

# Table of Contents

## Introduction



- Field of research, focus (research products, DH and Italian CH as case study), state of the art, gap and research questions.
- ATLAS project overview: goals; methodologies; workflow; deliverables.
- Contents of the white book.

## Abstract

In the last twenty years, numerous Web-based Digital Humanities projects have emerged to collect data in the field of Italian Cultural Heritage. However, the research outputs produced by these projects risk obsolescence if not well-documented and based on shared standards and best practices. Moreover, these research products are often difficult to discover, and currently, no dedicated comprehensive catalogue exists. The ATLAS project—a joint effort of the University of Bologna, the Ca' Foscari University of Venice, and the CNR Institute of Computational Linguistics—aims to fill this gap. It seeks to create a catalogue that leverages Semantic Web technologies to interlink scholarly outputs and enhance the Italian Cultural Digital Heritage. The main expected outcomes of the ATLAS project are:

- a knowledge graph representative of Italian heritage, accessible online via dedicated services and preserved in CLARIN;
- a web application that allows users to explore the catalogued scholarly products and their contents, starting from the pool of selected research products used as pilots for the project;
- a whitebook for the curators of the catalogue that also analyses the state of the art and offers a set of guidelines and best practices to produce FAIR

(findable, accessible, interoperable and reusable) and high quality scholarly data.

This is the first version of the ATLAS whitebook, which focuses on the initial outcomes of the project: the results of the pilot evaluation, complemented by a set of guidelines for producing FAIR data (Chapter 1), and the metadata for the catalographic description of research products (Chapter 2). The refined version of the whitebook (planned for M23) will present the knowledge extraction systems, the ATLAS web application, and the web services available for browsing the catalogue (Chapter 3).

## Chapter 1: pilot analysis



- Selection criteria and limitations
- Analysis criteria
- Analysis results
  - Text collections
  - Digital scholarly editions
  - Softwares
  - Linked open data
  - Ontologies

### Abstract

The ATLAS project began with evaluating a selected pool of research products. These scholarly products were incorporated into the knowledge graph and used to identify metadata and strategies for accessing and manipulating data from various digital artifacts. The pilots fall into five main categories:

1. collections of text documents ([ALIM](#), the Archive of the Italian Latinity of the Middle Ages; [Biblioteca italiana](#); [Musisque Deoque](#); [BUP - Digital Humanities](#));

2. digital scholarly editions ([VaSto](#), [VArchI STOrIa fiorentina](#); [Codice Pelavicino Digitale](#); [Leges Langobardorum](#); [Digital Edition of Aldo Moro's works](#));
3. software solutions for text processing and visualisation ([EVT](#), [Edition Visualization Technology](#); [Voyant tools](#));
4. linked open data ([Zeri & LODE](#); [DanteSources](#); [LiLa - Linking Latin](#); [Biflow - Toscana Bilingue Catalogue](#));
5. ontologies for the Cultural Heritage ([CIDOC-CRM](#); [HiCO](#); [SPAR](#)).

The pilots listed above were selected based on two main criteria. First, they represent advanced Digital Humanities projects in the Italian Cultural Heritage domain, covering Italian and Latin sources and spanning from classical to contemporary periods. Second, for each pilot, at least one member of the project team has been directly involved in its development, providing access and in-depth knowledge.

The pilots were evaluated according to existing standards and best practices established in the field of Digital Humanities, as well as the FAIR principles. The results of this evaluation are presented in the first chapter of the whitebook, which is divided into five sections. Each section is dedicated to one of the types of digital artifacts listed above.

At the end of each section, we summarize the trends and practices that emerged from the pilots' analysis for that particular type of digital object. We then suggest how to best implement that kind of artifact. The analysis and recommendations focus on the FAIRness of the scholarly data, not on their effectiveness from a theoretical standpoint. For example, in the case of a digital scholarly edition, we do not evaluate the pertinence of the philological approach applied. The recommendations outlined in this chapter aim to fill the current lack of a comprehensive set of guidelines for producing FAIR scholarly data in the field of Digital Humanities. [To add value to this project outcome, the guidelines will also be published as a separate document.]

## Chapter 2: data model



- Overview: relationship with existing models
- Research project
- Research product (aggiungi le proprietà)
  - General properties
  - [Administrative properties]
  - Text Collection
  - Digital Scholarly Edition
  - Software
  - Linked Open Data
  - Ontology
- Mapping with OpenAIRE and Schema.org

## Abstract

The second chapter focuses on the data model of the ATLAS knowledge graph. This version of the whitebook presents the outcomes from the project's first phase: the catalographic metadata for describing various types of digital artifacts (identified through the pilots' evaluation), and the framework serving as the foundation for the data model.

The framework was constructed by comparing and mapping existing models, particularly Schema.org, RO-crate, DCAT, and OpenAIRE. It is based on two main concepts: `Research Project` and `Research Product`. The latter represents all types of digital objects produced by scholarly research projects. The `Research Product` class offers a suite of properties that help identify the digital object from both catalographic (e.g., `title`, `description`, `release date`, `version`, `licence`) and technical (e.g., `format`, `access points`) perspectives.

To distinguish different types of digital artifacts, from the `Research Product` have been derived five sub-classes: `Text Collection`, `Digital Scholarly Edition`, `Ontology`, `Linked Open Data`, and `Software`. Each sub-class expands on the `Research Product`

class, allowing for the description of specific types of research products with additional properties. The second chapter presents these classes and their related properties, illustrated with examples from the pilot research products.

Finally, the chapter concludes with a detailed description of how the ATLAS-DM relates to OpenAIRE and [Schema.org](https://www.schema.org/), including alignment tables and other relevant materials.

## Chapter 3: the DH-ATLAS catalogue



- Data entry and data access
  - Data entry
    - Editorial rules
    - Records creation
    - Infrastructural and technical note
    - Knowledge extraction
  - Data access
    - Interfaces: search, indexes, statistics (WIP)
    - API
    - SPARQL endpoint
    - data dump
- The catalogue
  - Statistics (?)
  - examples by type

### Abstract

The third and final chapter of the whitebook presents the ultimate outcome: the ATLAS knowledge graph and its associated services. This chapter serves as a

guide for potential users and curators of the catalogue. It begins with a set of guidelines for data entry, including instructions on creating new records and using the knowledge extraction system. Next, it details the data access services—namely, the GUI, API, SPARQL endpoint, and data dump. The chapter concludes with a comprehensive description of the catalogue.

In this version of the whitebook, the ATLAS platform and backend services are briefly introduced. The definitive version will provide a more detailed description of these components.