



A knowledge model for supporting digital exploration of historical industrial archives

Monica DeMartino, Marina Monti, Chiara Rosati

Institute of Applied Mathematics and Information Technology "Enrico Magenes", CNR



URLA 2014 Istanbul 17-20 September 2014

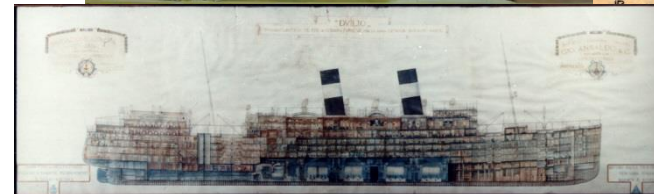
□ Project R.I.C.E.R.C.A. 2013-2015

“Risorse Industriali Culturali: Estrazione e Raccolta dei Contenuti per Accessibilità”, Italian project funded by European Commission

□ Activity objective

□ Approach

□ Outcomes





Project R.I.C.E.R.C.A.

□ OBJECTIVE:

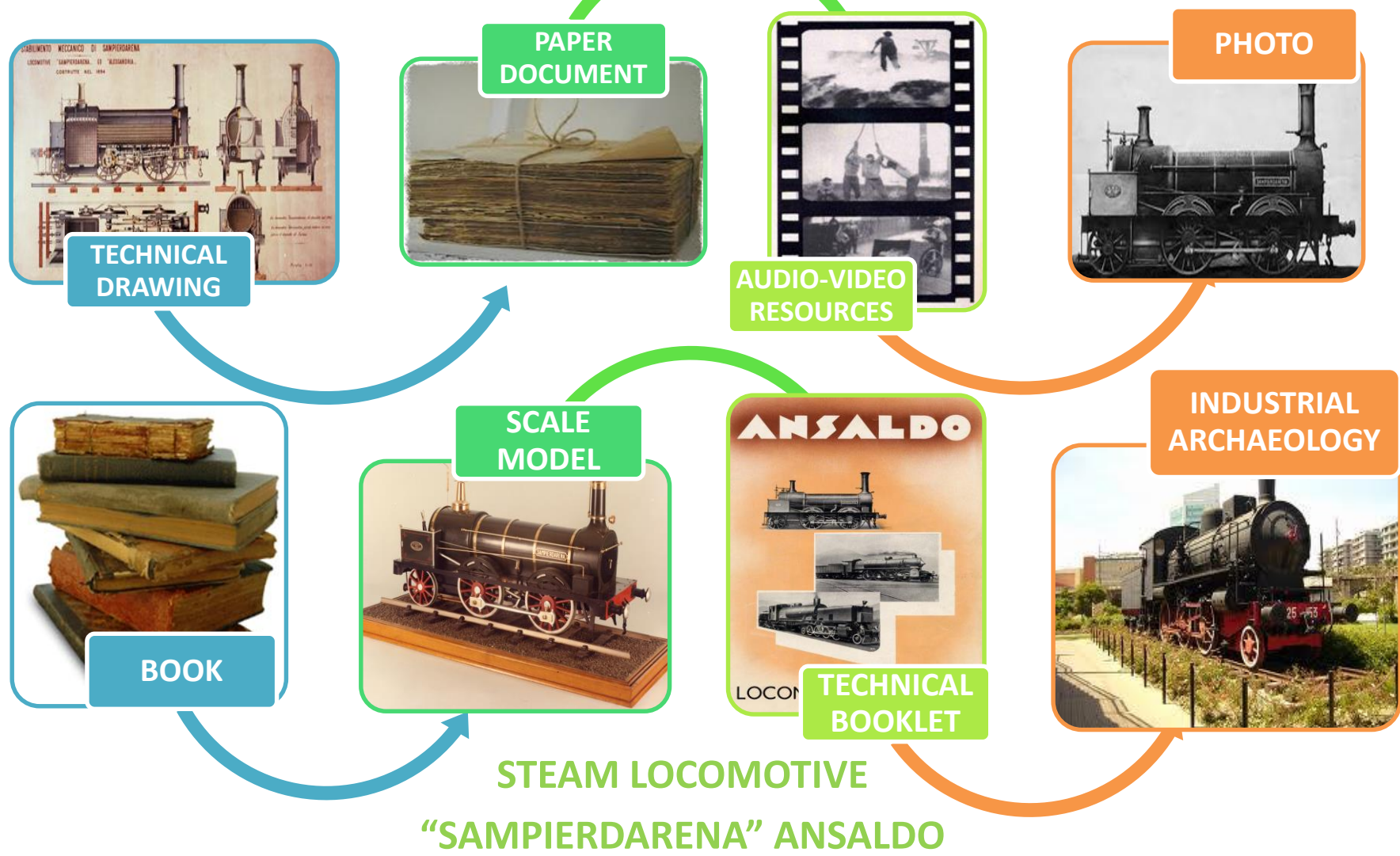
To improve the capacity of cataloguing, accessing and searching in Ansaldo Foundation's Historical Business Archives

- Define the Archive management model
 - *Resource preservation*
 - *Archival documents description*
 - *Archive enhancement*

- Design a web platform for archives exploration
 - *Semantic information*
 - *Interactive navigation tools*

- Make resources available and easily accessible
 - *Heterogeneous users*
 - *Different targets*

Semantics exploitation





Ansaldo Foundation



MISSION

- Preserve and enhance the Historical Business Archive produced by industrial-economic subjects of Genova and Liguria in XIX-XX century

ARCHIVAL HERITAGE

- Over 15 km. of archival documentation, 400.000 original photos, 5000 original movies, industrial artefacts collections, Library specialized in industrial-economic issues, an original photographic laboratory

HISTORY

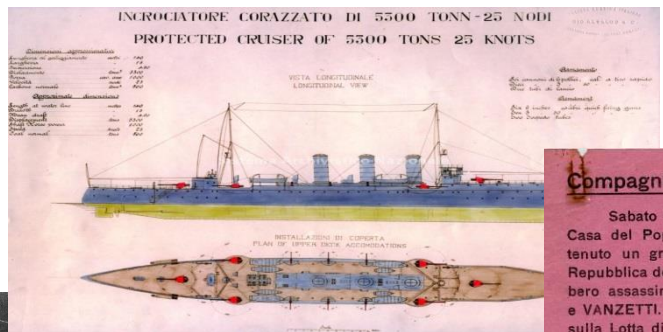
- Founded in 1980 as Archivio Storico Ansaldo, in 2000 became Ansaldo Foundation and in 2007 Ansaldo Foundation - Finmeccanica Group

PARTNERSHIP



URLA 2014 Istanbul 17-20 September 2014

Photo



Paper

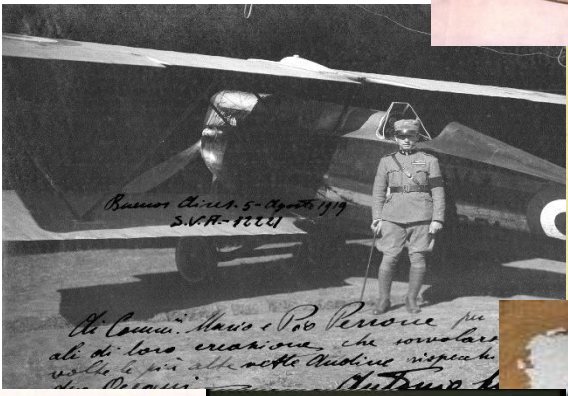


Compagni Lavoratori!

Sabato sera alle ore 20,30 nell'atrio della Casa del Popolo (già Palazzo d'Istruzione) sarà tenuto un grande Comizio di Protesta contro la Repubblica del Dollaro che il 1° Novembre ha osato assassinare i due nostri Compagni e VANZETTI. - Farleranno oratori di tutti sulla Lotta di Classe.

Partito Comunista d'Italia. - Partito Socialista Federale. Giovanile Comunista - Unione Anarchica. Camera del Lavoro Confed. - Camera del Lavoro.

Per entrare al Comizio occorre presentarsi sabato sera ad una delle Organizzazioni.



Book



ALDO-FOSSATI GENOVA - SESTRI ITALIA



Movie

Industrial Artifacts

Management of “Documentation produced or acquired by creators (corporate body, family, person) during the development of its activity, destined to permanent preservation to ensure public consultation”

□ STRENGTHS

- Evidence of the economic-social history of a country and valued sources for historical research
- Important role for corporate marketing, economic growth and strengthening of company public image

□ MANAGEMENT

- Appropriate methods to manage and preserve different media types
- Appropriate metadata profile to describe different document types
- Archival-historical analysis and contents contextualization
- Company communication policies/copyright/trade secret
- National legislation about consultability right/privacy protection



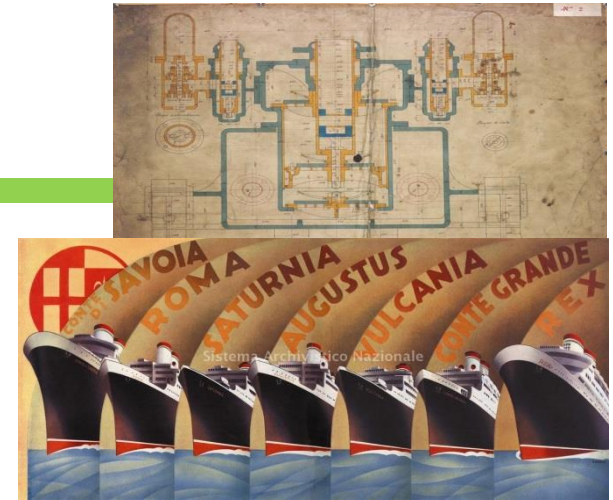


Industrial Archives

Historical documentation produced by companies, families or persons carrying out an economic activity in a specific industrial sector

□ FEATURES

- Different from other archives type (churches, municipal, political parties archives)
- Similar business activity (mechanical or metallurgy industry sector)
- Similar creator's typology (industrial company, owner, manager)
- Similar documentation types (social scriptures, accounting record, technical document)
- Influence of document production context



Historical period
National legislation
Industrial sector management

□ SPECIFICITY

High level of heterogeneity

Media
types

Text document
photo & movie
industrial artefact
book

Material
types

Paper, glass,
film, iron, wood

Document
types

Technical drawing
map & plant
company magazine

Content of
documentation



Challenges

□ HETEROGENEOUS DATA

Huge amount of documents
Heterogeneous document and media types
Content and knowledge to extract and contextualize

□ ARCHIVAL DESCRIPTION

- Standard metadata is insufficient to provide appropriate document and content description
- Need to be uniformed to facilitate resources access and make explicit relation among document and content

□ ARCHIVE ACCESS AND DATA SEARCH

Archive content understanding might be difficult for a wide range of users with diverse competence

- Different cultural background and skill levels

Professional users (historian, archivist)
Common users (citizen, company employee, student)

- Multicultural/Multilingual issue



Activity objective

Provide appropriate description of the archival resources

□ KNOWLEDGE MODEL

Definition of a knowledge model for historical industrial archives

Embedding information on archive content and structure

□ OBJECTIVE

- Provide common understanding of the archives content
- Support the diverse user's communities in the exploration of the archive and its contents by exploiting semantics

□ RESULT

Formalization of ontology able to support advanced tools for archival data searching



The methodological approach

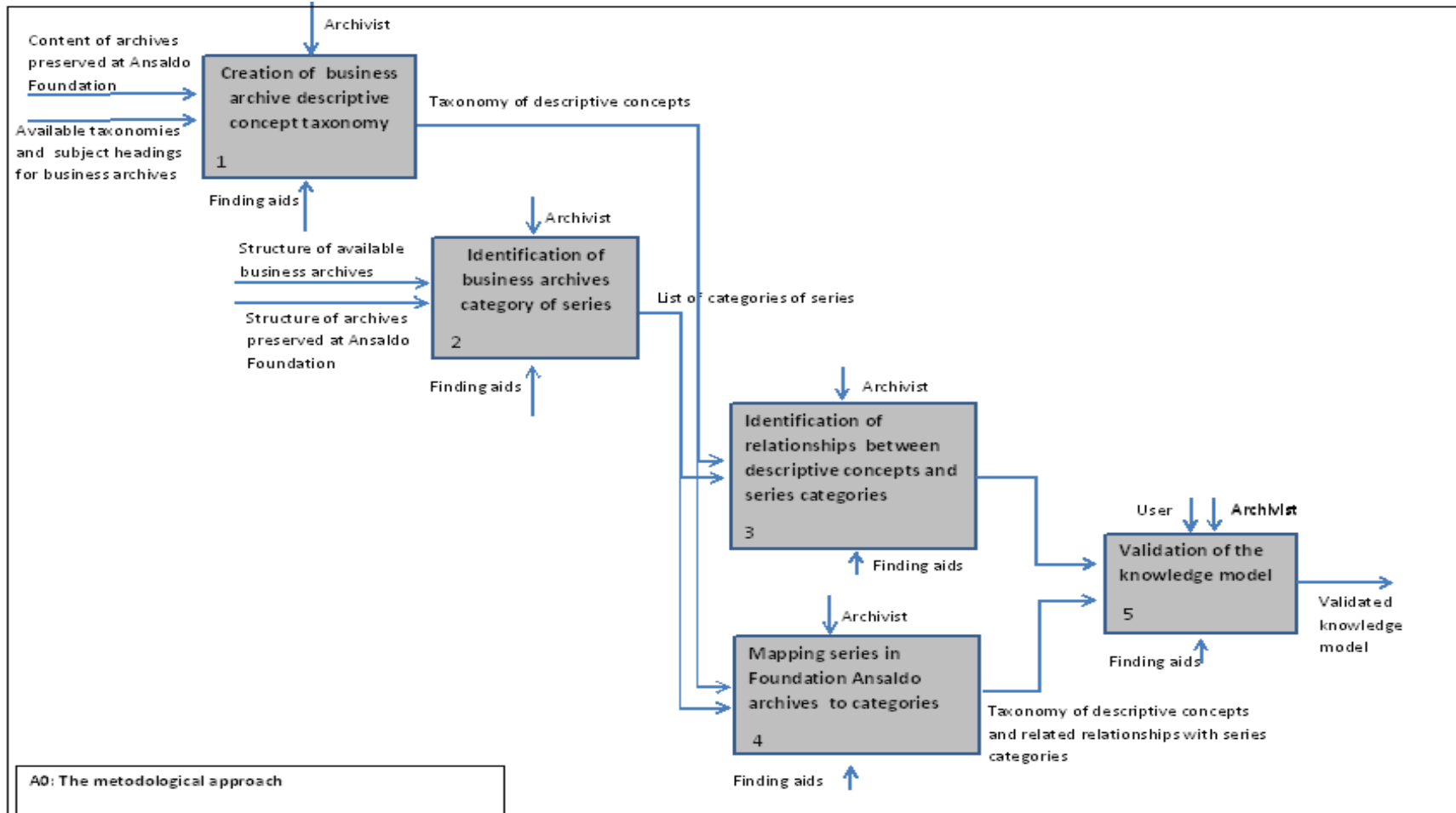
□ ANALYSIS OF THE STATE OF PLAY

- Existing standards metadata for archival description
- Knowledge organization systems available in Business archives domain

□ DEFINITION OF KNOWLEDGE MODEL FOR INDUSTRIAL ARCHIVES

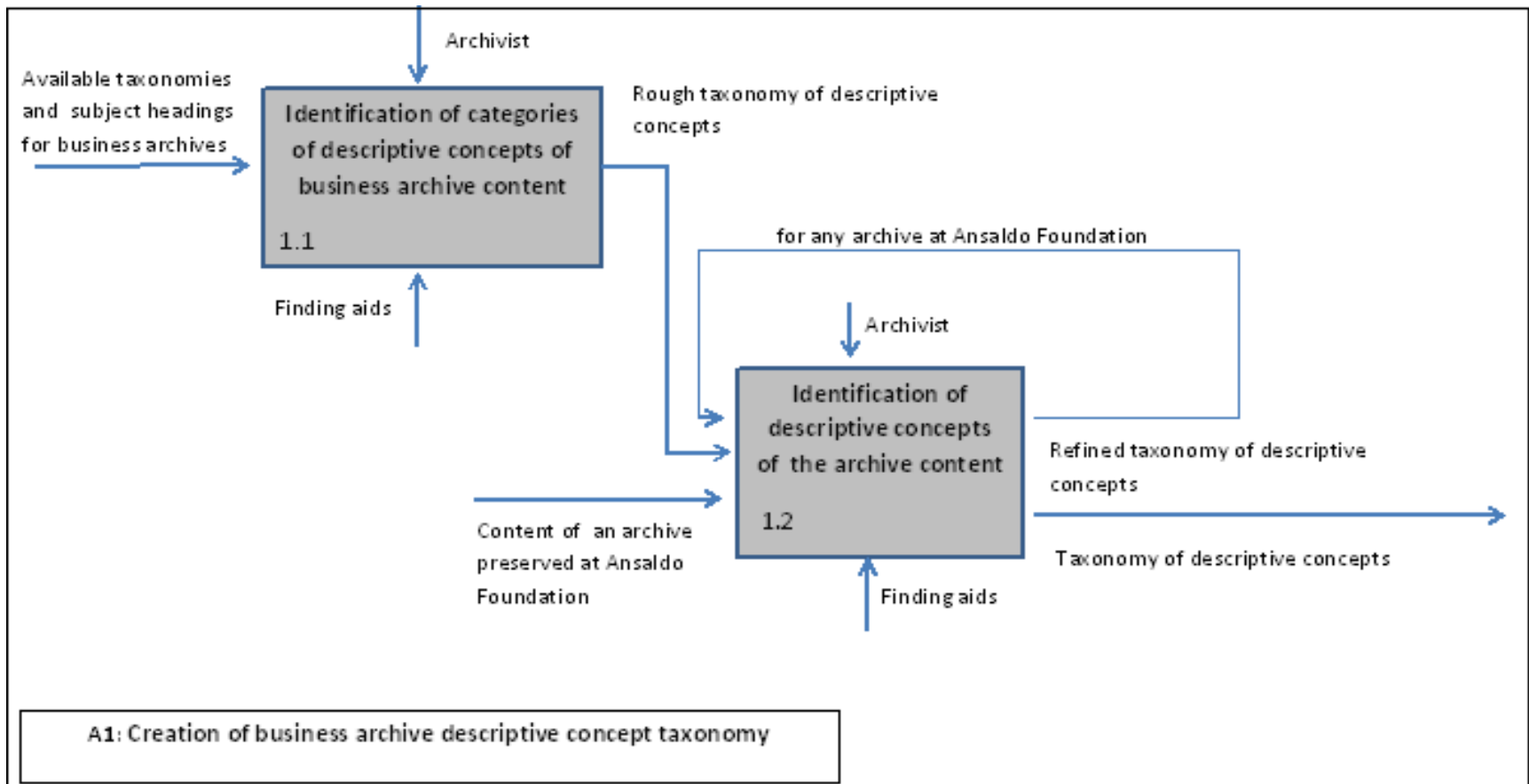
- 1 Identification of taxonomy of content descriptive concepts**
 - 2 Identification of Industrial Archives categories of series**
 - 3 Identification of relationships among concepts and series**
 - 4 Mapping of the existing series with the series categories identified**
 - 5 Validation of the knowledge model**
-

The methodological approach



The methodological approach

1 Identification of taxonomy of content descriptive concepts



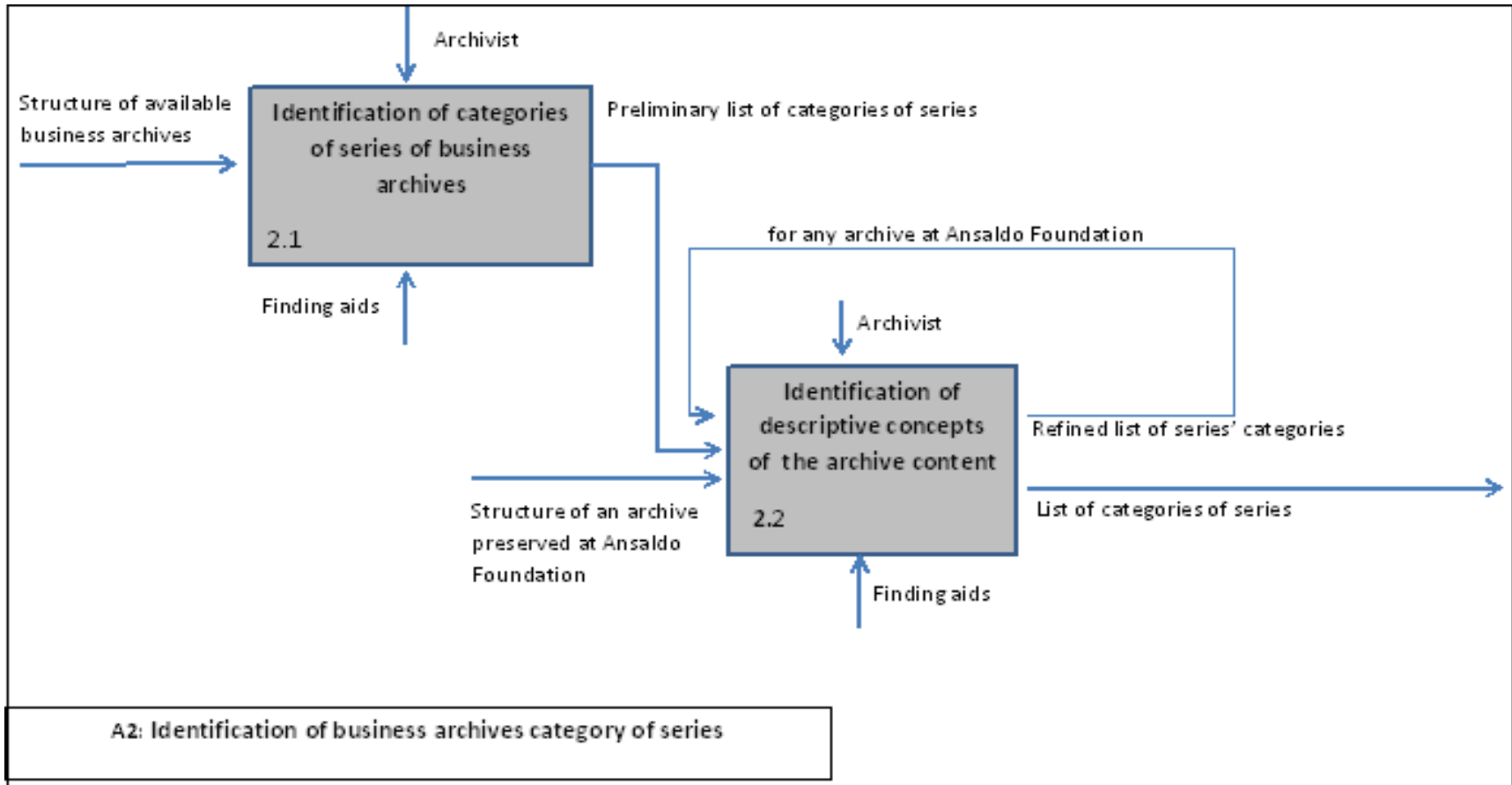


Identification of taxonomy of descriptive concepts

- ❑ Identification of preliminary list of descriptive concepts recognized as significant in Business archives domain
 - Archivist expertise and cultural background in BA management
- ❑ Validation of the effective significance of the identified categories of descriptive concepts
 - Subject headings
 - Outcomes of projects at national and international level
- ❑ Preliminary list is refined and concepts are added to adequately cover the content description
 - Analysis of any Ansaldo Foundation archives and its finding aids
- ❑ **RESULT:** Identification of taxonomy of descriptive concepts recognized as common and relevant in the content description of Business Archive documentation

The methodological approach

2 Identification of Industrial Archives categories of series





Identification of categories of series

- ❑ Identification of a preliminary list of series categories recognized as significant in Industrial archives domain
 - Archivist expertise and cultural background in Business archives management

 - ❑ Validation of the effective significance of the identified categories
 - Finding aids compiled by archives creators and institutions with archival holdings dedicated to Business archives management

 - ❑ Preliminary list is refined and series are added
 - Analysis of the existing series in any Ansaldo Foundation archives and its finding aids

 - ❑ **RESULT:** Identification of series categories recognized as common and relevant in the industrial archives structure
-



The methodological approach

3 Identification of relationships between descriptive concepts and categories of series

- Association of one/several series categories to each descriptive concept identified based on the documents actually preserved in each series

The criterion is based on the Archival science practice

- Series contain specific document types reflecting specific sector of the company management

Example

Series *Accounting records* preserves documents about the company accountancy administration and not technical documents about industrial product design or personnel management



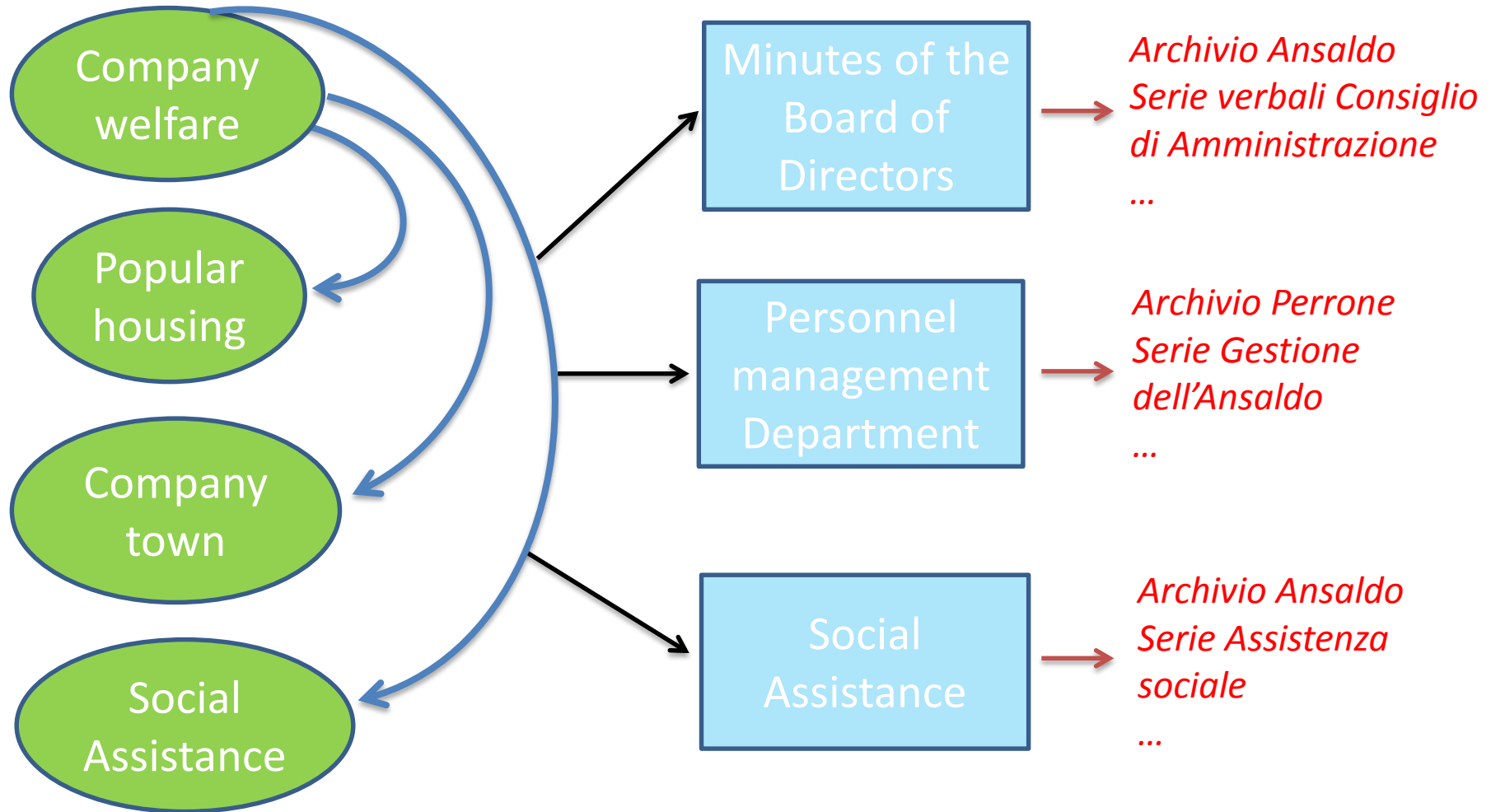
The methodological approach

4 Mapping of the existing series with the series categories identified

- Each existing series of each archive preserved at Ansaldo Foundation is linked with one of the identified series category according to its real content

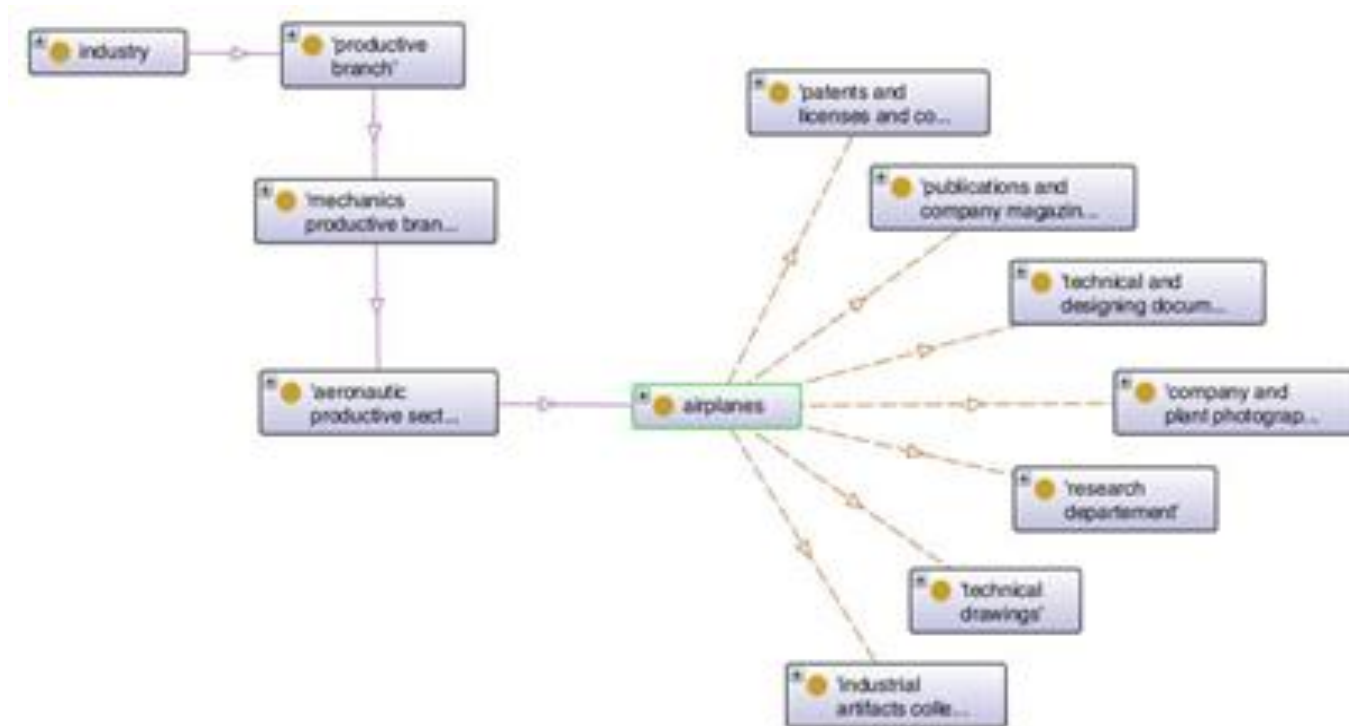
5 Validation of the knowledge model

- Simulation of the search for documentation based on the research topics emerged as significant in the use cases analysis conducted on the common users of Ansaldo Foundation archives
-



□ THE ONTOLOGY MODEL

The knowledge model has been formalized in ontology



□ ONTOLOGY MODEL

- Ontology is defined according to OWL 2.0 language
- Ontology editor Protegé 4.3 is used to create the classes, restrictions and properties

□ ONTOLOGY FORMALIZES

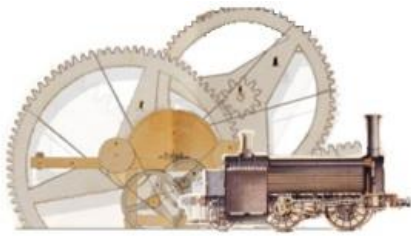
- 330 items in the taxonomy of descriptive concepts
- 210 items in the taxonomy of series categories
- 110 instances of Ansaldo Foundation archives

Descriptive concepts have been associated with *series categories* and with *instances* relative to each particular archive, making explicit the relationships between concepts and series



Conclusions

- ❑ The defined ontology model is used as knowledge base of a semantic visual exploration tool allowing effective and easy exploration of Ansaldo Foundation archives
- ❑ The knowledge model is specifically addressed to Ansaldo Foundation archives, but the methodology used is applicable to the definition of knowledge model related to other types of Business archives



R.I.C.E.R.C.A

THANKS!

<http://ricerca.ge.imati.cnr.it>
www.fondazioneansaldo.it

 **imati**

 **FondazioneAnsaldo**

softeco
sismat
information technology

All above pictures reproduced courtesy of Fondazione Ansaldo. Copyright © All rights reserved

URLA 2014 Istanbul 17-20 September 2014