

SICHELIA: Sicilian Intangible Cultural Heritage Exploration Leveraging Open Data

Giulia Speranza^{1,2,*}, Raffaele Manna^{1,2,†} and Maria Pia Di Buono^{1,2,†}

¹Dahlia s.r.l, 219, Via Duomo, Naples, 80139, Italy

²University of Naples "L'Orientale" - UNIOR NLP Research Group, 60/61, Via Chiatamone, Naples, 80133, Italy

Abstract

This paper presents the reuse, enrichment, ontology-driven mapping, and format conversion of open data to support the creation of SICHELIA, a web application designed to promote the Intangible Cultural Heritage (ICH) of the Sicily Region (Italy). Leveraging diverse open datasets, the project integrates cultural, geographic, and historical information into a unified, interoperable knowledge base. SICHELIA enables meaningful discovery and visualization of data through an interactive, user-friendly, and customizable map. The project demonstrates how the strategic and creative reuse of LOD can enhance and preserve cultural heritage, while also promoting socio-economic impact in regions facing the disappearance of ancient traditions, customs, local knowledge, and artisanal skills.

Keywords

Open Data, Semantic Web, Data Enrichment, Intangible Cultural Heritage (ICH)

1. Introduction

Linked Open Data (LOD) play a crucial role in the management, publication, and enhancement of cultural heritage, particularly in Italy, a country with one of the world's most extensive and diverse tangible and intangible heritage assets. The publication and dissemination of open data enable the reuse, linking, and integration of heterogeneous information, thereby facilitating the creation of new knowledge pathways and innovative opportunities for cultural engagement and territorial development [1, 2, 3]. The current state of LOD in the Cultural Heritage (CH) domain in Italy is characterized by significant national and local initiatives aimed at digitizing and making accessible relevant information on CH, in line with the standards defined by the World Wide Web Consortium (W3C). For example, the Italian Ministry of Culture (Ministero Italiano della Cultura - MiC) provides access to numerous open datasets concerning Italian cultural heritage through the dedicated platform¹. This platform is the result of a cooperative process involving the MiC's Central Institutes and General Directorates, and it interlinks datasets originating from different sources: the database of Cultural Sites; registries of Archives and Libraries; the Cultural Heritage Catalogue database; and other documentary and photographic databases. Each dataset can be accessed in a tabular format, directly via its URL, or downloaded in various distribution formats (XML/RDF, text/Turtle, application/JSON, etc.) [4].

IRCDL 2026: 22nd Conference on Information and Research Science Connecting to Digital and Library Science, February 19-20, 2026, Modena, Italy

*Corresponding author.

[†]These authors contributed equally.

✉ gsperanza@unior.it (G. Speranza); raffaelemanna@unior.it (R. Manna); mpdibuono@unior.it (M. P. Di Buono)

🆔 0000-0002-9249-492X (G. Speranza); 0009-0006-6285-8557 (R. Manna); 0000-0001-7284-6556 (M. P. Di Buono)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://dati.cultura.gov.it/>

Notably, as part of the Mission 1 Component 3 Investment 1.1. “Strategie e piattaforme digitali per il patrimonio culturale”² of the National Recovery and Resilience Plan (PNRR), institutions responsible for cultural heritage are required to implement, between 2022 and 2026, the digital transformation processes promoted by MiC, particularly concerning the digitization and publication of data.

Italian local administrations such as regional and municipal governments have also begun to publish their datasets in LOD format on the dedicated Italian national portal³, which includes various datasets on different domain and also dataset about CH. Open Government Data (OGD)⁴ represent a fundamental component toward digital transformation. These publicly funded datasets, made available without confidentiality or privacy constraints, are intentionally released for open access, enabling unrestricted reuse, redistribution, and integration into new digital services [5, 6, 7, 8].

Access to and reuse of LOD published by cultural institutions and governmental administrations represent a valuable resource for ensuring interoperability —namely, the ability to integrate multiple datasets while limiting the proliferation of data silos— as well as for designing new, interactive modes of engagement with CH, generating both social and economic value.

A recurring challenge in the open data ecosystem concerns the substantial investment of financial resources, time, and technical effort in producing and publishing large volumes of open data, without a corresponding development of strategies, tools, or skills for their effective and creative reuse. While institutions increasingly adopt open data policies and produce datasets in interoperable and machine-readable formats, these resources often remain underexploited due to limited awareness of their potential applications, insufficient integration mechanisms, or the absence of user-oriented services capable of transforming raw data into meaningful knowledge. As a result, the value generated by open data remains largely unrealized, highlighting the need for methodological frameworks, technological infrastructures, and socio-technical practices that encourage data reuse, foster innovation, and ensure that the initial investment translates into tangible cultural, social, and economic benefits [9, 10, 11].

Indeed, a strategic and creative reuse of LOD in the context of tangible and intangible cultural heritage can be particularly effective in enhancing and promoting sparsely visited areas, such as inland or peripheral territories affected by depopulation, demographic decline, population ageing, and limited infrastructure: factors that may contribute to economic impoverishment and, in some cases, the disappearance of traditions, customs, local knowledge, and artisanal skills.

Another major obstacle to the development of applications in the CH domain based on open data is the existence of data silos that do not interoperate and often rely on heterogeneous representation models [12]. Data federation ensures the integration, unification, and governance of data stored in distinct data repositories. Within this process, the ability to harmonize the diverse data models used to represent information is of primary importance. The federated system, being independent of specific data sources, guarantees flexibility and scalability.

Through the development of a methodology capable of supporting open data federation in the CH domain, the SICHELIA Project aims at reusing Open Data and improving access to the Sicilian Intangible Cultural Heritage (ICH) within the framework of the DataHighway Project⁵[13].

²<https://pnrr.cultura.gov.it/misura-1-patrimonio-culturale-per-la-prossima-generazione/1-1-piattaforme-e-strategie-digitali-per-laccesso-al-patrimonio-culturale/>

³<https://www.dati.gov.it/>

⁴<https://www.w3.org/TR/gov-data/>

⁵<https://www.datahighway.it/>

2. Case Study

2.1. The Sicilian Intangible Cultural Heritage

ICH represents the ensemble of traditions, practices, knowledge, products, expressions, and crafts transmitted across generations, sometimes orally, which collectively shape a community's cultural identity. As a consequence, ICH tells the story of places and people and encompasses:

- **Oral traditions:** narratives, legends, myths, and songs transmitting history, values, knowledge, dialects, and local linguistic varieties;
- **Social practices and rituals:** religious festivals, ceremonies, seasonal or commemorative celebrations, wedding customs, and hospitality practices;
- **Craftsmanship and artistic techniques:** know-how associated with the creation of traditional objects such as textiles, ceramics, and metalwork;
- **Culinary and gastronomic traditions:** local food products, recipes, and food-processing practices;
- **Artistic expressions:** dance, music, theatre, and other performances that embody a community's cultural essence.

ICH is recognized for its intrinsic value in preserving cultural identity and human diversity, which is why UNESCO⁶ safeguards it as a fundamental component of collective memory and global cultural richness [14, 15, 16]. Moreover, ICH plays a vital role in promoting and sustaining emerging forms of tourism, contributing significantly to both economic and social development [17, 18].

By enabling a strategic reuse of OGD, SICHELIA provides an innovative and interactive access experience about the Sicilian ICH, enhancing visitor engagement with physical sites, fostering a deeper understanding of their cultural significance, and simultaneously supporting the preservation and transmission of lesser-known traditions, knowledge, and practices that are at risk of disappearing. SICHELIA provides a tailored exploration experience of Sicily's rich ICH, ranging from local food and wine products to arts and crafts, festive celebrations, and cultural aspects linked to the local language.

2.2. Sicilian Open Data Selection

The preliminary investigation activity, involved an in-depth assessment of the current state of available Open Data pertaining to Sicilian ICH across various knowledge bases and catalogues. The objective was to map the coverage, structure, and quality of existing data, identifying potential gaps as well as opportunities for enrichment and interlinking.

To obtain a comprehensive overview of the structure, variety, and informational potential of the available datasets, a qualitative analysis was conducted on the LOD published on the official Open Data portal of the Sicilian Region⁷. The portal -serving as the primary access point to regional OGD- hosts more than 180 datasets organized into multiple thematic domains, including: Agriculture, Fisheries, and Food Products; Environment; Economy and Finance; Energy; Justice and Security; Government and Public Sector; Education, Culture, and Sport; Population and Society; Regions and Cities; Health; Science and Technology; International Affairs; and Transport.

⁶<https://ich.unesco.org/en/what-is-intangible-heritage-00003>

⁷<https://dati.regione.sicilia.it>

Among these, datasets potentially relevant to the SICHELIA case study were identified, specifically those containing information useful for documenting aspects of Sicilian ICH, with particular attention to data describing oral traditions, social practices, artisanal knowledge, ritual celebrations, linguistic expressions, and typical products and food processing techniques.

The selection phase followed a qualitative, multi-phase approach. First, employing thematic category filters on the platform we narrowed down the domain of interest to ICH. The most relevant categories were found to be “Education, Culture, and Sport” and “Agriculture, Fisheries, Forestry, and Food Products”.

Subsequently, criteria of semantic web compatibility were applied, with particular emphasis on the availability of publication formats oriented towards the Semantic Web (notably RDF/Turtle, JSON-LD, XML-OWL, etc.), capable of supporting explicit representation of entities and relationships through the use of URIs, ontologies, and standardized vocabularies.

Finally, a preliminary evaluation of the suitability of each dataset for the case study was performed, taking into account aspects such as data granularity, the degree of semantic structuring, the presence (or absence) of interlinking to external datasets, and ontological consistency with nationally and internationally recognized conceptual schemas.

This approach made it possible to isolate the datasets most closely aligned with the domain of interest and best suited for integration within a semantically interconnected perspective following Linked Data principles. The following datasets from the Sicilian Region’s open data portal were therefore selected:

- **Luoghi Siciliani del Gusto (LSG)**⁸: a georeferenced collection of the main locations where essences, foods, and traditional Sicilian products are made.
- **Prodotti Agroalimentari Tradizionali (PAT)**⁹: a collection of Sicilian products included in the national list of the Prodotti Agroalimentari Tradizionali (P.A.T) (Traditional Agri-Food Products).
- **Registro delle Eredità Immateriali (REI)**¹⁰: data from the Regional Register of Intangible Heritage of Sicily (Registro delle Entità Immateriali Siciliane - R.E.I.S.)

2.3. Data Quality and Interoperability Challenges

The qualitative analysis of the selected Sicilian Open Data reveals that the datasets selected are largely complementary and provide a rich and articulated coverage of heterogeneous ICH elements. Table 1 reports the quantitative and qualitative characteristics of the selected datasets.

Information in the datasets include typical Sicilian agri-food products (e.g., *Maiorchino*, Sicilian cheese, *Bastarduna di Calatafimi*, Sicilian prickly pear produced in Calatafimi) and various local preparations and food processing techniques, as well as information on places closely associated with the production or consumption of these products (e.g., *Anciòva sutta sali*, salt-packed anchovies), or linked to emblematic regional recipes. In addition, they document intangible practices related to food and conviviality, popular and ancient celebrations, religious rites (e.g., *Lu Signuri di li fasci*, local celebration on the Holy Friday), artisanal traditions (e.g., *Barche in Legno Trezzote*, wooden boats), and orally transmitted knowledge.

All the selected datasets are released in different formats, including the .ttl format (Turtle – Terse RDF Triple Language), which expresses data in RDF in a textual, readable, and compact form compliant with

⁸<https://dati.regione.sicilia.it/catalogo/2e14f4a7-3c8a-4269-a885-7ed1257cc28e>

⁹<https://dati.regione.sicilia.it/catalogo/prodotti-agroalimentari-tradizionali>

¹⁰<https://dati.regione.sicilia.it/catalogo/33a33ca5-a7d3-46f4-8b62-a0f3457765bb>

Table 1

Overview of the selected datasets

Dataset	# Entities	Information	Licence	Formats
LSG	308	Entity Name, Location Name, Location ID, Coordinates, Description, Image	CC-BY-4.0	GeoJson, json, .csv, .ttl
PAT	260	Entity Name, Production Area, Protected Name	CC-BY-4.0	.json, .csv, .ttl
REI	239	Entity Name, Location Name, Location ID, Description, Webpage	CC-BY-4.0	.json, .csv, .ttl
Total	807			

W3C standards for the representation of structured data on the Web, and are structured according to the Sicily Region Ontology¹¹.

Although machine-readable and partially structured according to Semantic Web principles, the datasets still pose some challenges since they do not fully implement effective interlinking mechanisms with authoritative external resources. The limited connection to controlled vocabularies, shared ontologies, or reference datasets (e.g., DBpedia, Wikidata) - with the exception of the use of the controlled vocabulary for Italian municipalities - reduces the degree of semantic interoperability and hinders the integration of these data within distributed environments. Additionally, some datasets lack essential information, making the represented data incomplete and, in certain cases, only partially usable for advanced processing and integration tasks. Missing attributes, such as detailed descriptions, precise geospatial references, temporal information, or standardized identifiers for cultural entities, limit the ability to perform accurate semantic alignment. These gaps also reduce the potential for meaningful cross-dataset comparisons, automated reasoning, and the creation of user-oriented services, since key contextual elements needed to fully characterize cultural heritage items are absent. Furthermore, links to external multimedia resources, such as images, are often non-functional, with the sole exception of web references to the Regional Register of Intangible Heritage.

According to Tim Berners-Lee's five-star Linked Open Data evaluation model¹², these datasets meet the requirements of the fourth level: they are published in open and structured formats (e.g., RDF/Turtle), yet they do not fully participate in the Web of Data, as they do not establish explicit links with other resources in the Linked Data cloud.

As a result, the datasets require targeted enrichment efforts, including format alignment, data integration to address incompleteness and the lack of interlinking, as well as conceptual alignment and structural interoperability, in order to restore missing information, ensure conceptual coherence, and enhance their overall interoperability within a broader LOD ecosystem. This scenario highlights significant room for improvement, both in terms of ontological design and in the dataset's openness toward the global LOD ecosystem.

¹¹<https://dati.regione.sicilia.it/lode/extract?url=https://dati.regione.sicilia.it/owl/regionesiciliana.rdf>

¹²<https://5stardata.info/en/>

3. Methodology

To solve the challenges posed by the datasets (Cfr. section 2.3), we developed a methodological pipeline composed of sequential and complementary steps. The datasets undergo a series of processing stages: (i) interlinking, (ii) semantic mapping and format conversion and (iii) enrichment.

Once these stages are completed, the processed datasets converge within the SICHELIA Context Broker, which continuously interacts with the DataHighway Project Context Broker. This bidirectional exchange of federated datasets ensures that the data remain up-to-date, harmonized, semantically interoperable, without data duplication.

Finally, we developed the SICHELIA Web App which retrieves the semantically enriched open data from the SICHELIA Context Broker, where they are exposed in NGSI-LD format. The application then leverages this harmonized knowledge base to deliver advanced functionalities —such as contextualized storytelling, geolocated cultural itineraries, and interactive exploration of intangible heritage elements— thus reusing open data with the aim of meaningfully offer digital experiences about ICH for end users (citizen, tourists, local governments and administrations, travel agencies, cultural heritage institutions).

3.1. Interlinking

Interlinking represents a fundamental strategy for enabling semantic connections between heterogeneous datasets. By establishing explicit relationships between entities across different sources, interlinking facilitates data integration, enhances discoverability, and allows for richer, more complex queries. Typical strategies include the use of shared identifiers, standardized vocabularies, and common ontological classes, as well as the alignment of entities through geographic, temporal, or thematic properties. In the context of CH, interlinking enables the creation of a unified view of related tangible and intangible assets, supporting both research and interactive applications [19].

The LSG and REI datasets can be readily interconnected through a shared value represented by the IRI: <https://dati.regione.sicilia.it/onto/regionesiciliana/FeatureLocation>.

Our methodology relies on leveraging information on entities present in both datasets based on the property `hasFeatureLocation`, which represents a geographic entity defined in the Sicily Region Ontology. This class present as value a unique numeric code established by the Controlled Vocabulary of Italian Municipalities¹³.

For example, in both the LSG and REI datasets, there are entities that share the value 087002 for the property `hasFeatureLocation` which corresponds to the municipality of Aci Castello in the province of Catania (CT), Sicily, which can be linked thanks to the shared location.

Regarding the PAT dataset, the available information is relatively sparse and does not contain any reference to the `FeatureLocation` class. Consequently, PAT lacks geographic information for its entities. However, to address this critical gap, Section 3.3 presents an enrichment strategy from external sources, enabling subsequent interlinking with the LSG and REI datasets.

3.2. Mapping and Conversion

Format conversion strategies are essential to ensure that heterogeneous datasets, originally published in diverse formats, can be integrated into a coherent Semantic Web framework. These strategies typically

¹³<https://github.com/italia/daf-ontologie-vocabolari-controllati/blob/master/VocabolariControllati/territorial-classificazioni/cities/cities.ttl>

involve transforming data from CSV, JSON, or XML into RDF-based formats such as Turtle, RDF/XML, or JSON-LD, which support machine-readable semantics, standard URIs, and explicit modeling of relationships. Format conversion not only enhances interoperability but also allows datasets to fully participate in Linked Data applications, enabling advanced querying, reasoning, and integration with external semantic resources.

As part of the DataHighway project, datasets collected for the SICHELIA Projects must be compliant with the NGSI-LD¹⁴ standard [20] in order to be interoperable in the data federation ecosystem.

To support the conversion of the selected datasets from the .ttl format to the JSON-LD format of the NGSI-LD standard, a conceptual and semantic mapping was developed between the entities and properties of the Sicily Region Ontology and the entity types and attributes of NGSI-LD, enabling a subsequent semi-automatic transformation of the considered datasets. This activity involved:

- Aligning ontology entities with NGSI-LD types (@type)
- Mapping RDF properties to NGSI-LD attributes (Property, Relationship)
- Structuring semi-automatic transformation rules using Natural Language Processing (NLP) techniques for IRI, literal values, and geographic references (GeoProperty)

Our methodology focuses on re-using entity types and attributes that are already defined in the NGSI-LD Context Broker of the DataHighway Project (e.g., “image,” “website,” “description,” “location,” etc.). In other cases, new entity types and attributes were introduced for mapping purposes, with a preference for reusing elements coming from standard ontologies, such as schema.org and DBpedia.

Although the NGSI-LD format, as described in the ETSI specifications¹⁵, provides comprehensive coverage of some domains, such as environment, no precise models currently exist for the CH domain, making conversion more challenging. Furthermore, the diverse and complex forms of ICH (aspects related to movement, dance, music and lyrics, site-specific performances, food products) often demand multiple ontological frameworks to adequately describe all associated entities [1].

Consequently, ad hoc entities were proposed and subsequently linked to corresponding DBpedia entities through an initial automatic matching process followed by manual validation.

Table 2 and Table 3 presents some examples of the semantic mapping between the entities of the datasets selected (SicilianOntology), the NGSI-LD entity types and attributes, and their correspondence with DBpedia¹⁶ and Agrovoc¹⁷ entities.

In the context of SICHELIA Project, a methodological approach was adopted to ensure semantic interoperability across data sources. Specifically, during the transformation of the original datasets into .ttl format, each occurrence of the property `hasFeatureLocation` was systematically replaced with the `isLocatedAt` relationship, modeled according to the Relationship type defined by the NGSI-LD standard. This enables direct referencing of geographic entities (cities and provinces) already available within the DataHighway Broker. Semantic interlinking is achieved through the use of pre-existing URNs, allowing direct and unambiguous connections to the corresponding NGSI-LD entities, such as `urn:ngsi-ld:City` or `urn:ngsi-ld:Province`. Conversely, the entities (types) stored in the SICHELIA Broker, provided by the DataHighway project team, can be accessed and linked within the DataHighway Broker. This enables the implementation of a federation between NGSI-LD brokers, ensuring decentralized data availability and allowing queries and entity retrieval from multiple sources.

¹⁴<https://ngsild.org/>

¹⁵https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.05.01_60/gs_cim009v010501p.pdf

¹⁶<https://www.dbpedia.org/resources/ontology/>

¹⁷<https://www.fao.org/agrovoc/linked-data>

Table 2
Entity Mapping Examples

Sicily Region Ontology Entities	NGSI-LD Entity Types	Schema
<i>Tecnica artigianale</i> (lit. Artisanal Technique)	"type": "Craft"	http://dbpedia.org/resource/Craft
<i>Tesoro umano vivente</i> (lit. Living Human Treasure)	"type": "LivingNationalTreasure"	http://dbpedia.org/resource/Living_national_treasure
<i>Parlata</i> (lit. Dialect)	"type": "Dialect"	http://dbpedia.org/resource/Dialect
<i>Paste fresche e prodotti della panetteria, della biscotteria, della pasticceria e della confetteria</i> (lit. Fresh pasta and bakery, biscuit, pastry, and confectionery products)	"type": "BakedGoods"	http://dbpedia.org/resource/Baked_goods
<i>Luoghi della produzione gastronomica</i> (lit. Gastronomic production sites)	"type": "ProductionLocation"	http://aims.fao.org/aos/agrovoc/c_6205

Table 3
Property and Attribute Mapping Examples

Sicily Region Ontology Properties	NGSI-LD Attribute Types	Schema
<code>rdfs: comment</code>	"description": { "type": "Property" "value": "[...]"}	http://schema.org/description
<code>regionesiciliana: hasFeatureLocation</code>	"isLocatedAt": { "type": "Relationship" "object": "urn:ngsi-ld:City:[...]"}	http://dbpedia.org/ontology/location
<code>regionesiciliana: hasGeometry</code>	"location": { "type": "GeoProperty", "value": { "type": "Point", "coordinates": [X,Y]}	http://dbpedia.org/ontology/location
<code>regionesiciliana: hasImage</code>	"image": { "type": "Property" "value": [link]}	http://schema.org/image
<code>regionesiciliana: hasProvince</code>	"hasProvince": { "type": "Relationship" "object": "urn:ngsi-ld:City:[...]"}	http://dbpedia.org/ontology/Province

3.3. Enrichment

The enrichment of datasets represents a critical step in the development of high-quality, interoperable, and semantically meaningful knowledge bases. Dataset enrichment addresses potential gaps by integrating complementary information from authoritative documents, external datasets, controlled

vocabularies, or domain ontologies. This process not only increases the completeness and descriptive depth of the original data but also enhances their usability, interoperability, and relevance for advanced applications.

Within the SICHELIA Project, enrichment plays a pivotal role in transforming raw and fragmented cultural heritage data into a coherent, semantically rich information layer. By adding missing attributes, harmonizing representations, and expanding the descriptive content of the selected datasets, the enrichment process ensures that the resulting knowledge graph can effectively support exploration, recommendation, and the creation of innovative digital services for the valorization of Sicilian ICH. Enrichment steps were performed for the REI and PAT Datasets.

For the REI dataset, the enrichment process was implemented starting from the *Description* (Description) textual field, containing unstructured narratives related to ritual and festive practices of Sicilian ICH. The processing pipeline is structured in three successive phases with incremental validation criteria.

First, a multilingual NER model¹⁸ was employed for automatic identification of entities belonging to four semantic categories: PER (persons), LOC (locations), ORG (organizations), and MISC (miscellaneous).

Subsequently, for each extracted entity, a linking procedure to encyclopedic resources was performed through Italian Wikipedia API queries. The matching between textual mentions and Wikipedia article titles was validated through the computation of Normalized Levenshtein Similarity [21, 22], applying an acceptance threshold of ≥ 0.80 . Subsequently, for each extracted entity, a linking procedure to encyclopedic resources was performed through Italian Wikipedia API queries. The matching between textual mentions and Wikipedia article titles was validated through the computation of Normalized Levenshtein Similarity [21, 22], applying an acceptance threshold of ≥ 0.80 . Out of a total of 3,160 extracted entities, 2,836 (89.7%) were successfully linked to their corresponding Wikipedia articles, while 324 (10.3%) remained unlinked. A manual inspection of the unlinked textual spans revealed that the majority correspond to entities not automatically resolvable, such as dialectal toponyms or local micro-toponyms (e.g., *Chiesa du u Ratu*, *Chiantu di Mari*), which constitute candidates for subsequent manual curation interventions.

Finally, in order to reduce false positives arising from referential ambiguity, a contextual control phase was introduced on the 2,836 linked entities. For each linked entity, the metadata of the corresponding Wikidata item were analyzed, verifying the presence of explicit references to the Sicily Region or its related provinces and municipalities in structured fields (P131 - located in administrative territorial entity, P17 - country, descriptions). This verification enabled disambiguation of homonymous entities by privileging interpretations consistent with the geographic context of the corpus.

The contextual filtering process revealed that 2,751 entities (97%) either corresponded to geographical references already present in the knowledge broker (cities, localities, provinces) or pertained to entities outside the project's thematic scope (e.g., *Impero Romano*). The remaining 85 entities (3%) constituted novel, domain-relevant entries successfully integrated into the knowledge base. Among these, a representative example is *Sampaoloni*¹⁹ (dialectal: *Sampauluna*), the traditional papier-mâché processional giants depicting the eleven apostles that characterize the Easter Sunday celebration in San Cataldo (Caltanissetta).

For the PAT dataset, a significant source for extension and enrichment is offered by the PAT Sicilia

¹⁸<https://huggingface.co/Babelscape/wikineural-multilingual-ner>

¹⁹<https://it.wikipedia.org/wiki/Sampaoloni>

2023 technical sheet²⁰, published in PDF format on the institutional portal of the Sicilian Region within the section dedicated to Traditional Agri-Food Products (23rd Update of the List of Traditional Agri-Food Products). This document contains the same 260 entries included in the PAT dataset, complemented with additional information that can be used to enhance and complete the dataset.

The enrichment strategy leveraged the product name field to perform Wikidata entity linking. This approach was enabled by the existence of a dedicated Wikipedia page²¹ listing Sicilian traditional agri-food products, which served as the primary source for entity matching. A strict validation criterion requiring 100% string matching between the PAT object name and the corresponding Wikipedia entry was applied, resulting in successful enrichment of approximately 35% of PAT objects (91 out of 260 elements).

Challenges still persist in the enrichment phase; for example, the *production area (territorio interessato alla produzione)* is provided as a free-text string (e.g., Calatafimi, Vita) and not as an identifier code according to a controlled vocabulary. As a strategy to overcome this issue, we automatically retrieved the corresponding geographic coordinates and searched for the associated city or municipality within the DataHighway Broker, establishing a direct relationship to the corresponding URN.

Furthermore, since in the free-text sources we retrieved also alternative names in the Sicilian dialect about local products (e.g., Italian: *Gambero rosso* (red shrimp), Sicilian Dialect: *àmmaru rùssu, ammarùni*), we employed the attribute type `alt_name` (<http://schema.org/alternateName>), which is already available within the DataHighway context.

4. SICHELIA Web App

The final phase of the project, focused on the implementation of the SICHELIA web application prototype²² that leverages the enriched LOD in JSON-LD format from the SICHELIA Context Broker.

The Web App is designed to provide an intuitive and accessible user interface for the visualization and exploration of open data. As highlighted by many scholars [23, 24, 25], data visualization allows a better discovery and exploration and facilitates access to data for users. For this reason we chose to display the selected, enriched and converted open data by means of a geographic representations of data through an interactive map.

Users can select one or more domains of interest from a predefined filter-list (e.g., Music, Food and Beverages, Craftsmanship, etc.) or perform a free-text search through a dedicated search bar. Based on the expressed preferences and interests, the corresponding points of interest are displayed on the map, each of which is interactive. By selecting a point, users can access detailed informative cards containing descriptions, curiosities, and historical insights related to culinary traditions, artisanal techniques, popular and religious festivities, local knowledge, and traditional customs linked to the Sicilian ICH. The interface, thus, offers a personalized and flexible user experience by providing information, descriptions, and contextual insights about the ICH of Sicily.

SICHELIA targets experiential tourists and slow travelers interested in the authentic discovery of local areas, traditions, food and wine heritage, and craftsmanship. The target audience includes both national and international visitors who are attentive to sustainability and conscious tourism, as well as

²⁰<https://www.regione.sicilia.it/istituzioni/regione/strutture-regionali/assessorato-agricoltura-sviluppo-rurale-pesca-mediterranea/dipartimento-agricoltura/servizi/brand-sicilia-marketing-territoriale/riconoscimento-prodotti-qualita/prodotti-agroalimentari>

²¹https://it.wikipedia.org/wiki/Prodotti_agroalimentari_tradizionali_siciliani

²²https://nlpgroup.unior.it/dahlia_nlp/sichelia_discovery/

Sicilian local communities wishing to rediscover their own heritage and strengthen proximity-based economies. The solution is also relevant for cultural and creative enterprises, local administrations, and tourism operators, as it provides an integrated tool for enhancing territorial assets, promoting public–private collaboration, and supporting initiatives for regeneration and economic development.

From a social impact perspective, SICHELIA fosters cultural storytelling and local development by preserving collective memory, and reinforcing community identity. It further ensures continuity and visibility for traditions at risk of disappearing, especially among younger generations. Furthermore, SICHELIA promotes a model of sustainable and distributed tourism centered on the enhancement of local economies, helping redirecting visitor flows toward less-traveled areas and small municipalities, generating direct economic benefits for micro-enterprises, artisan workshops, and local farmers, while raising awareness of circular and environmentally responsible production models.

4.1. Web App usability test

The application prototype was subjected to structured tests aimed at evaluating its performance, stability, and overall usability. The tests included both functional checks (for the correct behavior of the main features) and usability tests, carried out through sessions with real users, with the goal of observing the interaction with the interface and identifying any obstacles in the user experience by simulating realistic usage scenarios.

The test was designed following the theoretical models and best practices in the field of usability evaluation. Specifically, established tools such as the Questionnaire for User Interface Satisfaction (QUIS) [26], the USE Questionnaire [27], and the Website Analysis and Measurement Inventory (WAMMI)²³ were used as references.

The usability test (see Appendix A) was conducted via a written anonymous online survey administered to focus groups composed of real users, mainly students in the 20–25 age range. Participants were invited to use the web app intensively in both a controlled context, with direct supervision, to observe their interaction with the web app, and in an uncontrolled environment. At the end of the sessions, they were asked to complete the survey, which was structured to gather both quantitative and qualitative feedback. The survey included closed-ended Likert scale questions to measure the degree of satisfaction, ease of use, interface clarity, and perceived overall efficiency, in addition to open-ended questions designed to gather free observations, suggestions, and reports of any issues encountered.

The main objective of the survey was to identify areas for improvement in the Web app from the user experience perspective, by evaluating the intuitiveness of the design, the consistency of navigation, and the comprehensibility of the features. The collected data were subsequently analyzed to highlight recurring patterns.

The survey results (see Appendix A), based on feedback collected from 100 testers over 1 month, confirm a good level of general user satisfaction with the web app. The majority of participants reported a positive user experience. The main features were considered very easy to locate (~58%), the interface was perceived as intuitive (~61%), and the response time was judged to be very fast (~62%). Furthermore, most respondents indicated that the application mostly (~40%) or completely (~36%) met their expectations, and that they would recommend it to other users and would like to see it expanded to other Italian regions. No technical problems or bug were reported (~92%) however, a small number of users experienced issues when zooming in and out on the map while using tablet devices. The

²³<https://www.wammi.com/questionnaire.html>

most appreciated features were the exploration of descriptions, information, and fun facts related to ICH (~77%) and the visualization of points on the map (~75%). Regarding suggested improvements, approximately ~30% of participants proposed the inclusion of multimedia content (e.g., images, videos, and audio), ~15% recommended enhancements to the visual design, and ~10% suggested the use of additional color schemes for category differentiation. Survey results' will be taken into consideration to improve the application in the future version releases.

5. Conclusion and Future Works

The integrated activity of interlinking, mapping, conversion, and semantic enrichment led to the production of datasets fully compliant with the NGSI-LD data model, which have been exposed and published within the Context Broker provided for the SICHELIA project.

Through this process, the original open data were thoroughly restructured and enhanced, resulting in datasets that were in the end:

- linked to external entities already present in the information system, such as cities and provinces, through the use of shared URNs;
- interoperable, as they are modeled according to a shared standard and ready to be consumed by NGSI-LD-based services and applications;
- application-oriented, with particular relevance to smart tourism, territorial valorization, and intelligent reuse of CH data and OGD;
- semantically enriched.

The semantic enrichment activity resulted in a significant increase in the number of RDF triples present in the original .ttl datasets. This growth in informational content not only reflects a substantial improvement in data quality but also enables new scenarios for integration and reuse, in line with the principles of semantic interoperability and open data.

The federation between NGSI-LD brokers ensures decentralized data availability and allows queries and entity retrieval from multiple sources. This strategy supports the system's horizontal scalability, reduces the risk of bottlenecks, and promotes a distributed governance model for territorial data, in line with the principles of semantic, syntactic, and organizational interoperability. It serves to interconnect multiple data brokers so that they can cooperate, exchange information, and respond to distributed queries while remaining independent and autonomous.

The testing phase in a relevant environment (realistic simulation) ensures that the project reached a TRL 5 and demonstrated that the technology works as expected under conditions simulating real use, preparing the ground for full-scale testing and demonstrations.

The SICHELIA project, through the implementation of a user-friendly web app, aims at tackling the long-standing challenge of open data reuse by demonstrating how creatively leveraging datasets produced by governments and institutions can generate new cultural, social, and economic value.

Future enhancements will focus on improving the interface design for more intuitive navigation, personalized recommendations, and richer visualization, as resulted from the survey responses.

Finally, the SICHELIA methodology provides a replicable model for other case studies on ICH, showing how semantic enrichment, interlinking, and federated data infrastructures can be applied to other contexts, fostering knowledge sharing, preservation, and sustainable valorization of CH.

Acknowledgments

This project was carried out within the framework of the DataHighway Project, “Creation of a Linked Open Data platform for the development of innovative applications for tourism and cultural promotion,” funded by the Programma Operativo Complementare (POC) Sicilia 2014–2020 – Axis 1 “Supporting competitiveness and sustainable, innovative digital transformation,” Action 1.1.1 “Qualified investments for the development and strengthening of research and innovation capacity” – Completion of the ranking for Action 1.1.5 of the PO FESR Sicilia 2014/2020.”

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-4 and Grammarly in order to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

- [1] M. Ziku, Digital cultural heritage and linked data: Semantically-informed conceptualisations and practices with a focus on intangible cultural heritage, *LIBER Quarterly: The Journal of the Association of European Research Libraries* 30 (2020) 1–16.
- [2] E. Jones, M. Seikel, *Linked data for cultural heritage*, London: Facet (2016).
- [3] E. Hyvönen, *Publishing and using cultural heritage linked data on the semantic web*, volume 3, Morgan & Claypool Publishers, 2012.
- [4] M. Danese, N. Masini, M. Biscione, et al., *Gli open data per il patrimonio culturale: aspetti teorici ed esperienze in italia* (2024).
- [5] H. Park, R. P. Smiraglia, Enhancing data curation of cultural heritage for information sharing: A case study using open government data, in: *Research Conference on Metadata and Semantics Research*, Springer, 2014, pp. 95–106.
- [6] Á. Gomes, D. Soares, Open government data initiatives in europe: northern versus southern countries analysis, in: *Proceedings of the 8th international conference on theory and practice of electronic governance*, 2014, pp. 342–350.
- [7] J. Attard, F. Orlandi, S. Scerri, S. Auer, A systematic review of open government data initiatives, *Government information quarterly* 32 (2015) 399–418.
- [8] A. Quarati, Open government data: Usage trends and metadata quality, *Journal of Information Science* 49 (2023) 887–910.
- [9] E. Ruijter, S. Grimmelikhuijsen, J. Van Den Berg, A. Meijer, Open data work: understanding open data usage from a practice lens, *International Review of Administrative Sciences* 86 (2020) 3–19.
- [10] V. Weerakkody, Z. Irani, K. Kapoor, U. Sivarajah, Y. K. Dwivedi, Open data and its usability: an empirical view from the citizen’s perspective, *Information Systems Frontiers* 19 (2017) 285–300.
- [11] G. Santos-Hermosa, A. Quarati, E. Loría-Soriano, J. E. Raffaghelli, Why does open data get underused? a focus on the role of (open) data literacy, in: *Data cultures in higher education: Emergent practices and the challenge ahead*, Springer, 2023, pp. 145–177.
- [12] R. Ramachandran, K. Bugbee, K. Murphy, From open data to open science, *Earth and Space Science* 8 (2021) e2020EA001562.

- [13] G. Tropea, L. Longo, A. Detti, Data-highway: a linked open data platform for e-tourism in sicily, in: 10th Italian Conference on ICT for Smart Cities and Communities 18-20 September, 2024 | University of Messina, Dept. Engineering - Messina, Italy, 2024.
- [14] U. I. C. Heritage, P. Rii, Convention for the safeguarding of the intangible cultural heritage, in: Proceedings of the Report of the Eleventh Annual Coordination Meeting of Category, volume 2, 2020.
- [15] U. I. C. Heritage, What is intangible cultural heritage, 2019.
- [16] S. Lazaro Ortiz, C. Jimenez de Madariaga, The unesco convention for the safeguarding of the intangible cultural heritage: a critical analysis, *International journal of cultural policy* 28 (2022) 327–341.
- [17] K. Wasela, The role of intangible cultural heritage in the development of cultural tourism, *International Journal of Eco-Cultural Tourism, Hospitality Planning and Development* 6 (2023) 15–28.
- [18] M. Zain El Shandidy, The power of intangible heritage in sustainable development, *International Journal of Advanced Studies in World Archaeology* 6 (2023) 92–97.
- [19] C. Bizer, T. Heath, D. Ayers, Y. Raimond, Interlinking open data on the web, in: Demonstrations track, 4th european semantic web conference, innsbruck, austria, 2007.
- [20] G. Privat, A. Medvedev, Guidelines for modelling with ngsl-d, ETSI White Paper 42 (2021).
- [21] I. Koumarelas, A. Kroschk, C. Mosley, F. Naumann, Experience: Enhancing address matching with geocoding and similarity measure selection, *Journal of Data and Information Quality (JDIQ)* 10 (2018) 1–16.
- [22] E. Kaldeli, A. Chortaras, V. Lyberatos, J. Liartis, S. Kantarelis, G. Stamou, Combining automatic annotation with human validation for the semantic enrichment of cultural heritage metadata, in: CHR, 2024.
- [23] S. Giacomini, A. Bardi, M. Buzzoni, M. Daquino, F. Fischer, C. Martignano, G. Rubin, F. Tomasi, et al., Atlas: Towards a knowledge graph of international scholarly research on the italian digital cultural heritage, in: Proceedings of the 21st Conference on Information and Research science Connecting to Digital and Library science, 2025, pp. 1–13.
- [24] G. Renda, M. Daquino, et al., Learning the semantic web with tools for information visualisation and data storytelling, in: CEUR WORKSHOP PROCEEDINGS, volume 3937, CEUR Workshop Proceedings, 2025, pp. 1–9.
- [25] T. Battisti, M. Daquino, Exploring data-driven narratives in digital humanities web-based projects: features and impact, 2025.
- [26] B. D. Harper, K. L. Norman, Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5, in: Proceedings of the 1st annual mid-Atlantic human factors conference, volume 224, 1993, p. 228.
- [27] A. M. Lund, Measuring usability with the use questionnaire12, *Usability interface* 8 (2001) 3–6.

A. Online Resources

GitHub Repository: All supporting materials, including the survey questions, the aggregated survey results, data transformation code and data examples are available at the project GitHub Repository: <https://github.com/dahliaspinoff/sichelia>