

# An Innovative Content Repository to Explore Aerospace Documents based on Document Warehousing Technology

ASSUNTINA CEMBALO, MICHELE FERRUCCI  
FRANCESCA MARIA PISANO, ROSA SANNINO, MAFALDA VIOLA\*

**ABSTRACT:** Italian research centres, and mainly SMES, have to face with the difficulty to manage their Institutional Repository (IR), so the SIA (Aerospace Information System) project focused on creating an innovative content repository where real-time particularly interesting information pertaining to the evolution of technology and to applied research for the aerospace community, can be found. The SIA project main aim is to help the aerospace community (made by engineers, scientists, managers, stakeholders, authorities, and so on) to capitalize their outcomes and facilitate collaboration through the sharing of information content, not always easily found. The idea developed in this paper concerns the enhancement of a content repository using Document Warehousing as a support tool in the exploration of documents (e-books, e-journals, etc.) from heterogeneous sources, all related to aeronautics and aerospace domains, by performing OLAP queries. The challenges of cleaning and integrating document information content in order to offer easy navigation are investigated and discussed, focusing on the issues of removing ambiguities from documents. Complex information objects and heterogeneous resources are handled through the innovative SIA content repository, created for the aerospace community and users, so that they can take advantage of interconnected information, interacting with a multiplicity of other users and, consequently, feeling part of a community.

**Keywords:** Document Warehousing, On-Line Analytical Processing, Extraction-Transformation-Loading, Data Integration, Aeronautics and Aerospace domains.

## 1. Introduction

How hard it is to find useful information disseminated in documents, digital libraries, digital folders, websites, blogs, etc., is a common daily experience, whatever the topic of interest is. The challenge is even harder when scientific and technological concerns are involved, because information sharing

\* Assuntina CEMBALO, Michele FERRUCCI, Francesca Maria PISANO, Rosa SANNINO – C.I.R.A., Italian Aerospace Research Center, via Maiorise snc, 81043, Capua (CE), Italy.  
a.cembalo, m.ferrucci, fpisano, r.sannino @cira.it  
Mafalda VIOLA – Consultant at C.I.R.A., Italian Aerospace Research Center, via Maiorise snc, 81043, Capua (CE), Italy.