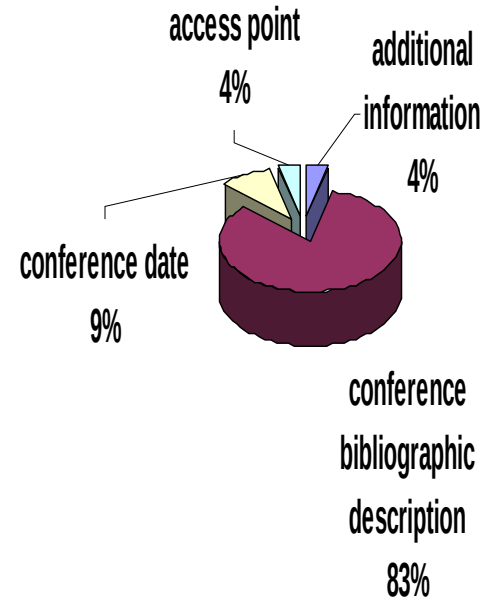
- Conference title and location
- Subject categories

### Missing fields:

- Conference date
- Conference bibliographic description (title, editors, publisher, ect.)

## Note field Content

26% of analyzed documents have the note field compiled



# Oral presentations

## Metadata

### Mandatory fields:

- Item types:
  - international or national
  - abstract/poster,
  - communication/relation,
  - invited
- Author & institution information:
  - author affiliation,
  - author research project,
  - author ID

### Optional fields :

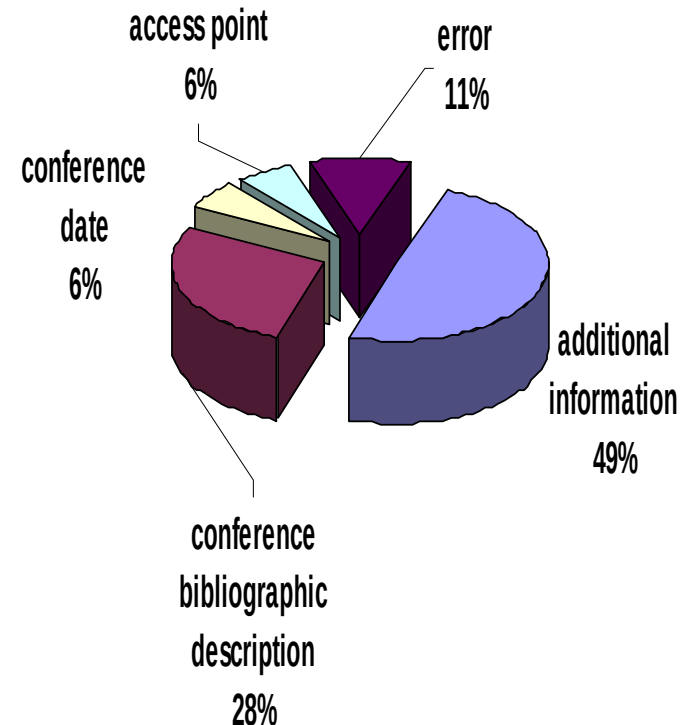
- Conference title and location
- Subject categories

### Missing fields:

- Conference date
- Conference bibliographic description (title, editors, publisher, ect.)

## Note field Content

20% of analyzed documents have the note field compiled



# Reports

## Metadata

### Mandatory fields:

- Item types:
  - technical reports, project reports,
  - guideline/manual,
  - multimedia products,
  - databases,
- Author & institution information:
  - author affiliation,
  - author research project,
  - author ID
- Format: cd-rom, floppy disk ect.

### Optional fields

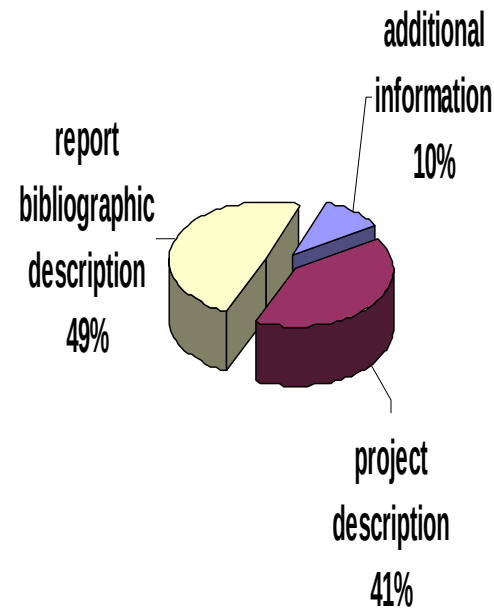
- Subject categories

### Missing fields:

- Report series and Report number
- Project description (project title , project number, contract number)

## Note field Content

52% of analyzed documents have the note field compiled



# In-house publications

Collection of documents published by CNR Units: high variety of document types

## Metadata

### Mandatory fields:

- Document types:
  - monograph & serials
  - catalogue & databases
  - excerpt
  - supplement ...
- Author & institution information:
  - author affiliation,
  - author research project,
  - author ID

### Optional fields

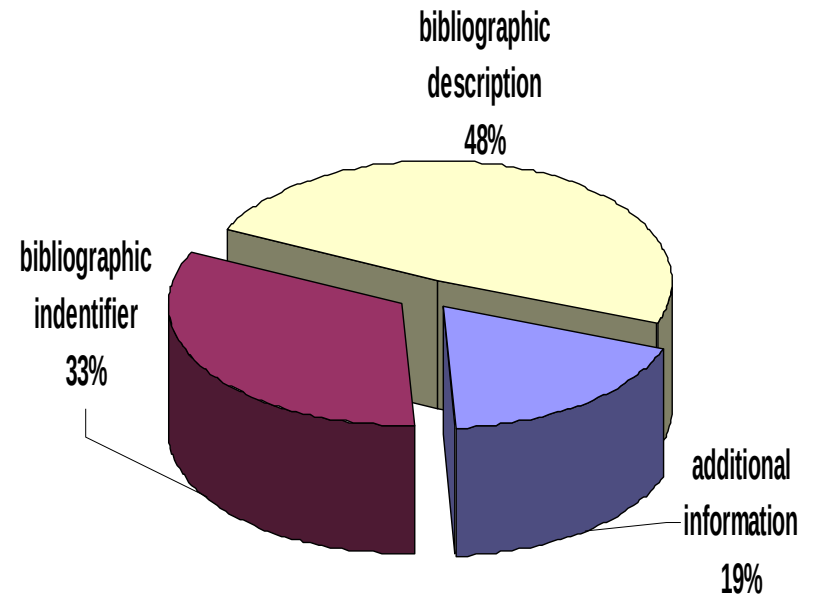
- Bibliographic references
- Subject categories

### Missing fields:

- No specific bibliographic description fields related to document types

## Note field Content

47% analyzed documents have the note field compiled



# First remarks

## Background of current document acquisition system

Designed within a set of databases oriented to collect managerial data containing also research outputs (i.e. CNR Annual report)

*Mandatory and daily* used by all CNR Research units: it now contains the research outputs starting from 2002 (= more than 90.000)

## Characteristics of the current document acquisition system:

Presence of a core set of metadata of the identified document types

Integration with other internal databases (i.e. authors IDs, CNR Projects, ...)

Provision of disciplinary categories & keywords

Special emphasis on data related to evaluation (ISI, invited presentations, ...)

Missing metadata for a correct GL document type description

# Conclusions and future works

- **High** percentage of the use of the Note field to describe GL documents
- **Higher** percentage for:
  - *Greyer* documents
  - GL documents described within no well defined collections
- The **highest** percentage refers to Reports where Notes supplement lacking metadata

Does homogenous description reported in the Notes depend on librarians inserting data?

Survey results are a contribution to the identification and organisation of collections - in particular GL - and are the basis of the development of guidelines that enhance data quality as well as help information providers to submit research outputs