

# MAPPING ITALIAN GREY COMMUNITIES. WHAT IS THERE BEYOND THE ACADEMY?

---

Silvia Giannini, Stefania Biagioni, CNR-ISTI, Pisa Italy

Sara Goggi, Gabriella Pardelli, CNR-ILC, Pisa Italy



# Summary

## The survey examines Italian Grey Communities beyond the Academy

- Objective
  - Creation of a map of non-academic communities and stakeholders involved in the management of grey material
- Method
  - Focused Google queries and acquisition of results
  - Building of the Corpus
    - Analysis of documentation and categorization (whom, what, how...)
  - Analysis of results and production of taxonomies
- Conclusions and Remarks



# The method: search and retrieve

## Focused Google queries and acquisition of results

- Selection of the web portals as resulted from the query:  
"letteratura grigia" OR "letteratura non convenzionale" OR  
"documentazione grigia" OR "materiale grigio"

## Google results

- a. Number of indexed documents considered as "interesting" with respect to the terms of the query (about 42000)
- b. List ordered by relevance of URL (never more than 1000)
- c. 800 URLs retrieved for building the sample (Corpus)



# On the scene



Several different communities involved in the management of grey material, each with quite different

- Institutional nature
  - public and private, ecclesiastic ....
- Fields to which these grey communities belong
  - cultural & social/welfare interest, healthcare, research, education and beyond

Several types of “grey material”, each with quite different

- Formats and dissemination
  - various types of documents published in pdf format or simple descriptions on web sites, thematic bibliographies, newspaper articles, multimedia, catalogues, repositories
- Organization, representation, storing, preservation ...



# The method: categorization

The Corpus is built by grouping the gathered data in 12 informative classes assigned to each URL

Classes	Example
Stakeholder	Association
Type of Stakeholder	Social Centre
Name of Stakeholder	“Lilith - Archivi delle donne e del femminismo”
Infrastructure/service	Archive
Field	Social
Topic	Feminism
Type of material	Leaflets, bibliographies, private correspondence, manuscripts, minutes ...
Format	html
Access	free
Domain	it
Description	GL appears in the description of the archive property



# Analysis of the data

How do we behave when facing a classified list of results?

- The Corpus has been “cleaned up” based on the relevance of content:



- exclusion of academic web sites, web sites advertising/selling products of grey colour, urls duplicates...
- The data have been ordered on the bases of the stakeholders' nature:
  - e.g. public, private, business, personal ...
- The communities have been collected on the bases of similar features and content:
  - e.g. childhood, music cinema and theatre, women, welfare, professional associations, healthcare ...

Which criteria should be used for assembling the communities?

- subjects of relevant social interest
- richness of content
- organisation, representation, storing, preservation of data/documents
- ...



# Stakeholders

Who is most interested in GL in Italy beyond the Academy?

There are stakeholders of public, private and hybrid nature

- The public sector discloses information homogeneously because it should respond to institutional standards as far as the treatment of data is concerned:
  - State entities and agencies
  - Public Institutions/Ministries/Regions/Municipalities
  - Clerical entities
  - Research centres
  - International agencies
- The private sector discloses information in a heterogeneous way according to its own purposes:
  - Companies/corporations
  - Public and personal blogs
  - Personal homepages
  - Associations/organizations/foundations
- The hybrid sector (public/private)
  - Associations/foundations
  - Spin-offs



# Communities

Significant examples have been identified on the web sites presenting a heterogeneous set of material which could prove to be especially interesting to researchers, scientists, professionals and simple fans and lovers of the various subjects if ever made available and usable.

Environment

Women

Health

History

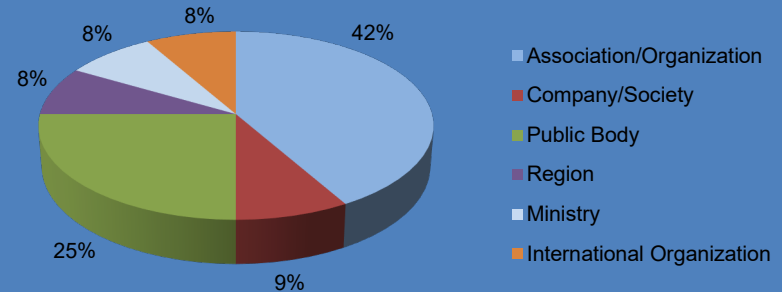
Several communities handling and disclosing “grey” documentation in various ways come to light from this analysis. The ensemble of information collected allowed to identify and delimit 4 illustrative communities.



# Communities - Stakeholders

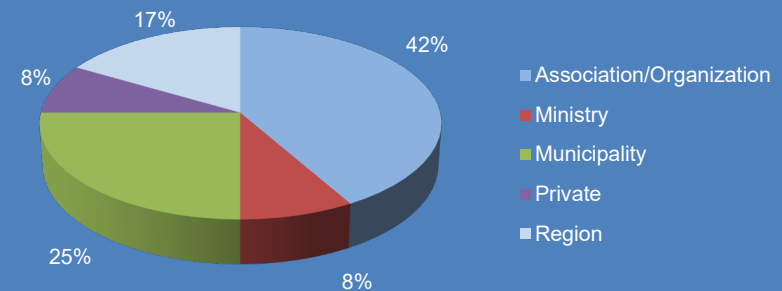
## ENVIRONMENT

- Sustainable cities
- Geomorphology
- Mediterranean sea
- Aero-spatial research
- Mussels' production chain
- ....



## WOMEN

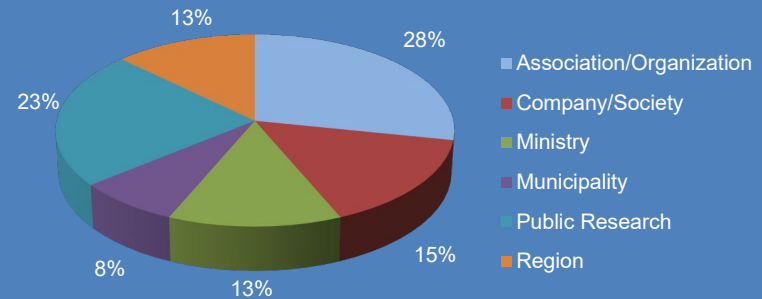
- Archives of the history of women
- Voices, visions and actions of women
- Violence on women
- Archives of feminism
- ....



# Communities - Stakeholders

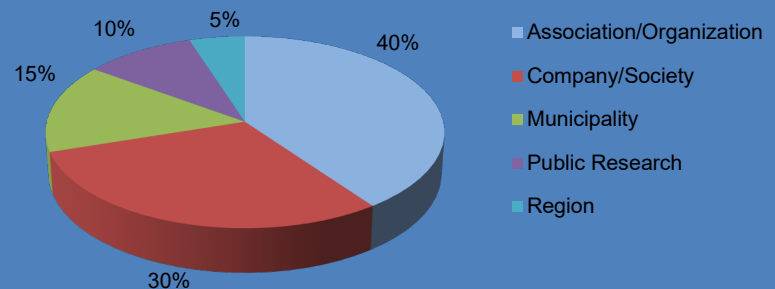
## HEALTHCARE

- Drug dependences, mental diseases and youth disorders
- Work, safety and social politics
- Alcohol-related pathologies
- Eating disorders
- ....



## HISTORY

- History of the Church
- History of the Italian Risorgimento
- Archives for the history of Education
- Historical archives of political parties
- ....



# Conclusions & Remarks

Using Google-search as a source for building the GL Corpus has been useful for evaluating the state-of-the-art ...

- Disadvantages
  - There is an “explosion” of information ... who is producing it? a Person, a Society, a Region, a Foundation?
  - Pages containing large amounts of unrelated material (downloads, graphics, non-pertinent material ...)
  - Many web sites cannot be accessed / change their URLs / are not updated / information is either cancelled or duplicated
  - It is impossible to have more than 1000 results
- Advantages
  - Google finds the unobtainable...
  - Google is a outstanding source of information and of access to data
  - Google allows to easily download the retrieved information



... and starting sketching a map of communities and stakeholders

# Conclusions & Remarks

The processing has been time-consuming and little fulfilling:

- there is a big amount of disorganised and badly-structured material
- there are many web pages with access denied
- the term “letteratura grigia” often appears simply as a description of the material but the content is not available (e.g. catalogues, lists of references, glossaries, encyclopaedia entries ...)

There is a great variety of non-academic communities, each of them having a different approach to GL material:

- grey documentation does not have any frontier and belongs to the most diverse fields
- grey literature is more miscellaneous and more variously represented

The Web helps in disclosing what is available as well as in pointing out the difficulties which the communities still face in organising the grey material