Universitatea
Transilvania
din Brașov

INTERDISCIPLINARY DOCTORAL SCHOOL

Faculty of Product and Environmental Design

Nicolaie CIUBOTARU (Constantinescu)

Systems and APIs for aggregating and presenting research data and metadata using Open Web Platform technologies

SUMMARY

Scientific supervisor

Prof. dr. eng., dr. marketing Angela REPANOVICI

BRASOV, 2023

Universitatea Transilvania din Brașov

Mr. (Ms.) ....

STRUCTURE
PhD committee
Appointed by order of the Rector of Transilvania University in Braşov
No. 12886 of 11.12.2023

Support Committee:

Prof. Dr. Eng. Codruța JALIU President, Dean of the Faculty of Product and Environmental Design, Transilvania University of Brașov

Prof. Dr. Eng., Dr. marketing Angela REPANOVICI Scientific manager, Transilvania University of Brașov

Assoc.Prof. Dr. Joumana BOUSTANY Official Lecturer, Universite Gustave Eiffel, Paris, France

Prof. Dr. Emeritus Mircea REGNEALĂ Official Referee, University of Bucharest

Prof. Dr. Eng. Anca DRĂGHICI Official Referee, Politehnica University Timișoara

Date, time and place of the public defense of the doctoral thesis: February 19, 2023, time....., room .....

At the same time, we invite you to take part in the public meeting to support the thesis of doctorate.

Thank you.

Introduction

The technologies that are used to achieve communication between digital resources are the component parts of the Internet - World Wide Web (WWW) tandem. After the beginning of the World Wide Web at the beginning of the 90s of the last century, the academic environment looked for models of organization, management and exploitation for the digital representations of the scientific works they were creating. The field of cultural heritage had the same concerns from the desire to exhibit the digital representations of the great works of humanity. Both areas needed an organizational model, which can be reduced to the need to describe the resource using metadata, uniquely identifying the resource, as well as connecting it to related or complementary ones.
With the maturation of the World Wide Web comes a consolidation of the standards that govern the components. Most of these are known as Open Web Platforms.
The research carried out focused on the communication mechanisms of these resources. Through communication we have reduced the whole spectrum to the elements that allow the interconnection of digital resources, which in the administrative contexts of the so-called digital repositories, become digital objects. The aspects concerning Application Programming Interfaces were studied , i.e. APIs as the first tools for creating digital repositories and which allow the implementation of mechanisms to connect to entire collections of digital objects without the need for intermediation through graphic interfaces. Application Programming Interfaces, i.e. APIs, are currently the basic model for exploiting digital resources. Having the same generic API name, we also have tools at hand that allow interaction with software packages that expose interfaces that facilitate programmer access to functionality.
This paper primarily deals with the technologies and working models/architectures involving Web APIs, but where they are in extension or twinned with software APIs, they will be described without making a deliberate distinction. Software APIs can be viewed as a dashboard through which we can interact with complex software packages without having to understand the entire architecture that is *hidden* behind an interface. Web APIs can be viewed as the seaports of the planetary ocean ready to receive resources/data transported through the Internet.
Web APIs and those exposed through different services of some software components is the main theme of the thesis.

Research objectives

Researching the current systems and APIs that provide access to scientific research results as well as cultural heritage data is a must in the context of emerging changes to how data and metadata specific to exposed resources are accessed using the World Wide Web Web (WWW).
At this moment, the changes refer on the one hand to the way resource representations are created, following recipes prescribed by national or institutional policies, and on the other hand, the technologies available to make implementations that facilitate access to a varied range of representations are modified by incorporating those specific to the Semantic Web.

Other technologies that have developed in the last 15 years, such as peer-to-peer and blockchain, are directly concerned with the articulation of new types of information spaces, where the focus is on authenticity, security and unique identification not only of resources, but also of transactions realization in specific chains of a suite of protocols that are the basis of these new spaces.

The main objective of the PhD thesis is to create a model of an information space for the exploitation of data and metadata through APIs that communicate with Web2 and Web3.

To achieve the main objective, the following specific objectives are targeted:

O1. Literature review to determine the role APIs play in developing research and cultural heritage data and metadata management systems.

O2. Investigating the information ecosystem where data in the form of digital objects are described and distributed through specialized solutions. The investigation seeks a good foundation of the existing functional components.

O3. Studying existing APIs and the context in which they are used.

O4. Studying the most suitable Web3 architecture for data distribution and making a model that benefits from the functional advantages.

O5. Exploring opportunities for data and metadata communication through software implementation of a functional model for distributed storage using Web3 technologies applying Open Web Platform tools.

, web APIs, as well as their implementations were investigated in order to identify models of representation of some hypermedia resources (data and metadata) currently used in the communication of agents (humans or machines). [1]It was also sought to identify a possible model that would provide a new way of interaction using the technologies that hypermedia peer-to-peer protocols (Web3 – Decentralized Web) make available. The W3C defines the Open Web Platform as "a collection of open (no-cost) technologies that make up the Web." The primary technologies listed by W3C are those that allow the creation of hyperdocuments: HyperText Markup Language (HTML), CSS (Cascading Style Sheets), DOM (Document Object Model), SVG (Scalable Vector Graphics), MathML, EcmaScript/JavaScript, Web APIs, HTTP , URI and MediaAccessibilityChecklist. The work will not investigate the aspects related to the programming of web APIs, limiting itself only to highlighting the constructive blocks at the architectural level, as well as the suite of protocols and technologies that are necessary to realize these blocks.

Main features of the current models that existing implementations provide have been inventoried and highlighted, as well as the inventory of data (resources) and metadata (representations) to

[1] https://www.w3.org/wiki/Open_Web_Platform

develop a possible future interaction model using the technology stack of the current Web with a view over the opportunities that Web3 makes available (objectives 2, 3 and 4).

Finality is a possible model of organization and access to data and metadata, using not only the technologies of existing web APIs, either as means of access and interconnection, or as means for creating management interfaces, but also of distributed technologies on which Web3 has been proposing for more than ten years already (objective 4).

The new model aims to meet the needs of data interconnection in a more efficient way. The data from the field of research and cultural heritage are the target of these studies in the context where at the European level the preservation of a true digital heritage is being discussed, and the field of scientific research needs to work with FAIR (Findable, Accessible, Interoperable and Reusable) data. It will be analyzed what is the nature of a digital object, what it is when it is created/used/transacted in research processes (FAIR data objects – research data objects) and how digital heritage objects are used. The technologies that are used to store/describe and interconnect these objects will be investigated to create the technological context and current best practices.

The resulting model brings a possible reduction in the number of software applications and participants in the transaction and exploitation of web resources or those in value chains of distributed technologies (objective 4). In the case of metadata, a possible capitalization of the experiences in working with the data that the Semantic Web exposes will be pursued. In this sense, the technologies of the Semantic Web will be examined and a perspective of how these technologies play a role in data trading will be outlined.

The current systems are investigated in order to extract experiences of an organizational nature, as well as efficiency models that will be incorporated in the final model, as well as in the transition model realized in objective five. The highlighted technologies were implemented to build a working model of an application that can manage digital objects using the Open Web Platform suite of technologies.

Thesis structure

The work is structured on three main plans divided into five chapters.

The first plan was dedicated to the investigation of the scientific literature using bibliometric analysis tools accompanied by an investigation of the foundations of the existing systems that are engaged in the communication of digital objects. This step was necessary to delimit the research and to understand the context of concerns for the theme in the international context of scientific research dedicated to the theme (chapter 1). Digital objects are investigated starting from the types of data, their nature from the perspective of management and digital preservation , the metadata necessary for the description, as well as the evolution of the architectures and solutions necessary for their aggregation and digital preservation (chapter 2).

The second plan concerns the technologies through which the communication components of digital repositories are technically realized. A point of interest is also how the efforts to open up "data silos" are supported by policies at the European level. The investigation begins with existing standards, looking at how web APIs are modeled through specific protocols and architectural patterns. Standardization is not only viewed at the level of Open Web Platform technologies, but goes further to the standardization needs for the Web APIs themselves (chapter 3). A good insight into existing APIs was achieved by investigating a set of implementations currently in use for accessing scientific research and cultural heritage data and metadata (chapter 4). Chapter 3 exemplifies how APIs can be accessed through calls that return the desired data/metadata that can be processed in true chains of operations.

The third plan concerns the standards used to describe digital objects that are accessible through Web APIs, as well as the opportunities that Web3 technologies expose to realize new models of connection and dissemination. This plan aims to investigate the Semantic Web to identify patterns used in the description and identification of digital objects. The focus is on the aggregation opportunities that semantic graphs provide in a standardized way at this point (chapter 3 and 4). From the panoply of technologies developed under the generic name of Web3, IPFS (Interplanetary File System) was analyzed. The analysis included how the data is aggregated (Merkle Dags) and identified (CIDs – Content Identifiers). At the same time, a functional model for APIs capable of being extended in the Web3 domain was also created (chapter 5).

Organization of the thesis by chapters

The first chapter is introductory, specifying the purpose and objectives of the paper. It is the chapter in which the bibliometric analysis is made on an important sample of scientific research works (5089 articles) of current affairs from the Euro-Atlantic space. The analysis revealed a good connection of the API concepts with the rest of the descriptors extracted from the articles. This indicates that APIs are only apparently not a central topic for the field of data management and research results. In fact, in most cases APIs are mechanisms that are in the subsidiary, their presence does not need to be highlighted, a fact that gives them seemingly little importance. Coupled with a study presented in the paper, which analyzes a set of APIs whose description and characteristics have been completed as descriptive attributes, these connection mechanisms become central elements in data distribution/interconnection. As a final conclusion of the scientific research literature review, APIs are a component of data management services that is present in the context of obtaining metadata describing digital entities/objects. APIs are built as component parts of software solutions. Each API is locally adapted to specific needs. There is no harmonization of implementations other than prescriptions provided by W3C or industry.

Chapter 2 sought to outline a complete picture of technology and its evolution over the past thirty years. The main standards used in data serialization were brought into analysis because they are the ones used to compose descriptions in the form of metadata. Many of the working concepts have been clarified, which provides a rich context for in-depth understanding of the data exposure

model using Web APIs. The standards of good practice regarding the connection of data via the WWW have been brought to the fore. This point was all the more useful as metadata is actually data itself. This data, which has different roles, from administrative to descriptive, is the data that is most often exposed through API endpoints. It was necessary to investigate the standards and practices involving vocabularies and ontologies available to write metadata that achieves an increased level of interoperability. This is the chapter clarifying the concept of digital objects as a premise for clarifying the nature of FAIR digital objects. In the context of the valorization of scientific research and digital cultural heritage, these entities are called research objects *or* Cultural Heritage Objects (CHO). The chapter is complemented by an insight that also has a correspondent in the practical application made to support the research. This is Reusable Learning Objects (RLO). The practical dimension in the field of education involved a deepening of the study on the architectures of digital repositories, which are the applications whose mission is to aggregate, describe, manage, as well as ensure access to digital objects. This was necessary because in the last years of the second decade and the beginning of the third, the community is moving towards more flexible tools, better interconnected data. The ultimate goal is that they can be easily found and quickly inserted into complex processing flows. Of course, the part that will catalyze these changes is directly related to how the components of digital repositories that implement web APIs will be reactive enough, easy enough to integrate into the processing flows of third-party tools. From this perspective, it was necessary to investigate technologies for data communication and metadata. A more comprehensive study of existing tools (open source based on adoption rate) could not ignore the dimension of digital preservation, which provides the most sought-after attribute for a digital repository at the moment: trust.

The third chapter is dedicated to the way in which data interconnection is achieved. This is also the most technical content, looking in detail at the architectures and protocols underlying software implementations of web APIs and more. APIs based on request-response, but also those that are based on events, were introduced into the analysis. This analysis was coupled with the study of RFCs (Request for Comments), the true standards that underpin WWW technologies that are managed by the IETF (Internet Engineering Task Force, https://www.ietf.org/). A review of IETF activity revealed the existence of a group dedicated to standardizing machine-to-machine communication using Web APIs: *the Building Blocks for HTTP APIs Working Group* . Current models of metadata communication in libraries and organizations maintaining digital repositories were investigated: Z39.50 and OAI-PMH. For the European continent, the services provided by national libraries have been evaluated. This European tour was necessary to outline existing practices regarding communication between catalogs, but also remote data access. This tour of the European horizon reveals the success of the OAI-PMH standard, but the story it really tells is one of persistence in using the technologies once they have been adopted. Moreover, for memory institutions, especially libraries, what chapter three reveals is that once a technology has been adopted it will be kept alive even if others are adopted in parallel that enjoy similar success or which offers a clear superior competitive advantage.
The study is also completed by evaluating the deployment models of modern web applications that are hosted and managed using cloud services. The internal architecture of these applications indicates a preference for microservices in tandem with virtualization technologies.

This chapter also clarifies the details of API standardization efforts by investigating the solution that first established itself as a best practice under the name Swagger and more recently the unanimously accepted specification called OpenAPI. The technical side of APIs is only useful if it can be framed in policies that provide the perspective of continuity and thus funding. From this perspective, the study to this point considers European-level policies aimed at achieving a single market and for data with a perspective regarding blockchain technologies. One of the important aspects in creating new informational spaces where data is articulated through APIs refers to the application of ontologies in the context of knowledge graphs, as well as the principles of Linked Data.

Chapter four considers building a new model involving Web2 and Web3 technologies. At the same time, this exploratory model has its foundations in the conclusions drawn following an analysis of a data set that includes the most used APIs at the moment.
The main concepts of distributed systems that are part of Web3 are analyzed. This path leads to the understanding of the possible opportunities to expand the services of digital repositories as long as they become nodes of Web3. The range of opportunities that the Interplanetary File System offers for faster integration and access is analyzed, eliminating as many intermediaries as possible, aiming for a safer exploitation environment. The advantages and architecture of IPFS are evaluated. Special attention has been paid to how data is handled and especially identified in the context of IPFS.
The proposed model looks at a digital object from the perspective of the needs for it to be dynamic. Have the ability to connect and natively respond to capitalization requirements. A new concept is introduced that aims to transform a passive digital object, as current models treat it, into a reactive one. Achieving this level of *reactivity* involves blockchain technologies such as smart contracts, but also capabilities to interact through APIs with the already existing infrastructure.

Chapter five aims to create a functional model for APIs integrated with Web3. In order to reach the necessary conclusions from the point of view of software implementation, I chose to explore two solutions for the integration of digital objects. The first is the implementation of a specialized digital repository at the National Physics Library from the National Research-Development Institute for Physics and Nuclear Engineering "Horia Hulubei" - IFIN-HH, and the second is a solution created specifically to aggregate digital objects from the field of education (Open Educational Resources). The two approaches, the first related to the integration of existing resources using an open source application and the second built from scratch, provided a complete insight into the capabilities as well as the limitations of the Open Web Platform technology stack.
Two software solutions were presented that I worked on throughout the doctoral program. The first involves the implementation of an existing software solution that is widely adopted to semanticize the aggregated contents (Omeka S), and the second solution is an original one dedicated to the aggregation of Open Educational Resources.

Chapters six and seven are devoted to discussions and final conclusions.

Research methodology

The thesis is based on two main studies. The first is an extended bibliometric analysis on several collections of scientific research articles from the activity of prestigious conferences and journals in the field, and the second on a set of APIs with the aim of finding the elements that define existing implementations .
In the first study, more than five thousand research articles were investigated to outline a state of the art and to assess as accurately as possible the area of concern for APIs in the context of the exploitation of scientific research and cultural heritage. These are:

- ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL - https://www.jcdl.org/)
- Journal on Computing and Cultural Heritage (JOCCH - https://dl.acm.org/journal/jocch)
- Semantic Web in Libraries (SWIB - https://swib.org/)
- Research and Advanced Technology for Digital Libraries (ECDL/TPDL - https://link.springer.com/conference/tpdl)
- International Journal of Digital Libraries (JODL - https://www.springer.com/journal/799)
- Code4Lib - https://code4lib.org/

These benchmarks were chosen for the value of the published papers, for the increased complexity of the solutions presented in the research articles, as well as for the historical dimension. This last aspect proved very useful to outline a context that mediated the understanding of some aspects regarding the evolution of solutions and some systems used today.
All abstracts were read, after which manual indexing was done (the so-called *golden standard*), and the main concepts treated in the article/research paper were extracted. No means of automated extraction were used due to the need to fully and thoroughly explore the entire body of scientific literature. The focus was on identifying all scientific literature to be described by the following concepts: Application Programming Interface, API, RESTful or APIs.

To carry out the investigation, existing software packages such as Zotero, VosViewer and Obsidian were used, but because the data needed to be modeled at some intermediate steps, two software packages were written which are available as open source. The open source software packages that have been used are:
- **Obsidian** (https://obsidian.md/) - an application used for personalized note taking, but which was used to collect, curate and annotate research article data. Dataview, DB Folder, JSON/ CSV Importer plugins were used to save data ;
- **Zotero** (https://www.zotero.org/) - an application used to collect, organize and format research paper citations. This software package was used in a dual role. The first being filtering the bibliographic data extracted from Obsidian after it has been collected. The second role was that of organizer of bibliographic notes for the present work;
- **LibreOffice Calc** (https://www.libreoffice.org/) – work package with spreadsheets to quickly reformat the content of CSV files;
- **OpenRefine** (https://openrefine.org/) - package used for data cleaning and formatting;
- **Gephi** (https://gephi.org/) was used as a clustering solution and visualization of the resulting graph;
- **VOSviewer** (https://www.vosviewer.com/) was used to quickly investigate bibliographic data.

The software applications that were written specifically to process the bibliographic data needed for the analysis are five of which two are private, and the remaining three can be consulted and used freely by accessing the following Github repos that contain software created specifically for the needs of transformation and analysis of this study:
- ***obsidian-auto-metadata*** (https://github.com/kosson/obsidian-auto-metadata), being an application that enriches Obsidian notes with metadata for conference or journal articles. Unfortunately, Osidian is not enough with its own tools when it comes to generating metadata. For this reason, I created this application that based on the structure of an Obsidian note, being a Markdown file, collects its primary data (title, authors and abstract) which it writes in the metadata area in YAML format, thus enriching the note that only after this processing is ready to be exposed to further additions and then the harvesting of this data to constitute the required set in CSV format;
- ***csv-to-ris*** (https://github.com/kosson/csv-to-ris) converts a simple bibliographic data structure in CSV format into a RIS file for import into either Zotero or VOSviewer. This version allows inclusion of the abstract if desired;
- ***GnodEdge*** (https://github.com/kosson/GnodEdge) transforms a bibliographic dataset into a graph ready to be explored with the specialized Gephi package. This package applies repeated transformations to the data to detect which descriptors (keyword/concept) identify the most research papers.

The bibliographic data had a course dictated by the final need for an analysis assisted by the Gephi and VOSviewer packages involving the bibliographic metadata. In order to achieve the most complete perspective on the investigated research resources, a local knowledge base (personal knowledge base) was created using the Obsidian application (obsidian.md). This application was chosen for its simplicity and expressiveness in the effort to create and analyze personal study notes. Beyond the technical details, Obsidian was used as an ad hoc database to investigate bibliographic resources, organize research resources, extract notes useful to the investigation, and obtain a highly connected knowledge graph by creating additional resources : a database of acronyms, a glossary and a database dedicated to researchers involved/discovered during the traversal of the scientific literature. Each journal has its own directory, each article benefits from its own note in which apart from the metadata area which is vital, it contains at least the abstract. A very important reason why the Obsidian solution was chosen is related to the analysis capabilities it offers through various data filtering mechanisms, but especially for the data and metadata import and export capabilities. The data was imported into Zotero so that Zotero could then be used as a data dispatcher formatted using different formats for distributing bibliographic data. For example, for data analysis using VOSviewer, I exported data from Zotero in RIS format.
The topics chosen for the choice of the field of investigation were related to the context in which APIs find their utility: digital repositories, digital libraries and the technologies used to make software implementations for them. From this perspective, I have chosen the sphere of metadata, which has a solid pillar in the field of semantic web technologies. All abstracts were read, after which manual indexing was done (the so-called *golden standard*), and the main concepts treated in the article/research paper were extracted. No means of automated extraction were used due to the need to fully and thoroughly explore the entire body of scientific literature. The focus was on

identifying all scientific literature to be described by the following concepts: Application Programming Interface, API, RESTful or APIs.
This effort was not singular, but was accompanied by additional investigations using applications that filter information by targeted indexing terms.

Data exploration and visualization was also done with Gephi, a software package recognized for synthetic insights into a graph in which data is organized. To build a viable graph in Gephi, it needs two separate files, one containing all the entities in the graph, and the second all the edges of the graph, specifically which entity connects to another and what are the attributes of this connection . Since there is no tool for bibliographic records analysis that creates the above files, I created a specialized software called GnodEdge, which can be consulted at: https://github.com/kosson/GnodEdge .
The consolidated dataset was subjected to the transformations mentioned in the guide at https://github.com/kosson/GnodEdge/blob/main/DOCS/preparingdata.md and then went through a manual processing step, after which, according to the created application guide accessible from https://github.com/kosson/GnodEdge/blob/main/DOCS/operations.md , the resources needed to build the graph in Gephi were available ready to be transformed into a useful graph.
Once imported into the application, the data were enriched with statistical indices obtained by applying algorithms characteristic of graphs. The purpose of this processing was to understand the context in which API-related topics establish links with the central descriptors discovered when indexing the body of scientific literature.
In order to outline these conclusions more thoroughly, I sought to dynamically investigate Wikipedia as well, treating each page dedicated to a topic of interest as a potential node of a graph capable of revealing links that confirm or diminish the findings. For this purpose we used an application called Seealsology which can be accessed from the following link https://densitydesign.github.io/strumentalia-seealsology/ . As a working data set, three lists that differ in the revealing potential of the set of connections were provided.

Regarding the analysis of the existing APIs, Google Forms was used to create a special form, which in the administrative area creates a Google Sheet with the collected data. Thus, Google Forms in addition to Google Sheets were used to create a mini database. This solution was chosen for the flexibility and native analytical capabilities of Google Forms.
In order to create a comprehensive and well-grounded context, the standards of Open Web Platform technologies with extension to the Semantic Web and Web3 were researched. For the same purpose, good practices in exposing and connecting to data using the World Wide Web were also investigated. These true premises catalyzed a study that centered on the concept of the Digital Object and by extension the FAIR Digital Objects.
The data was collected over five months, starting in February 2023 and ending in early June of the same year. A good balance was sought regarding the institutions involved in this study. The domains of academic communication and cultural heritage were chosen because the data coming from their APIs have the potential to connect and become part of exploitation chains/workflows. All APIs were investigated by going to their institutions' websites, reading the help and FAQ pages. For most of the APIs discovered, accounts have been created that explore the possibilities of direct

access to data and metadata. After accounts were created for the outlets requiring this approach, a collection was created in the specialized Postman software to extract sample data for further analysis. Analysis of these samples was done to find the identifiers used in the metadata returned as data. These identifiers would constitute as many points of connection with entities that would enrich the context, be they bibliographic records, or digital objects belonging to well-known domains (namespaces) or even directly accessible data.
These APIs were probed against the metadata exposed in sample records, and in cases where the documentation did not provide them, accounts were created following the vendors' guidance to obtain access tokens for authentication. The following providers were explored following the documentation provided to obtain metadata samples using the well-known Postman software: arXiv, Orcid, Springer Nature, FAIRSharing API, DPLA (Digital Public Library of America), New York Times, and Elsevier.
The data was collected to position the API service as best as possible in the context of the digital services in which it is integrated and for this reason, data was collected regarding the location of the documentation, the direct link to the API service where it exists because sometimes the access is through to the specifications in the documentation, the guardianship institution or the department where applicable and a description. These were followed by more details of a technical nature which we will analyze as follows. National and international data portals dedicated to cataloging data from institutions providing information in the public sector were not included in this study in order not to create an overlap of the profile of memory and research institutions with those creating aggregation points for general interest data of all fields of human activity.

Highlights of the research park

The research effort can be marked by several conclusions regarding each of the stages covered in each chapter. These conclusions punctuate the investigative effort that targeted all technical, best practice and policy benchmarks.

The analysis of the scientific literature revealed a good connection of the concepts regarding APIs with the rest of the descriptors extracted from the articles. This indicates that APIs are only apparently not a central topic for the field of data management and research results. In fact, in most cases APIs are mechanisms that are in the subsidiary, their presence does not need to be highlighted, a fact that gives them seemingly little importance. Coupled with a study presented in the paper, which analyzes a set of APIs whose description and characteristics have been completed as descriptive attributes, these connection mechanisms become central elements in data distribution/interconnection. As a final conclusion of the scientific research literature review, APIs are a component of data management services that is present in the context of obtaining metadata describing digital entities/objects. APIs are built as component parts of software solutions. Each API is locally adapted to specific needs. There is no harmonization of implementations other than prescriptions provided by W3C or industry.

The HTTP protocol acts as a fundamental link of communication in the WWW, a fact that positions it at the center of the information space of which web APIs are also a part. HTTP/1.1 was published

in 1998, and two years later Roy Thomas Fielding of the University of California, Irvine proposed a new architectural style for software systems that performed hypermedia distribution. This new architectural style called Representational State Transfer - REST is becoming an essential architecture for all web APIs made in the last 20 years. Analysis of APIs cannot be done without an analysis of how the data that an API makes available is serialized into a text arrangement whose syntax follows the rules of universally used standards. We can say that the web resources that we can access through an API are representations of the data, a representation of the state it was in before it was requested. Before being sent to the requester, the data is transformed into text documents through the serialization process, leaving it up to the programmers to choose a format that is expressive enough to encode records whose complexity is sometimes very high.

Digital objects are the foundation of understanding the modern architectures of the components of the networks for the exploitation and distribution of scientific research results, as well as the objects belonging to the digital cultural heritage. Understanding them proves absolutely necessary to realize the context of the services and curatorial actions necessary for the management and to achieve a digital preservation of them.

Entities involved in data transactions using APIs are serialized digital objects that use *the protocol stack*[2] and components *of the OSI Model*[3] in order, together with Open Web Platform technologies, [4] to transfer data from one computing system to another remotely located, in order to satisfy the needs of information processing or display. In the context of the valorization of scientific research and digital cultural heritage, these entities are called research objects *or* Cultural Heritage Objects (CHO). Congruent with these efforts to outline a digital entity capable of satisfying more technical and conceptual requirements in the field of scientific research exploitation, there is a field that has already accumulated its own rich experiences in managing digital objects of increased complexity. This field is that of education, which has similar needs in terms of *packaging* educational objects, be they in isolated learning contexts or Open Educational Resources. In order to study the experience of working with digital objects of high complexity and heterogeneity, the educational field was investigated which revealed an experience from which many important conclusions can be drawn regarding their management, connection and distribution using APIs .

The results of scientific research are understood as digital objects, which regardless of format and size must become part of the valorization networks initiated by the European Commission through the establishment of the European Research Area, an internal market for research. The creation of this market, basically a turntable for research, is a specific objective of the European Union, the catalyst of which is Open Science, according to the communication *A new ERA for research and innovation* from 2020. According to the requirements for integration into the interoperable research structures of European Open Science Cloud, research data must comply with the principles

2 https://en.wikipedia.org/wiki/Protocol_stack
3 https://en.wikipedia.org/wiki/OSI_model
4 https://en.wikipedia.org/wiki/Web_platform

of FAIR [5](Find, Access, Interoperate, and Reuse), which transforms them into FAIR Digital Objects (EOSC Executive Board and Directorate-General for Research and Innovation (European Commission), 2020). Easy to see is the central role that metadata plays, having a primary role in achieving data interconnection. The role of these documents related to the integration of the FAIR principles in all activities related to scientific research of the European Union directly refers to the need to understand what a digital research object is as it emerges from the investigated texts. These digital entities can constitute the necessary parts of a processing flow (*workflow*). At this point, in order to arrive at the results of a research study, all stages of experimental data processing should be recreated. It would be useful if the entire processing flow could be memorized in some way to be initiated whenever it is necessary to reach the result. APIs can play an essential role in obtaining the resources that constitute the digital research object.
The Research Data Alliance proposes a set of measurable indicators derived from the FAIR principles to answer the question "what should be measured to see the degree of FAIR compliance of a digital object". Thus, forty-one indicators were created, which together constitute a model that can be further implemented in tools capable of evaluating a data set if it respects the principles. A close examination of the Model will reveal that all indicators are very important in the economics of accessing digital objects using APIs. We can say that they create a true audit trail for those who create APIs, as well as for those who access them beyond the technical solutions behind the implementations.

Metadata is the most valuable part of the economy of interconnecting datasets. Metadata are distinguished from each other by the fact that they serve different technical aspects. Some are aimed at the administrative level, others provide information on the structure of the data, and those that concern the nature, usefulness and context in which they can be used are descriptive. The importance of metadata is revealed by the fact that they are the ones that make the difference between a simple aggregation of electronic resources and a complex object that, through the way of making connections with the parts, becomes highly functional and thus, reusable. According to the FAIR principles , metadata must be rich, uniquely identified, include the identifiers of the digital objects to which it is attached, and be indexed using a mechanism that exposes it to interested parties or agents.

A digital repository is a searchable and queryable entity capable of storing, managing and curating (with administrative meaning) Digital Data/Objects (Research Data Alliance). Digital repositories are service structures that provide optimal solutions for organizing digital resources. In the context of the WWW information space, digital objects are called resources. Basic functionalities of a digital repository:

- search/retrieve
- communication and/or interconnection of objects through metadata

[5] https://www.eoscsecretariat.eu/working-groups/fair-working-group

- easy integration of objects from other repositories

Digital repositories are a set of capabilities and administrative models over one or more digital objects and datasets.
At the beginning of the 2000s, digital repositories concentrated development around institutions ("a «silo» trend") which then expanded thanks to international projects that had in mind the aggregation of collections over which there could be federated search services. This is the decade in which the architectures of software solutions such as Fedora, DSpace or EPrints crystallize. The Open Access movement is becoming a catalyst for effective research output management solutions.
Digital libraries are highly specialized services built on digital repositories that provide the basic services for managing digital resources. Digital resources are information structured in files according to their types and functionalities. All digital resources subject to the rigors of organization through digital repositories must be described according to specific standards. This is visible when we access a digital repository established to serve the community of archivists, another organization provides us with a digital repository made by the community of librarians. These possible forms of resource organization are not limited to the mentioned communities, but are extended to all natural and legal persons who need to rigorously structure their collections, whether physical or exclusively digital.

The roles and responsibilities of memory institutions in terms of research output and cultural heritage are evolving in the context of a difficult struggle with the growth of digital collections and the decrease of budget allocation. This paper seeks to find the salient aspects of how digital preservation practices are actively incorporated into digital repository management actions and policies. The investigation considered major documents issued by the most respected European and international institutions, with a particular interest in strategies, training frameworks and dedicated policy frameworks seeking solutions for long-term preservation.

Application Programming Interfaces (APIs) is a modern mechanism that allows the interconnection of data and discrete services in the context of the increased need to create articulated informational spaces through which data is circulated and/or processed automatically or semi-automatically. API - Application Programming Interface achieves communication between computers by exposing interaction mechanisms with their software. An API is not an interface that ensures the interaction of a user with software in the sense of graphical interfaces. It is a facility made available to programmers to integrate API data and services into their own applications. The deep nature of APIs is communication and data access. From this communication perspective, APIs are created according to the request-response paradigm (request/response APIs) or as event-driven APIs. Essentially, what is achieved using APIs is an integration of different software packages that communicate. Moreover, a communication of distributed systems is realized. Technically speaking, APIs greatly simplify systems programming because they move the effort to an abstract plane and the focus is on the data/objects being processed/transmitted. This abstract plane is usually called

so because it isolates the mechanisms needed to interact with the data into simple commands/instructions. Beyond the interface through which you operate them, they have a life cycle of their own, including operated software packages, initialized processes, etc. Abstractions in the context of APIs are simple actions that hide complex processes. APIs act as a point of contact for programmers with a software system. The API hides the details of a particular implementation (functions, procedures, classes). What it exposes is a small set of well-documented mechanisms by which one can interact in a time-consistent manner with a software application. The way the implementation is done may change over time, but if the API call mechanisms remain the same with the same effect, a stable and predictable level of exploitation is achieved. Thus, APIs can become the components of a larger system, and its maintenance is much easier over time. Every API has a specification that states how it works. At this point, the near-universal adoption of APIs as a data exposure model has led to the organization of software applications as microservices. This model, along with the maturity of virtualization technologies and the development of cloud computing, constitutes the prevailing paradigm for web application development. Instead, many of them are based on an architecture characteristic of a cloud application, which involves a good orchestration of components currently called *services* . Each service can represent a function of the application. It can benefit from its own software implementation which allows a degree of flexibility for programmers. All these components communicate using REST API protocols or by brokering messages , using services offered by third parties. Thus, the monolithic aspect (*tight integration*) that we were used to begins to yield to the pressure of much more flexible (*loosely coupled*) and easier to maintain architectures.

Web APIs have become increasingly popular over the last ten years, which has led to the need to standardize how they are built. The most widely used standard at the moment is OpenAPI which is being investigated to outline the aspects that have propelled its success. The Open API Specification (OAS) [6]defines a standardized way to describe the interface through which an HTTP - based API is accessed in order to allow clients to understand what the API provides without having to access the service's source code or related documentation. Creating an API that conforms to the Open API Specification (OAS) standard eliminates ambiguities and efforts to approximate the details needed to make successful HTTP calls to service *endpoints* .

The European level of political support

The European Strategy for Data aims to transform the European Union into a leader of the data-based society by creating a single market dedicated to them, in which their free inter-sectoral movement is guaranteed with direct benefits for the business environment, research and public administration. The most important thing to keep in mind is that people must be at the center of all concerns about this strategy in the wider context of the *Digital Decade* [7]. The European data strategy also involves the creation of data spaces *in* the different fields and sectors of activity. According to the recently created Data Spaces Support Center (https://dssc.eu/ - DSSC) to support the implementation of these data spaces, these are infrastructures that mediate data transactions.

---

6 https://spec.openapis.org/oas/latest.html and GitHub: https://github.com/OAI/OpenAPI-Specification

7 https://digital-strategy.ec.europa.eu/en/policies/europes-digital-decade

Recital 27 of Directive 1024 of June 20, 2019 separates research data from scientific research articles in the context of Open Access publishing, but the most important aspect is the importance that the data management plan acquires it is intended to be integrated into scientific practice. There is no shortage of data qualification that must comply with FAIR principles. Moreover, in recital 28, Member States should oblige themselves to adopt policies on Open Access to research data. Another very important clarification is related to the fact that the data do not necessarily have to be aggregated in specialized digital repositories, the provisions of the directive being extended to those that are attached to research articles published in a magazine or even in serial publications dedicated to the publication of documents with data included. Recital 30 is dedicated to establishing a definition for the term *document , and in Recital 31,* [8]APIs are indicated as the access mechanism for dynamically collected data . APIs take on the value of the tool for capitalizing on *economic potential* . This step is very important because it also includes attributes that can describe the data very well. Recital 32 provides the definition of the API: *an API is a set of functions, procedures, definitions and protocols for communication between machines and the unrestricted sharing of data* . An important emphasis is placed on the quality of API documentation, and it is even indicated without specifically mentioning the standardization of APIs using Open API.
The Joint Research Center (JRC) published in 2020 a Report entitled Application Programming Interfaces in Governments: why, what and how (Vaccari et al., 2020) and in 2022 a Technical Report entitled API strategy essentials for Public Sector Innovation (Posada et al ., 2022). The report Application Programming Interfaces in Governments: why, what and how presents the results of the Application Programming Interfaces for Digital Government (APIs4DGov) study undertaken by the European Commission to understand the role of APIs in the public sector as well as the possible widespread adoption by the governmental institutional arch of each member state. The objectives of the 2020 report were dictated by the need to assess the relevance and adoption of APIs in the broader context of governments' digitization needs. European policies have in mind the creation of digital ecosystems (A Europe Fit for the Digital Age), aiming at a transformation of relations between participants in the context of a transformed economy qualified as a dynamic data-based economy (data-agile economy). Several conclusions of the 2020 report state the need to introduce APIs as a component of digitization due to their modularity and high degree of reusability. At the European level, the adoption of APIs should be done in a coordinated way to avoid the negative effects that ad hoc adoption brings to long-term exploitation. One aspect well captured in the report is that APIs allow institutions to avoid the data "silo" effect. The study also provides a definition of API that we will adopt as a benchmark for this paper: Application programming interfaces (APIs) are digital interfaces that facilitate the exchange of data and services (functionalities).

The role of APIs is growing along with the infrastructure used to manage metadata. Infrastructure must meet growing concerns about long-term digital preservation. For all research carried out under Horizon 2020, researchers must submit papers (in a machine-readable format) and data in

[8] *This Directive defines the term 'document' and that definition should include any part of a document. The term "document" should include any representation of acts, facts or information and any compilation of such acts, facts or information, regardless of the storage medium (on paper, in electronic form or in the form of an audio, video or audiovisual). The definition of "document" is not intended to include computer programs. Member States may extend the application of this Directive to computer programs.*

"trust deposits" according to Article 17 of the model grant agreement (Annotated Grant Agreement) . This requirement is also embedded in the publication guidelines of Open Research Europe, the European open access publishing service offered to researchers involved in the Horizon 2020 project. For this reason, the dataset provided by CoreTrustSeal was considered for the analysis. CoreTrustSeal is a set of requirements for repositories fully named CoreTrustSeal Trustworthy Digital Repository (TDR) Requirements. This is the collaborative effort of the Data Seal of Approval (DSA) and the World Data System, part of the International Science Council (WDS), working together under the Research Data Alliance (RDA). This is a tool by which digital repositories are analyzed for compliance. As long as all criteria are met, the digital infrastructure is declared capable of providing long-term preservation services.
The data tells an interesting story about the importance of developing and maintaining an API. Almost a third of CoreTrustSeal repositories have an API. The rest have no public endpoints.

Highlights and discussion of APIs and models that include Web3 technologies

APIs are the mechanisms most useful in making functional connections between the various data silos on the Web. In the 2020 report (Vaccari et al., 2020) there is a characterization of APIs that restores the understanding of the value of APIs from another perspective: APIs are technical contracts that can be seen as software products that have a value chain. The most important point that can be captured is that of the need to look at APIs as communication mechanisms between machines. But for this to be possible the APIs would need to be built and perform transactions based on real contracts. The RFC (Request for Comments) technical notes have proven to be the most useful sources of information, which have the role of standardizing how technologies should be implemented on the Internet and in the World Wide Web information space. From the wide palette of APIs, web APIs were singled out for the thesis study because they are used for communication and dissemination of data and metadata. Another important criterion for focusing on web APIs is related to the fact that they implement technologies that are part of the Open Web Platform.
After outlining a useful profile of web APIs complemented by their standardization, the contexts in which they are used were sought. Relevant was the revelation that APIs do not have a flanking position in an architecture, but have themselves become central components in European digitization, for example. APIs are centrally positioned even in the central policies concerning the realization and functioning of the Single Data Market.

Web3 is a new organization of specialized information spaces through the use of protocols focused on restoring trust in communicated data, based on cryptographic means. The main attribute of a Web3 network is **trust** . This trust is derived from the fact that Web3 technologies allow the verification of the status of transactions taking place on the network at any time. Thus, the fields of application are that of finance, but the most sensitive would be that of identity management. The premise is that no actor in the network can be trusted. Achieving the level of trust is done through the use of blockchain. The distributed model is needed to update the blockchain with the current state or better said, with the last state of a transaction. Until yesterday this role was fulfilled by a

database. Participants in a blockchain are servers, not clients. Specialized services have emerged that mediate clients' access to the blockchain through an intermediary layer made using dedicated APIs.

One of the important mechanisms that the study aims at is smart contracts. To communicate with a smart contract, an intermediary is needed that takes the signals indicating a necessary change to send them in the form of data for the smart contract to understand and create the necessary changes. The biggest advantage of a smart contract is that it reacts to changes, being software that executes when those changes occur. Changes are irreversible and easy to track due to the constructive nature of the blockchain. For this reason, smart contracts can be considered at least for the record of research results.

Another important component that was taken into account for the realization of the oe model that the thesis proposes, is IPFS - Interplanetary File System. The Interplanetary File System [9]is a suite of protocols that solves the problem of access, addressing and truly unique identification of content. IPFS can be seen as a file system that instead of storing data conventionally by maintaining localized records, uses a DHT – Distributed Hash Table [10]in a peer-to-peer architecture. The DHT is a way to couple key values in a distributed model. In IPFS it is the core of the request routing system through which a particular hash is discovered. To make it easier to cheat, we can say that it is a catalog of the resources that reside in the active nodes of IPFS. The data model is a Merkle DAG tree [11]that uses the mechanisms of the decentralized web. IPFS is a global network of nodes with which resources are traded.

To interact with this network several APIs1 and a desktop application are provided by Protocol Labs (https://protocol.ai/).

Returning to the interaction model, three can currently be identified using the Internet: centralized, decentralized, and distributed.

The distributed model solves the following problems:

- the data belong to their owners - they do not reside on servers beyond their control;
- the data is very easy to verify thanks to the cryptographic signature (content identifier - CID), which means a high level of security;
- data distribution is easily scalable;
- data is resistant to passing through unsafe environments;
- data that is duplicated is managed without being multiplied;
- there are no central points in the network that once compromised produce a cascading effect.

At this moment the interaction with this distributed network is carried out through the specific APIs of several programming languages. To understand work entities, we will go through the building blocks of IPFS with the aim of outlining the functional characteristics, but especially the opportunities that this technology offers to work with digital objects of which more diverse.

In order to have an overall perspective, it was necessary to investigate the current solutions and models. Beyond the active parts already discussed, APIs are the mechanisms for interacting with Web 3 implementations. The key to a new possible growth model proposed below is rooted in the

9 https://ipfs.tech/

10 https://docs.ipfs.tech/concepts/dht/

11 Merkle Acyclic Graph (DAG)

technologies offered by Web 3 so that API transactions use data stored in graphs that are hosted by distributed solutions and not in silos in light of the practices mentioned in the paper. We have seen that in the intended behaviors for digital repositories of the future there is a requirement that updating a resource triggers notifications about what has changed. This need will be reflected in the future by developing some reactive mechanisms that will ensure this behavior. One of these mechanisms can be a smart contract in a certain blockchain. If the structure of the research object changes (versions), the smart contract will be notified and it will be executed triggering a cascade of events provided for each type of notification. These notifications can be related to the modification of the research object, its introduction into a workflow or even its access through interaction with third-party services. What's more, whenever the research object is cited, the smart contract can receive a notification and thus execute it.
In the not too distant future it is possible that scientific research articles will turn into the metadata of the research object that will be a reactive entity. This can be done by transforming the article, either by encoding using TEI (Text Encoding Initiative), or by fully transforming it into a notebook (Jupyter Notebook). The closest model we've come across that comes close to the reactive entity we're describing is RO-Crate, which provides a sufficiently flexible data support structure but still passively waits to be integrated into a workflow.
At this moment, Web3 technologies are the ones that allow the transformation of inert digital objects of research or cultural heritage, whose level of interaction is achieved through the application that manages them, into real reactive digital objects.

In the general model presented, the premise is that digital objects (reaction entities), as well as metadata, are considered objects uniquely identifiable by their own CID, but also that of the aggregation of which they are a part. Thus, evidence mechanisms will benefit from the exposure of increased granularity, but more than that, from new ways of distributed storage. It must never be forgotten that we are going beyond the limits of the WWW model. The aggregation of resources in complex digital objects such as RO-Crates will be reflected by a distinct graph of the Interplanetary Linked Data type, which will be able to be exploited and interconnected with other similar graphs or currently used through implementations of Semantic Web technologies . Beyond the digital objects themselves, even classes and subclasses of ontologies, and even ontologies themselves will be able to be represented by IPLD graphs. The vocabularies will be the first to benefit from a representation in the Interplanetary Linked Data information space. Each term will get its own CID, and the complex relationships they establish can be represented by IPLDs like thesauruses, for example. The most important gain is related to the fact that the unique identification will be given by a CID, not a link. The CID is not tied to any web location, however its only dependency is the IPFS architecture. Management can be instrumented by creating specialized Dapps. Dapps will play a role in orchestrating these resources. It will regulate access, determine authenticity, and allow interaction with different digital objects based on sets of rules written in the smart contract. Because previously we designed the future research object as a reactive one, from the Web3 perspective, each of these can be a smart contract. This will ensure their own independence and an increased level of interaction. We can imagine a scenario where a scientific research article is actually a smart contract, i.e. a reactive entity whose life cycle is governed by rules. This entity can receive signals regarding certain metrics such as how many times it is cited. This data will be

written in the blockchain and will be secure, tamper-proof. Another interaction may be between the research article and its accompanying data set. The data set can in turn be a reactive entity, that is, its life cycle can be regulated by the rules written in the smart contract. In a scenario of interaction of the scientific article with the dataset, we can very easily get very useful data on how the dataset has been used. If this is also a reactive entity, we can write very useful usage metrics into the blockchain. Moreover, as in the case of RO-Crates, we could introduce the rules of use in the allowed processing flows into the smart contract.

Summary

The paper creates the full context needed to understand how APIs (Application Programming Interfaces) are currently being used to access the data and metadata of digital objects.
The investigation of the scientific literature as part of the introductory chapter revealed that APIs are not a direct concern of parties involved in the management of digital objects. The binding potential they have for the different data silos is limited only to the function of operating with the resources, from their creation to their modification. APIs are seen as a mechanism implemented in software to provide a programmatic data access mechanism in addition to the graphical interface, but there are too few cases where the specialists involved have seen their potential as connectors between data sets, between digital objects FAIR. At the moment, they only serve as connectors of data silos.
The role that APIs have seems to be secondary or assimilated to the functional requirements of software implementations. They are made following the industry standards and models that the RESTful practice has imposed since the year 2000.
Also in the introductory chapter, section 1.3 dedicated to the analysis of scientific research literature reveals the fact that specialists also prefer to work with external APIs to solve specific tasks, either those related to data processing or to access data.
In the second chapter, which deals with the Fundamentals of systems for aggregation and presentation of research data and metadata, the necessary context for understanding the components involved in working with APIs was built. It started from the analysis of data with serialization formats (the form in which the data ends up being extracted from the computer's memory, being stored in a file) in which they end up being the raw material of the APIs. The most used serialization formats were identified and the standards and best practices that currently govern the distribution and interconnection of data using WWW and Open Web Platform technologies were searched. From the analysis of data it has moved to the investigation of how it creates digital objects. It is these entities that are traded in research and digital cultural heritage valorization networks. The analysis included
a theorizing part of digital objects, where a better outline of what a digital object means was sought. The individualization of the attributes of these digital objects was sought. This conceptual delimitation was followed by a study of some digital objects that are used in the educational field, a research space that revealed many models dictated by the need to organize training materials that follow a pedagogical purpose. These learning objects are relevant in the way they are created and the purposes they serve and research objects that follow the principles of FAIR (Findable, Accessible, Identifiable, Reusable). The last section of the chapter is dedicated to metadata

accompanying digital objects, where the most visible schemes used are analyzed. Metadata are still data, but they have the role of contextualizing the digital objects they accompany and describe.
For a correct framing of the APIs, the digital repositories were also analyzed, as well as the aspects related to the creation of digital libraries. A good understanding of these leads to a clear outline of expectations regarding the implementation of a related API. Thus, the most relevant definitions were searched, the connection of digital repositories was analyzed, a reference model was investigated that was adhered to by the majority of those who carry out software implementations and finally, we investigated what it means for a digital repository to be one of trust. An important part of the aspects related to the management of digital objects is that related to digital preservation.
The second chapter concludes with a brief analysis of a software realization model that implements the functional requirements of digital repositories/digital libraries.
In the third chapter the study turned its attention to how data connects through APIs, but also how graphs become the preferred data storage model. In the opening of the chapter, an evolutionary path is marked, web APIs being analyzed from a functional point of view. Because APIs are considered even by the European Commission to be a link between the services of a digitized society, the most important European public policy documents were analyzed to delimit and identify where APIs become relevant and especially what their value is . They play an essential role in the realization of interoperational plans and are indispensable in the recommendations and implementations of policies and initiatives. They are indeed an integral part of the directives which gives them a key position in the architectures that will create European digital services.
In the descriptive fabric of the mantle created by metadata, the technologies of the Semantic Web are combined, whether we are talking about vocabularies, or whether we encounter implementations of some ontologies necessary for the organization of an informational space that serves a certain field or purpose. For this reason, chapter three concludes with an analysis of these technologies and how to organize data using graphs as a preamble to the model that chapter four proposes.
The last chapter is the one in which a set of fifty-one inventoried APIs in the field of research and exploitation of digital cultural heritage is studied. This data set was investigated from the perspective of the serialization formats used, the vocabularies used, the way in which access to the data is allowed, according to the profile of the institution that created it and according to the field in which they operate.
After this analysis, which created the exploratory context conducive to the advancement of a possible new model, the new informational space generically known as Web3 was analyzed. The World Wide Web and Web3 are two information spaces that currently work symbiotically. The proposed model is based on the technologies of both. In the analysis of Web3, the technologies that compose it were inventoried, being investigated data graphs, smart contracts, Dapps, the function of databases in this model, as well as IPFS (InterPlanetary File System) in the role of integrator model for data uniquely identified by cryptographically generated keys based on content analysis.
Finally, a possible model is developed that aims to transform digital objects from passive objects that are the subject of operations into reactive digital objects, which provide mechanisms that allow increased interaction, even a degree of autonomy based on operations algorithms that smart contracts offer. The advantages of such a model are presented. Most of them derive from the way

Web3 works, but which cumulatively, by assigning a digital object, can transform it into a dynamic, highly functional entity through the mentioned level of reactivity, benefiting from the best safety and provenance features on which distributed systems such as blockchain make available to them. The chapters of discussions and conclusions come to outline the work's approach towards the future of reactive digital objects, semi-autonomous digital entities caught in administrative and managerial mechanisms that provide the means so that they can be orchestrated in a new informational space where man and machine to be first class citizens.
Finally, a software application with a demonstrative role was created, which by using Open Web Platform technologies, succeeded in implementing the communication of digital objects from the World Wide Web in Web3. This is an important step towards achieving the level of interaction that FAIR digital objects will necessarily acquire in the future. Digital repositories must no longer behave like data silos, but rather facilitate interaction between reactive digital objects.

Personal and original contributions

The thesis investigates for the first time the context of technologies and policies in which FAIR Digital Objects are situated. This leads to possible policy developments at national or consortia level. For the first time, blockchain and distributed technologies are being investigated to find opportunities to leverage scientific research results beyond the *silos* that current digital repositories still maintain as a paradigm. This study also created the experimental basis necessary to make a demonstration model through the original software created.

The contribution to the development of scientific knowledge can be evaluated by the contribution brought from different perspectives of scientific research. These are embodied in:
A. Synthesis contributions
◘ Theoretical research on the nature of data and metadata that APIs transact/distribute;
◘ Identifying the position that APIs have in the economy of digital repositories used in the field of research and for the valorization of cultural heritage;
◘ Investigating mechanisms and typologies of digital objects in currently used APIs.
◘ Analysis of APIs to identify common characteristics and to look for potential solutions for leveraging and interconnecting through the exposed metadata;
◘ Analysis of technologies and software solutions underlying systems that provide access to metadata and data through APIs;
◘ Analysis of the evolution and the way digital objects are represented in the field of research and cultural heritage;
◘ Exposing the context of European policies that require and mediate the development of APIs.
B. Theoretical and experimental contributions
◘ Realization of a general model for aggregating data and metadata using the World Wide Web and Web3;
◘ Realization of a functional model for the integration of APIs with Web3 by storing digital objects in IPFS as part of the presentation software solution
C. Contributions with a scientific curricular character
◘ Elaboration of scientific research reports within the doctoral research program;

- Completion of the doctoral thesis;
- The current state of research.

D. The novelty of the doctoral thesis

The thesis aligns the objectives to outline the following aspects:

- Positions APIs as a concern in the FAIR research data ecosystem and the digital cultural heritage valorization sector through a vast bibliometric analysis of specialized literature;
- Clarifies what is meant by a digital object in general and a FAIR digital object in particular;
- Clarifies the value of APIs from the perspective of European concerns at the level of policies and initiatives;
- Builds a perspective on the last thirty years in terms of systems for capitalizing on scientific research using digital repositories;
- Proposes a new model that proposes the transition of data distribution practices, from digital repositories that behave like silos, to connecting nodes in a data continuum using IPFS (Web3).