DIGITAL HERITAGE (2025)
S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors)

Designing and developing *ICH Atlas*: A Trend Identification Platform for Digital Valorization of Intangible Cultural Heritage

Chiara Di Lodovico[†1], Arianna Bertolino[2], Davide Spallazzo[3], Ilaria Bollati[3], and Greta Cappellini[2]

[1]University of Milan, Cultural and Environmental Heritage Dept., Italy
[2] Icoolhunt S.p.A, Italy [3] Politecnico di Milano, Design Dept., Italy

Abstract

Social media-based trend identification platforms powered by Artificial Intelligence (AI) are gaining considerable attention from scholars and industry professionals alike. These platforms analyze large amounts of user-generated data, identify the emergence and evolution of practices, and support data-driven decision-making. Although automatic trend analysis is growing, limited attention has focused on developing platforms that target digital transformation practices in Intangible Cultural Heritage (ICH). This paper introduces the development process behind ICH Atlas, a digital platform that displays and enables dynamic navigation of emerging trends connected to ICH-related professional roles, skills, and digital technologies and their expected growth based on global trends. The platform was the result of nine-month remote collaboration between a university-based design research group with expertise in cultural heritage and a trend forecasting company specializing in artificial intelligence (AI) and big data analytics. Based on a qualitative analysis of meeting records, email correspondence, datasets, and platform prototypes, the paper outlines the platform's iterations. We trace the decision-making process development, encountered challenges, and coping strategies. Based on our reflections, we identify three tensions that might be of interest for industry-academia initiatives in the intangible cultural heritage sector: scaffolding vs ambiguity, interpretation vs granularity of data, tacit vs explicit knowledge.

CCS Concepts

• ***Computing methodologies*** → *Information extraction;* • ***Human-centered computing*** → *Collaborative and social computing systems and tools;*

1. Introduction

In recent decades, trend identification has become increasingly important for a broad spectrum of cultural actors, from industry professionals to academic researchers. A trend is commonly defined as *"a topic area that is growing in interest and utility over time,"*. Trends can be detected using systems that analyze large corpora of textual data to identify novel or increasingly prominent topics [KGP*04]. These systems — now frequently powered by artificial intelligence and big data analytics — allow for real-time monitoring of shifting dynamics across sectors, facilitating timely insights and supporting data-driven, informed decision-making [DPL17].

In the field of Intangible Cultural Heritage (ICH), researchers have begun applying trend detection to map scholarly interest evolution and identify knowledge gaps [AGDR21, LP23]. These studies usually draw on structured datasets (such as peer-reviewed publications) filtered through standardized keyword systems established within the research community. While this method works well for analysing academic discourse, it may fall short in capturing the informal, rapidly evolving practices and discourses surrounding ICH, especially those outside the academia boundaries.

To our knowledge, no prior study has applied trend identification systems to analyze social media data to detect emergent digital initiatives related to ICH safeguarding practices. Yet, social media platforms represent a fertile ground for such inquiries [GS22]. They provide access to informal, real-world conversations and practices - including those emerging from grassroots actors or underrepresented communities - which may signal broader cultural shifts [VGA12, ZSOW22, HHP18]. In light of the growing *"digital imperative"* [vH10b] and the increasing pressure on cultural actors to embrace the digital transition [Dig], social media–based trend analysis offers a promising opportunity for identifying emerging skills, technologies, and initiatives by scraping data from posts related to ICH and digital topics.

This paper presents the decision-making process behind *ICH Atlas*, a platform specifically conceived to address this purpose. The platform is the outcome of a remote interdisciplinary collaboration

† Corresponding author

DOI: 10.2312/dh.20253189

between a design research group from Politecnico di Milano with expertise in heritage and Icoolhunt, a trend forecasting company specializing in AI and big data. Over the nine-month period, the two teams followed a learning-by-making, iterative approach. Each iteration of the platform was refined through the combined expertise of both teams.

The paper is structured as follows. The *Related Works* section provides an overview of the digital transformation in ICH safeguarding practices and introduces the need for resources able to capture emerging jobs and skills in this evolving field, followed by an overview of the features of trend identification platforms. Next, we outline the scope of the paper and the methods used to gather insights into the platform design process. The *Results* section presents the latest version of the platform, traces its iterations, and offers a critical reflection on the collaborative dynamics between different disciplinary domains, highlighting challenges and coping strategies. The paper concludes by discussing a series of tensions that emerged during the development of the platform.

2. Related works

2.1. The digital turn in Intangible Cultural Heritage safeguarding practices

Safeguarding Intangible Cultural Heritage (ICH) presents complex challenge due to its inherently immaterial and dynamic nature. Unlike tangible heritage, which is preserved through physical conservation, ICH encompasses living practices, oral traditions, performing arts, languages, and craftsmanship that are constantly evolving within communities [UNE03]. While ICH often manifests in tangible outputs, its core meaning and value transcend physical form [Bou03]. UNESCO's *"Representative List of the Intangible Cultural Heritage of Humanity"* and *"List of Intangible Cultural Heritage in Need of Urgent Safeguarding"* reflect the growing international recognition of the need to protect these ephemeral forms of cultural expression [vH10a].

Historically, ICH transmission and survival relied on oral transmission and embodied practice, passed down from generation to generation within communities of practice. However, globalization, urbanization, and shifting social structures have increasingly disrupted this model, raising concerns about the long-term survival of several ICH practices [LLX25, Ste25]. In response to these challenges, a growing body of research has explored the potential of digital technologies to support the documentation, storage, presentation, dissemination, education, and revitalization of ICH, while also fostering new forms of community engagement [ABKT*17, RR15, CDS*18, HKP*22, LDC*23].

While early efforts often focused on the creation of digital archives to preserve and share cultural materials, more recent projects are experimenting with advanced technologies and cross-sectorial collaborations, expanding the possibilities offered by digital preservation. For example, the Icelandic government recently partnered with OpenAI and local language technology companies to better integrate Icelandic into conversational AI systems. The project uses Reinforcement Learning from Human Feedback (RLHF) involving local volunteers to train OpenAI's multimodal large language model, GPT-4. The goal is to reduce grammatical errors, improve cultural accuracy, and keep Icelandic relevant in the digital space [Ope23]. Another initiative, the *Hong Kong Martial Arts Living Archive (HKMALA)*, developed through a collaboration between academic institutions and a cultural association, uses digital technologies such as motion capture and 3D reconstruction to document and analyze traditional Southern Chinese martial arts. The project, active since 2012, led to the development of a digital repository that supports research, educational activities, and public interactive exhibitions [CDS*18, HKP*22]. Furthermore, the *Quantum Temple* platform exemplifies the use of emerging Web3 technologies for heritage safeguarding. The platform promotes sustainable cultural tourism and equitable partnerships with Indigenous communities using blockchain-based systems. Its NFT collection *"The Paths to Alango"* - co-created with tradition bearers, cultural practitioners, and anthropologists - digitally represent Balinese rituals and traditions while generating income for local communities. Complementing this, the *QT Passport* allows travelers to trace their contributions to heritage preservation and community projects through transparent, blockchain-verified donations [AS23, Tem].

These initiatives are not isolated cases. They reflect a broader phenomenon known as the *digital imperative* [vH10b], which is also affecting the field of Intangible Cultural Heritage (ICH). Policies and funding frameworks increasingly prioritize digitization as a key strategy for enhancing the preservation, accessibility, and engagement of ICH. While this shift often overlooks the challenges and biases inherent in digital-driven practices [vH10b, ABKT*17], it prompts many stakeholders in the ICH ecosystem to embrace digital transformation and seek the skills and partners needed to support it. This process accelerated further during and after the COVID-19 pandemic [AAKWK21].

In response to this need, several initiatives within the EU Erasmus+ framework have been launched to identify emerging professional roles and digital skills potentially useful to support the safeguarding of cultural heritage in the digital era. The *CHARTER (European Cultural Heritage Skills Alliance) project* [CHA], for example, mapped core and transversal competences, including digital, technological and green adaptation skills in the fields of cultural heritage safeguarding and preservation, crafts and traditional knowledge, dissemination and communication, knowledge, planning and management. While the *Digicult project* [Dig] developed an e-learning platform to support professionals in upskilling in the digitization, presentation, and preservation of ICH assets.

While these initiatives — among others — are highly valuable, they primarily rely on qualitative inquiries and quantitative surveys, which are often time-consuming and centered on established actors. This raises a key question: *How might we develop a resource capable of automatically identifying emerging trends related to ICH, including emerging skills, job roles, and digital technologies, using a more diverse and global corpus of data that incorporates both formal and informal actors?* Before turning to the methodology and findings addressing the research question, the next section examines trend identification platforms in cultural and research contexts.

2.2. Trend Identification Systems

Trend identification is a critical practice in various fields, particularly in fashion, design, and visual culture, as it involves detecting emerging signals that may indicate the future evolution of tastes, practices, or styles. Traditionally, this task was carried out by professionals known as "coolhunters," individuals skilled at spotting new trends by observing subcultures and predicting their rise [P*10]. This practice, often referred to as trend spotting, is a subset of trend analysis and has been widely adopted within the fashion industry [STT22, DL20]. More recently, the scope of trend identification and forecasting has expanded to other fields, including the job market (e.g., The World Economic Forum's *"The Future of Jobs Report"* [For25]), technological innovation (e.g., McKinsey's *"Technology Trends Outlook"*), scientific research [BMA*21], and cultural heritage (e.g., *"IFLA's Trend Report"* by [DO24]).

In this context, the advent of social media and advancements in big data and AI have significantly transformed the trend identification process. On the one hand, the widespread use of social networking platforms allows users to share posts and topics - images, texts, hashtags — in an ongoing and potentially public manner, generating an unprecedented volume of data. On the other hand, advancements in AI technologies have enhanced the ability to analyze big data, identify correlations, patterns, and the evolution of tastes and practices over time, making trend identification faster and more efficient [DL20]. In this context, organizations and communities in academia and industry alike are leveraging this wealth of data to better understand communities' evolving interests and support data-driven decision-making [BMA*21].

Emerging Trend Detection (ETD) systems are tools that automatically identifies new patterns, topics, behaviors, or changes happening within a large amount of data over time [KGP*04]. These systems are usually developed through a systematic process encompassing iterative cycles of data collection, preprocessing, analysis, and visualization [STT22] and can operate as fully automatic or semi-automatic [KGP*04]. Fully automatic systems analyze textual data independently to identify and visualize potential emerging topics, which are then reviewed by humans. Semi-automatic systems, on the other hand, start from user-defined inputs and provide evidence to assess whether those topics are truly emerging.

The process typically begins with data collected from sources like social media, news outlets, and market reports. This data undergoes preprocessing to improve consistency, reduce noise, and eliminate redundancies [STT22, BMA*21]. Analytical techniques are then applied, combining statistical methods, machine learning, and Natural Language Processing tools [LGA15]. Among these, topic modeling and topic clustering are commonly used to uncover hidden patterns. Topic modeling is an unsupervised technique that automatically discovers topics within text by analyzing the co-occurrence of words in a collection of documents [AK24]. Topic clustering entails the use of algorithms to group similar documents or data points together based on their features or textual content [Gev02, MZ05]. To enhance trend identification, data analysis incorporates tracking the evolution of clusters over time [BMA*21]. Finally, results are visualized through dashboards and reports to aid decision-making.

For example, Sleiman et al. [STT22] developed a semi-automated system to detect emerging fashion trends from French online sources. The platform collects and structures textual data via web scraping, followed by pre-processing steps such as sentence segmentation, word normalization, and removal of duplicates. It employs NLP techniques and deep learning methods to extract syntactic patterns and word relationships. A specialized fashion dictionary categorizes the extracted terms by attributes like garment type, color, and material. Trend detection is based on analyzing term frequency and co-occurrence, with results visualized as network graphs showing the evolution and interrelation of fashion features over time.

While trend identification systems based on social media data have been widely adopted in sectors such as fashion [SC23], their application within the domain of ICH remains limited. Existing research has primarily focused on mapping ICH-related trends and knowledge gaps within academic literature [AGDR21, LP23], or on analyzing affective and cognitive dynamics in online heritage-sharing practices [GS22]. However, there has been little focus on the use of social media data to extract trends related to emerging skills and professional roles in the ICH sector.

We acknowledge the methodological and ethical challenges associated with leveraging digital content — such as the prompt-driven nature of data, which raises concerns about representativeness and temporality [ZSOW22, SC23, GSBC09]. Nevertheless, we align with Roigé et al. [RAUS21], who argue that, *"[d]espite these drawbacks, observation of the online space is very useful in a field such as ICH,"* given its prolific nature and the abundance of content it generates. However, while the volume of available digital content is vast, the development of effective platforms for social media-based trend identification remains far from straightforward [ZSOW22]. Therefore, in this paper, we provide insights into the design rationales, decision-making processes, and collaborative problem-solving strategies employed by two interdisciplinary teams in the development of a data-driven platform for ICH+Digital trend identification. Through a critical analysis of this process, we aim to offer a set of reflections that can inform the design of future trend identification tools in the ICH domain.

3. Methodology

This study critically reflects on the development of a platform designed to detect emerging trends related to professional roles, skills, technologies, and ongoing projects within the ICH domain.

The methodology combined data collection, system development, and iterative evaluation. The data collection process combined multiple sources to document the evolution of the platform. Meetings between the two teams were recorded and transcribed enabling the systematic reconstruction of the platform's development trajectory. These documents served also as a basis for ongoing reflection on the collaboration dynamics. Email correspondence, mainly used for exchanging datasets, meeting minutes, summaries of new features, and iterative reviews of prototypes and progresses, further contributed to documenting the assessments of platform performance and perceived accuracy.

The collected materials were analyzed to uncover both the platform's design decisions and the challenges encountered throughout its development. Using a structured coding approach [Sal21], we first reconstructed the development process chronologically, following the platform's iterative phases, and annotated the corresponding evaluation outcomes, highlighting the changes implemented and the rationale behind them. In a second phase, we conducted a qualitative analysis to identify recurring challenges and the corresponding coping strategies within the collected textual materials. Building on prior studies that conceptualize conflicting dynamics as "tensions" [Lod23], we developed a set of reflections that illuminate the opposing forces complicating both the platform's development and the process of trend identification. These forces were rooted not only in how the domain of Intangible Cultural Heritage is articulated by its diverse stakeholders, but also in the disciplinary differences between the collaborating teams.

4. Results

After a brief description of the *ICH Atlas* platform final version's features, this section reconstructs the platform development phases, showing how design choices and analytical strategies co-evolved through iterative refinement and cross-disciplinary negotiation.

4.1. The platform

ICH Atlas is a web-based platform that displays and enables dynamic navigation of emerging trends connected to ICH- related professional roles, skills, and digital technologies and their expected growth based on global trends. Currently, the platform provides insights derived from the analysis of 1,883 Twitter, Instagram, and Linkedin posts published over the past three years by 284 identified users. The social media content is sourced from publicly accessible profiles representing a broad range of actors within the ICH ecosystem. These include cultural institutions and associations, design studios and artistic collectives, policy and decision-makers, technology providers, journalists and magazine, professionals and researchers in ICH and digital humanities, academic institutions, and freelance practitioners. The selection of these categories is grounded in previous research that identifies them as key stakeholders in the evolving field of ICH and digital innovation [SC15, Lup21].

The dataset is compiled through an iterative multi-step process. First, posts are collected from public profiles. To ensure linguistic consistency across the corpus, all posts are translated into English. Then, based on a preliminary vocabulary of terms, a targeted query methodology is adopted to filter and extract data containing keywords related to ICH (its practices, representations, and domains) as well as emerging digital technologies. The resulting subset of posts is subsequently processed using GPT-4. GPT-4 was integrated into the extraction pipeline due to its capacity to address semantic heterogeneity at scale, its advanced contextual understanding for detecting indirect or emergent references, and its prompt-driven architecture, which enabled flexible and modular querying across diverse entity types. In our project, we adopted GPT-4 for semi-automatic extraction and labeling of key *topics* associated with ICH, professional skills, job roles, events, and technologies. In this context, a "topic" refers to a single word or short phrase (e.g., "creative technologist").

Specifically, due to the vast conceptual space and the evolving vocabulary of digital transformation, traditional rule-based approaches or simple pattern matchers were insufficient. Instead, we employed a hybrid approach involving:

- *GPT-4-based semantic extraction* to identify latent mentions of technologies (e.g., AR/VR, machine learning algorithms), skills (e.g., metadata curation), and events (e.g. webinars, conferences..).
- *Semantic post-processing* to validate and refine the outputs produced by GPT-4 and to ensure accurate alignment with domain-specific terminology. To do so, we employed text embedding models, which encode contextual meaning, to compute semantic similarity between extracted items and a controlled vocabulary. This process was essential for automatically confirming the relevance of candidate entities (skills, technologies, or events) to the domain of ICH.
- *Iterative prompt development* tailored for each task (e.g. data extraction, classification, clustering), refined progressively to balance specificity and generalization.
- *Iterative monitoring of hallucinations* (generation of semantically plausible but factually incorrect information) through manual review of GPT-extracted entities on a stratified sample and confidence scoring based on model output probabilities and semantic alignment.

Finally, the extracted topics function as "weak signals" reflecting early indicators of emerging discourse within the ICH community. To assess their broader significance, signals are analyzed using a proprietary trend forecasting system that tracks millions of social media interactions. This comparative analysis enables the identification of potential future developments relevant to the ICH sector.

The platform is organized into three primary sections, each designed to support different modes of exploration and analysis:

- **ICH Domains**. This section offers a static clustering of posts and topics categorized according to the official domains of intangible cultural heritage, as defined by UNESCO (2003) - namely, oral traditions, customs and social practices, performing arts, knowledge and practices concerning nature and the universe, and traditional craftsmanship. For each domain, the platform displays the number of posts tagged accordingly, a set of associated keywords, the most frequent contributors, as well as identified skills, job roles, events, and technologies. Additionally, it features a curated selection of ten representative posts per domain.
- **Search**. This section enhances transparency in the labelling process by displaying the extracted keywords from each post, categorized according to user category, skills, job roles, technologies, and events. It also allows performing keyword-based searches and accessing the corresponding posts in which the terms appear.
- **Trend Navigation**. This section presents the results of a Semantic Topic Clustering Analysis, enabling users to explore identified topics grouped by skills, job roles, or technologies, as well as in broader hybrid clusters that combine these elements. Users can adjust the granularity of the clustering, generating between 1 and 25 topic groups. Furthermore, the section highlights the relative growth rates of each topic over time. As representative visualizations, Figure 1 displays eight emergent trends within the

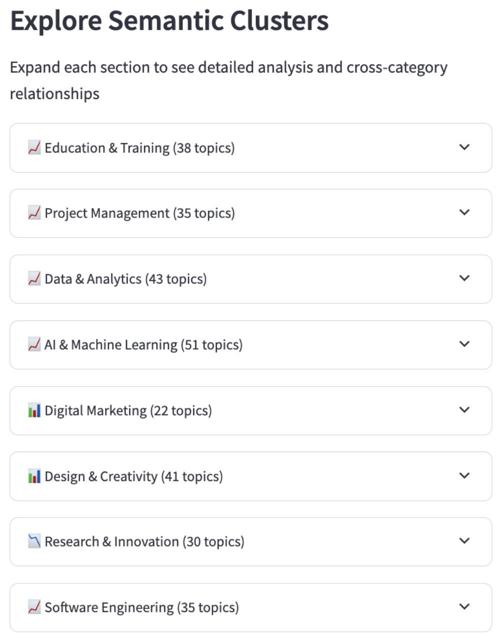

Figure 1: *Trend Navigation. Eight emergent skills clusters.*

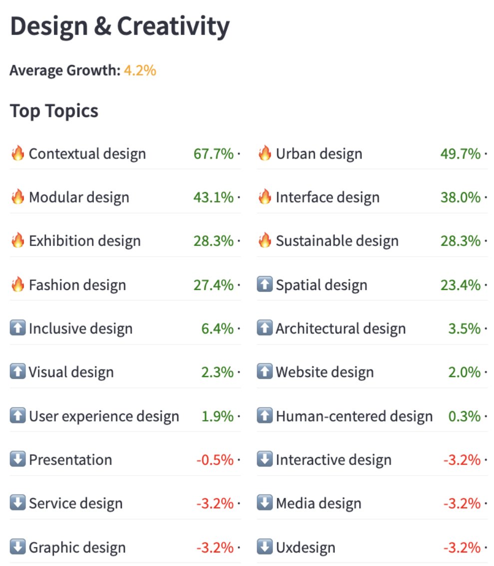

Figure 2: *Trend Navigation. "Design and creativity" skills cluster.*

Figure 3: *Trend Navigation. Contextual relationships within the "design and creativity" cluster, including relevant events and projects, professional roles, technologies and an exemplar post.*

skills domain; Figure 2 provides a detailed view of the "design and creativity" cluster, showcasing emerging topics and their projected growth based on global trend data; Figure 3 illustrates relevant topics linked to the "design and creativity" skills cluster, in relation to events and projects, professional roles, technologies, and an exemplary post.

4.2. The decision-making process behind the platform development

Following a review of platform development meetings and the collaboration between the research team and the technology company, we identified several key iterations that shaped the evolution of the trend identification platform. These iterations highlight how challenges were progressively addressed through a combination of technological adjustments and conceptual refinements.

4.2.1. Initial exploration: Twitter-based themes detection

The initial phase focused on establishing a shared understanding between the two teams (from now the "research team" and the "technology team"). Reciprocal presentations were organized: the research team introduced the conceptual background of intangible cultural heritage (ICH), while the technology team presented examples of their prior trend forecasting projects.

As a starting point for identifying emerging trends, on indication of the technology team, the research team provided a set of 50 public social media profiles (including Twitter, Instagram, Linkedin, and Youtube) and relevant ICH and digital-related keywords. The technology team subsequently focused on collecting posts from identified profiles, limiting the scope to Twitter for convenience.

The content extraction process was based on a targeted query methodology applied to Twitter and Instagram, leveraging a data extraction and analysis system previously developed by the company. Posts were initially filtered using a seed list of keywords explicitly referencing both “intangible cultural heritage” (e.g. ICH, intangible cultural heritage, living heritage, folklore, handicraft, cultural celebrations, rituals, heritage preservation) and digital-related concepts (e.g., digitization, virtual museum, 3D reconstruction, digital storytelling). After the first pass, false positives (irrelevant or ambiguous mentions) were identified and excluded through manual review and automated scoring filters. Subsequently, an iterative enrichment loop was applied: new relevant terms and concepts encountered in the validated posts were appended to the keyword list, allowing the system to progressively expand its sensitivity to the domain-specific discourse. Only terms confirmed through co-occurrence with known ICH markers and validated through semantic scoring were retained. Finally, data were analysed through topic modelling and clustering techniques.

This process led to the refinement of the keyword list, a distribution analysis of profile typologies (e.g., cultural institutions, policymakers, research organizations), and the identification of four emerging conversation clusters. These clusters included: (1) the evolution of ICH-related professional roles and skills; (2) growing efforts to engage broader and younger audiences with intangible cultural heritage; (3) the adaptation of digital tools to support and promote ICH; and (4) the increasing use of interactive and spatial exhibition formats to reimagine and transmit ICH. Each theme was illustrated through a GPT-4–generated description, a circular chart depicting the distribution of profile typologies, an exemplar post from the dataset, and a corresponding word cloud. For example, the word cloud for cluster (1) featured terms such as “Museum Professionals,” “Cultural Leaders,” “Capacity Building,” “Digital Skills,” “Mentorship Program,” and “Professional Development,” among others.

While the identified clusters aligned with the research team’s understanding of the state of the art in the sector, the insights provided and the format of the cluster analysis raised several questions. Specifically, the description of emerging themes often appeared to reflect tendencies toward overgeneralization and overinterpretation, leading to apparent inconsistencies. Furthermore, the most common questions raised during the feedback session by the research team were: “What is the process behind this insight?,” “How was it obtained?,” and “How can we be certain that this relates specifically to ‘Intangible Cultural Heritage’ and not to ‘Cultural Heritage’ in a broader sense?”. To address these uncertainties, the research team requested access to the ’raw’ data underpinning the analyses, while the technology team asked for a more refined set of keywords to guide the identification process toward more precise representations of intangible cultural heritage-related topics.

4.2.2. Community enlargement: Iterative profiles identification

Following the initial exploration, efforts were directed toward expanding the community of relevant social media profiles. The technology team broadened the search by validating an initial set of core keywords and developing targeted search queries on Twitter and Instagram to identify additional relevant profiles. These queries were iteratively refined based on emerging results. Through this process, the technology team compiled a dataset consisting of 158 Twitter and Instagram profiles and 390 posts, extracted leveraging GPT-4, which were then shared with the research team for review. The manual validation process curated by the research team identified 102 profiles and 54 posts as potentially relevant, with classifications distinguishing between those focused on ICH, digital content only, or both.

Despite these efforts, a significant limitation emerged. The dataset extracted solely from Twitter and Instagram, while providing spontaneous, temporally dynamic interactions reflecting public and organizational engagements, lacked the thematic depth and granularity necessary to capture the nuances of ICH and the specific technologies being adopted. In response to this challenge, the research team requested the incorporation of Linkedin data to enrich the dataset. Leveraging the profiles identified in earlier phases, the technology team implemented an ad hoc pipeline to extract data from Linkedin and conducted targeted search queries to retrieve potentially relevant posts related to ICH and digital technologies. A semi-automated process, assisted again by GPT-4, facilitated the selection of 556 posts and 353 profiles.

Following this filtering stage, the research team carried out a review and manual validation phase of the dataset. The review indicated that LinkedIn was particularly effective in capturing professional discourse, events, technological applications, and institutional narratives related to intangible cultural heritage (ICH). When combined with data from Twitter and Instagram, this platform mix provided the semantic richness and accessibility necessary for scalable textual analysis. Advantages that, according to the technology company’s prior experience, were not equally present on platforms such as TikTok and Facebook.

The review process led to the inclusion of 129 posts and 79 profiles, expanding the dataset further. Although this expansion improved the dataset's breadth, challenges persisted regarding the thematic alignment between post contents and ICH digital valorization practices. Specifically, core profiles and posts related to ICH remained scarce, and the dataset continued to include posts largely focused on tangible cultural heritage, or general cultural and artistic initiatives that, while containing some of the keywords used for the search (e.g., "artisans," "community identity"), lacked direct references to ICH manifestations. By the conclusion of the community enlargement process, the dataset comprised 284 profiles and 1,883 posts.

4.2.3. Back-end platform development and restructuring

In parallel with the community identification efforts, and in response to the research team's request for greater transparency, the technology team developed a preliminary back-end interface that enabled the former to autonomously navigate the dataset from multiple perspectives and levels of granularity. The platform consisted of three main sections:

- an overview of the *ICH Atlas* community, presenting distributions by profession and post volume over time;
- a topic detection section, allowing the exploration of 25 emerging themes clustered according to the categories of ICH, Digital, and ICH+Digital, each accompanied by a number (1–25), a title, a GPT-4 powered automatically generated summary, top keywords, a word cloud, and distribution by profession;
- a free text search section for posts, with filters based on thematic areas and metadata concerning users, events, organizations, locations, and mentioned skills.

A qualitative assessment conducted by the research team spotted interesting clusters (e.g. "Empowering Indigenous Communities Through Decentralized Governance and Blockchain Technology") but surfaced several critical issues. Overlaps between thematic clusters and insufficient granularity (e.g. the clusters "Digital Transformation in Cultural Heritage" and "Digitization and Preservation of Cultural Heritage"), and inconsistent representations of ICH topics (e.g. in clusters like "Machine Learning in Financial Fraud Detection and Cybersecurity") undermined the interpretability and utility of the insights. In response to the challenges previously identified, the research team proposed a series of improvements grounded in UNESCO's categorization framework and a descriptive-analytical model for ICH valorization cases.

This approach informed the restructuring of thematic clusters in alignment with recognized ICH Domains, and according to three phases of the ICH valorization process, as outlined in the CHARTER model [CHA]. In the latter case, the goal was to organize topics into three distinct categories: those pertaining exclusively to insights only related to ICH manufestations (*"ICH Recognition* phase); topics related to digitization efforts aimed at capturing, modeling, protecting, and conserving intangible heritage (*"ICH Conservation and Preservation"* phase); and practices focused on the revitalization of ICH, such as community engagement initiatives, digital and virtual exhibitions, and entrepreneurial or educational efforts, (*"ICH Access, Enhancement, and Use"* phase).

To gain deeper insights into professional roles and skills, the research team requested the triangulation of data across skills, professions, technologies, and events. Building on this refined conceptual framework, the platform underwent significant restructuring. Posts were reorganized into (1) *"ICH Domains"*, following UNESCO's classifications rationale [UNE03], and (2) *ICH Cluster phases* CHARTER-inspired phases described above [CHA]. Additionally, an enhanced (3) *"Search"* function was implemented, allowing filtering results by both valorization phases and ICH domains, thereby improving navigability and increasing the analytical depth of the platform.

4.2.4. Introduction of trend navigation and forecasting

In conjunction with structural reorganization, a new feature entitled *Trend Navigation* was implemented. This section provided an overview of more than 150 emerging topics specifically related to skills, professional roles, and technologies. Illustrative examples include "contextual design" and "metadata analysis" under skills; "digital heritage preservationist" and "XR developer" for professional roles; and "Extended Reality (XR)" and "mobile application" under technologies. Each topic was accompanied by an automatically generated summary and was connected to related skills, roles, technologies, and corresponding ICH domains, thereby enabling a more interconnected and fine-grained exploration of data.

Unlike earlier platform versions — where clusters emerged solely from the initial dataset — the technology team incorporated their proprietary trend forecasting platform, which monitors millions of social media posts (e.g., Twitter, Instagram, Reddit, Youtube, etc.) related to innovation, trends, and styles. Specifically, emerging topics identified within the project's limited dataset were treated as "weak signals" and were automatically compared with broader consumer conversations. Topics displaying resonances with ongoing wider trends were interpreted as an indicators of likely future relevance for intangible heritage. This approach was guided by the aim of overcoming the limitations associated with identifying "known insights," based on the principle that if a trend had already gained substantial visibility within ICH-specific discourse, it was likely approaching a state of saturation.

Despite these advancements, several critical issues persisted. Overlaps and inconsistencies were noted within the ICH Domains and Cluster Phases sections, resulting in fragmented representations of limited datasets. Notably, the limited volume of available profiles and posts, combined with the imposition of an "artificial" scholarly framework, emerged as a further constraint. The ways in which ICH was discussed in posts often did not allow for unambiguous codification into distinct valorization phases, rendering strict categorizations only marginally useful. On the other hand, while the automated topic summaries risked oversimplification, raising concerns about potential misinterpretations, the individual emerging trends and their possible aggregation into broader categories were regarded as promising and worthy of continued exploration.

4.2.5. Semantic clustering approach

In light of these results, both teams agreed to move away from clustering posts by ICH valorization phase. Additionally, they agreed to remove automated descriptions of topics in the "Topics navigation section". The technology team reorganized emerging trends related

Proceedings published by Eurographics - The European Association for Computer Graphics.

to skills, professional roles, and technologies into broader semantic clusters, with the possibility to dynamically adjust the number of clusters to display (1-25). Each topic, whether related to a skill, profession, or technology, was initially assigned to a macro-category (e.g., the skill "contextual design" was classified under the macro-category Design & Creativity; see Figure 2) based on semantic similarity. Within each macro-category, topics were further organized into smaller sub-clusters, again using semantic proximity as a guiding criterion. Each macro-category was enriched with metadata, including its title, average growth rate, top associated topics, growth trend comparisons, related entities (e.g., job roles, skills, and technologies), and links to relevant events and projects (see Figure 3). Representative social media posts were also included to provide accessible, contextual entry points into each trend. This semantic aggregation represented a significant advancement in the platform's development, fostering a more balanced exploration of emerging trends and topics. Furthermore, the iterative refinement of ICH-related keywords and dataset semi-automated labelling process contributed to improved — albeit still improvable — clustering within the "ICH Domains" section, thus expanding the analytical capabilities available to future platform users.

5. Discussion

Throughout the development of the platform, several key tensions emerged, particularly regarding the integration of theoretical frameworks with real-world data, over-interpretation vs high data granularity, as well as the interplay between explicit and tacit knowledge. These tensions were critical to the evolution of both the platform and the research process.

5.1. Scaffolding vs ambiguity

A central issue was the tension between the analytical scaffolding provided by established theoretical frameworks — such as the CHARTER valorization phases and UNESCO's ICH categories — and the ambiguous, heterogeneous nature of the ICH discourse in social media contexts. These frameworks, while offering necessary structure, often fell short in capturing the multifaceted ways in which ICH is referenced, enacted, and discussed by practitioners, institutions, and communities online.

In particular, social media conversations rarely adhered to formal distinctions embedded in institutional taxonomies. Posts frequently blurred boundaries between ICH and other domains, such as tangible heritage, contemporary art, or digital culture. This blurred representation complicated attempts to classify content accurately and consistently. For example, mentions of "craft," "language," or "community identity" often appeared without explicit reference to ICH categories or terminologies, challenging the operational application of such frameworks for automated tagging and clustering. Moreover, the analysis revealed that terms like "ICH" or "intangible heritage" were disproportionately used by policy-oriented actors or academic institutions, risking biasing the dataset toward profiles that shared institutional or scholarly vocabularies, while underrepresenting grassroots or community-led discourse.

A further layer of complexity concerned terminological ambiguity, an ongoing issue in text mining efforts [FJO08]. Indeed, the same keyword —such as "learning," "tradition," or "memory" — could vary significantly in meaning. This not only posed several challenges in terms of false positives, as well as during filtering and automatic clustering. To address this issue, several cycles of manual reviews were implemented to detect such data and reframe instructions to perform more refined topic extraction cycles.

These processes revealed important insights: (1) while frameworks serve as valuable cognitive and methodological anchors, their direct application to messy, real-world data may risk over-simplification or misclassification, (2) consistent effort should be dedicated to refine the dataset considering terminological ambiguity, especially in such a niche sector. Therefore, the platform's evolution gradually moved toward more flexible, data-driven clustering methods that could better accommodate semantic ambiguity and emergent patterns.

5.2. Data over-interpretation vs over-fragmentation

Another key tension emerged around the balance between interpretive synthesis and data granularity. As the dataset expanded to include a wider range of social media profiles and posts, maintaining the accuracy and relevance of insights became increasingly complex. One of the primary challenges involved avoiding the risk of over-interpretation or misrepresentation, particularly in the use of automated text generation tools — a concern raised in recent critical studies on AI-assisted analysis [AYAK25].

Indeed, the initial attempts to generate automated topic descriptions were critically assessed for occasionally producing over-generalized or over-interpreted description — results that, although plausible, lacked visible justification for domain experts. As perviously mentioned, this led to recurring feedback questions from the research team such as: "What is the process behind this insight?"

This kind of question points to a broader epistemological issue in AI-powered systems: the opacity of machine-generated insights often clashes with the users' need for traceability and context [DLTDCR25, ZPWHY23].

The iterative restructuring of the platform — removing automated summaries, enabling access to original posts and assigned labels, and implementing user-controllable clustering — was a direct response to this issue. Moreover, the teams opted to reduce the degree of interpretive automation, instead providing more granular data (in the form of "topics") around the skills, professional roles, and technologies categories. However, this shift introduced novel difficulties as highly granular data, while apparently precise, can become analytically overwhelming or cognitively inaccessible for users attempting to extract broader patterns. Finding the optimal balance between interpretive abstraction and data specificity proved to be one of the most demanding aspects of the platform's design. Ultimately, the teams implemented a dynamic aggregation mechanism and semanthic-based data triangulation and clustering to allow greater adjustment possibilities on the information level of detail and resolution.

5.3. Tacit Knowledge vs Explicit Knowledge

A further critical tension emerged around the relationship between tacit and explicit knowledge during the platform development process.

Initially, much of the technological work — such as data extraction, cleaning, and clustering procedures — heavily relied on the tacit expertise of the technology team. This reliance made it difficult for the research team to fully grasp the underlying criteria and assumptions informing data processing and analysis. In response, the research team progressively demanded clearer documentation and greater operational transparency. This led to the development of a dedicated exploratory platform, specifically designed to allow the research team to interact directly with the datasets, observe clustering mechanisms, and adjust parameters as needed.

However, despite these efforts, disciplinary differences and distinct forms of expertise between the two teams continued to hinder full mutual understanding. On the one hand, it became apparent that the technology team, responsible for the data coding system, struggled to fully comprehend the nature of intangible cultural heritage (ICH) — a domain characterized by blurred boundaries even for specialized professionals [Sch17]. Although the research team provided informational materials and conceptual frameworks, these resources proved insufficient for effectively conveying the embodied and situated knowledge held by the research team. In retrospect, they may have underestimated the complexity involved in transferring disciplinary expertise and overestimated the self-evident nature of the materials provided. On the other hand, the tacit knowledge held by the technology team also presented significant obstacles. Despite the sharing of procedural system details, the research team — lacking direct access to the very labeling and analysis system and unfamiliar with its underlying workflows — often found it difficult to conceive alternative approaches to data processing and visualization. As a result, their efforts were largely guided by trial-and-error methods and dependent on the mediation of the technology team.

Given their background in design research, the team later engaged with the more transparent version of the platform experimentally, using the tool itself to progressively discover and understand its possible issues and applications. While this experimental approach supported a deeper understanding of the available datasets, the inability to directly access the developer's system significantly limited their capacity to fully comprehend the system's affordances and ultimately contributed to slowing down the platform's development process.

Finally, throughout this iterative process, the platform and its various data exploration sections progressively came to function as a *boundary object* [LS10]: an artifact sufficiently flexible to be interpreted differently by each team, yet robust enough to maintain a shared structure that supported collaboration. The gradual increase in transparency regarding post classifications enabled both teams to progressively develop a deeper understanding of each other's domain, overcoming initial perceptions of ICH and AI as "black boxes" from the perspective of the respective groups. Interestingly, writing the current paper further contributed to enhancing transparency and shared understanding.

6. Study limitations

We acknowledge two main limitations in our research. First, the platform currently operates on a relatively small dataset, and the limited number of processed posts and profiles has, at times, influenced the depth and reliability of the analytical outputs. Second, we recognize that social media content offers only a partial view of the broader landscape of ICH-related initiatives, as it excludes actors who do not use these platforms to disseminate their work.

7. Acknowledgements

The paper presents one of the outcomes of the project *TRAMA. Cultural Transformations: Intangible Heritage and New Digital-Driven Professional Skills*, which was selected through a cascading funding scheme under the broader Italian PNRR project *CHANGES – Cultural Heritage Active Innovation for Sustainable Society*, funded by the European Union's Next Generation EU program. The cascading grant aimed to promote and enhance mechanisms of technology transfer and knowledge exchange between industrial and academic partners. The collaboration is governed by institutional agreements between the university and the company, which have been assessed and authorized by the university's ethics committee.

References

[AAKWK21] AMANKWAH-AMOAH J., KHAN Z., WOOD G., KNIGHT G.: Covid-19 and digitalization: The great acceleration. *Journal of business research 136* (2021), 602–611. 2

[ABKT*17] ALIVIZATOU-BARAKOU M., KITSIKIDIS A., TSALAKANIDOU F., DIMITROPOULOS K., GIANNIS C., NIKOLOPOULOS S., KORK S. A., ET AL.: Intangible cultural heritage and new technologies: Challenges and opportunities for cultural preservation and development. In *Mixed Reality and Gamification for Cultural Heritage* (2017), Ioannides M., Magnenat-Thalmann N., Papagiannakis G., (Eds.). 2

[AGDR21] ARTEAGA J. V., GRAVINI-DONADO M., RIVA L. Z.: Digital technologies for heritage teaching: trend analysis in new realities. *International Journal of Emerging Technologies in Learning (iJET) 16*, 21 (2021), 132–148. 1, 3

[AK24] AMIRI B., KARIMIANGHADIM R.: A novel text clustering model based on topic modelling and social network analysis. *Chaos, Solitons & Fractals 181* (2024), 114633. 3

[AS23] ADAMI L., SIGMARINGO N.: Quantum temple: A new vision for utilizing blockchain technology for the recognition and preservation of cultural heritage. In *International Conference on Human-Computer Interaction* (2023), pp. 3–14. 2

[AYAK25] ARAKAWA R., YAKURA H., AKUZAWA K., KUBO S.: Ai for meeting minutes: Promises and challenges in designing human-ai collaboration on a production saas platform. In *CHI'25 Extended Abstracts, Yokohama, Japan* (2025). 8

[BMA*21] BEHPOUR S., MOHAMMADI M., ALBERT M. V., ALAM Z. S., WANG L., XIAO T.: Automatic trend detection: Time-biased document clustering. *Knowledge-Based Systems 220* (2021), 106907. 3

[Bou03] BOUCHENAKI M.: *The interdependency of the tangible and intangible cultural heritage*. Tech. rep., ICOMOS 14th General Assembly and Scientific Symposium, 2003. 2

[CDS*18] CHAO H., DELBRIDGE M., SHAW J., KENDERDINE S., NICHOLSON L.: Kapturing kung fu: Future proofing the hong kong martial arts living archive. In *Digital Echoes: Spaces for Intangible and Performance-based Cultural Heritage*. Cambridge Scholars Publishing, 2018, pp. 249–264. 2

[CHA] CHARTER: European cultural heritage skills alliance. URL: https://charter-alliance.eu/. 2, 7

[Dig] DIGICULT: Digital presentation and preservation of intangible cultural heritage. URL: https://digicultproject.eu/en/. 1, 2

[DL20] DUBREUIL M., LU S.: Traditional vs. big-data fashion trend forecasting: an examination using wgsn and edited. *International Journal of Fashion Design, Technology and Education 13*, 1 (2020), 68–77. 3

[DLTDCR25] DI LODOVICO C., TORRIELLI F., DI CARO L., RAPP A.: How do people develop folk theories of generative ai text-to-image models? a qualitative study on how people strive to explain and make sense of genai. *International Journal of Human–Computer Interaction* (2025), 1–25. 8

[DO24] DEZUANNI M., OSMAN K.: *IFLA Trend Report 2024: Facing the future of information with confidenc*. Tech. rep., Digital Media Research Centre, 2024. 3

[DPL17] D'AVANZO E., PILATO G., LYTRAS M.: Using twitter sentiment and emotions analysis of google trends for decisions making. *Program 51*, 3 (2017), 322–350. 1

[FJO08] FAWAREH H. M. A., JUSOH S., OSMAN W. R. S.: Ambiguity in text mining. In *2008 International Conference on Computer and Communication Engineering* (2008), IEEE, pp. 1172–1176. 8

[For25] FORUM W. E.: *The Future of Jobs Report 2025*. Tech. rep., World Economic Forum, 2025. 3

[Gev02] GEVRY D. R.: *Detection of emerging trends: Automation of domain expert practices*. Master's thesis, Department of Computer Science and Engineering, Lehigh University, 2002. 3

[GS22] GINZARLY M., SROUR J. F.: Cultural heritage through the lens of covid-19. *Poetics 92* (2022), 101622. 1, 3

[GSBC09] GARCIA A. C., STANDLEE A. I., BECHKOFF J., CUI Y.: Ethnographic approaches to the internet and computer-mediated communication. *Journal of contemporary ethnography 38*, 1 (2009), 52–84. 3

[HHP18] HOLSAPPLE C. W., HSIAO S.-H., PAKATH R.: Business social media analytics: Characterization and conceptual framework. *Decision Support Systems 110* (2018), 32–45. 1

[HKP*22] HOU Y., KENDERDINE S., PICCA D., EGLOFF M., ADAMOU A.: Digitizing intangible cultural heritage embodied: State of the art. *Journal on Computing and Cultural Heritage (JOCCH) 15*, 3 (2022), 1–20. 2

[KGP*04] KONTOSTATHIS A., GALITSKY L. M., POTTENGER W. M., ROY S., PHELPS D. J.: A survey of emerging trend detection in textual data mining. In *Survey of text mining: Clustering, classification, and retrieval*. Springer New York, 2004, pp. 185–244. 1, 3

[LDC*23] LIU G., DING X., CAI J., WANG W., WANG X., DIAO Y., CHEN J., YU T., XU H., MI H.: Digital making for inheritance and enlivening intangible cultural heritage: a case of hairy monkey handicrafts. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (2023), pp. 1–17. 2

[LGA15] LAKHANPAL S., GUPTA A., AGRAWAL R.: Discover trending domains using fusion of supervised machine learning with natural language processing. In *2015 18th International Conference on Information Fusion* (2015), IEEE, pp. 893–900. 3

[LLX25] LI N., LI X., XIAO W.: Intangible cultural heritage threatened-level categories and criteria. *Humanities and Social Sciences Communications 12*, 1 (2025), 1–11. 2

[Lod23] LODOVICO C. D.: Exploring the tensions of self-tracking wearable technologies through design. In *International Conference on Human-Computer Interaction* (2023), Springer, pp. 30–46. 4

[LP23] LIU S., PAN Y.: Exploring trends in intangible cultural heritage design: a bibliometric and content analysis. *Sustainability 15*, 13 (2023), 10049. 1, 3

[LS10] LEIGH STAR S.: This is not a boundary object: Reflections on the origin of a concept. *Science, technology, & human values 35*, 5 (2010), 601–617. 9

[Lup21] LUPO E.: Design and innovation for the cultural heritage. phygital connections for a heritage of proximity. *AGATHÓN| International Journal of Architecture, Art and Design 10* (2021), 186–199. 4

[MZ05] MEI Q., ZHAI C.: Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining* (2005), Association for Computing Machinery, New York, NY, USA, p. 198–207. 3

[Ope23] OPENAI: *Government of Iceland. How Iceland is using GPT 4 to preserve its language*. Tech. rep., OpenAI, 2023. 2

[P*10] PEDRONI M., ET AL.: *Coolhunting. Genesi di una pratica professionale eretica*. Franco Angeli, 2010. 3

[RAUS21] ROIGÉ X., ARRIETA-URTIZBEREA I., SEGUÍ J.: The sustainability of intangible heritage in the covid-19 era—resilience, reinvention, and challenges in spain. *Sustainability 13*, 11 (2021), 5796. 3

[RR15] RODIL K., REHM M.: A decade later: looking at the past while sketching the future of ich through the tripartite digitisation model. *International journal of intangible heritage 10* (2015), 47–60. 2

[Sal21] SALDAÑA J.: Coding techniques for quantitative and mixed data. *The Routledge reviewer's guide to mixed methods analysis* (2021), 151–160. 4

[SC15] SEVERO M., COMINELLI F.: Mapping intangible cultural heritage in france. In *Valletta 2018 Conference "Cultural Mapping: Debating Spaces and Places"* (2015). 4

[SC23] SCHON V. R., COLOMBI C.: Ethical implications in ai-powered trend research platforms. *Fashion Highlight*, 1 (2023), 48–57. 3

[Sch17] SCHREIBER H.: Intangible cultural heritage and soft power–exploring the relationship. *International Journal of Intangible Heritage 12*, 44-57 (2017). 9

[Ste25] STEFANO M. L.: Threats and approaches to the safeguarding of intangible cultural heritage: A view from the united states. In *Handbook on Intangible Cultural Practices as Global Strategies for the Future*, Wulf C., (Ed.). Springer, Cham, 2025. 2

[STT22] SLEIMAN R., TRAN K.-P., THOMASSEY S.: Natural language processing for fashion trends detection. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (2022), IEEE. 3

[Tem] TEMPLE Q.:. URL: https://quantumtemple.xyz/. 2

[UNE03] UNESCO: Convention for the safeguarding of the intangible cultural heritage, 2003. Adopted at the 32nd session of the General Conference of UNESCO in Paris on 17 October 2003. URL: https://ich.unesco.org/en/convention. 2, 7

[VGA12] VAKALI A., GIATSOGLOU M., ANTARIS S.: Social networking trends and dynamics detection via a cloud-based framework design. In *Proceedings of the 21st International Conference on World Wide Web* (2012), pp. 1213–1220. 1

[vH10a] VAN HEUR B.: A definition of cultural heritage: From the tangible to the intangible. *Journal of cultural heritage 11*, 3 (2010), 321–234. 2

[vH10b] VAN HEUR B.: From analogue to digital and back again: institutional dynamics of heritage innovation. *International Journal of Heritage Studies 16*, 6 (2010), 405–416. 1, 2

[ZPWHY23] ZAMFIRESCU-PEREIRA J. D., WONG R. Y., HARTMANN B., YANG Q.: Why johnny can't prompt: how non-ai experts try (and fail) to design llm prompts. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (2023), pp. 1–21. 8

[ZSOW22] ZACHLOD C., SAMUEL O., OCHSNER A., WERTHMÜLLER S.: Analytics of social media data–state of characteristics and application. *Journal of Business Research 144* (2022), 1064–1076. 1, 3