Res Mobilis
Revista internacional de investigación en
mobiliario y objetos decorativos
Vol. 14, nº. 22, 2026

CRAFT HERITAGE IN EUROPEANA: DIGITISATION PHILOSOPHIES AND INTANGIBLE KNOWLEDGE REPRESENTATION IN ASTURIAS, LÕUNA-EESTI, AND VÄRMLAND

José Manuel Uría González*
Uriaxait S. L.

Abstract

This article presents a quantitative analysis of Europeana metadata for the regions of Asturias (Spain), Lõuna-Eesti (Estonia) and Värmland (Sweden), measuring the representation of craft heritage—in both its tangible and intangible dimensions—as conditioned by the predominant digitisation philosophy of each corpus. The three regions represent three contrasting models: text-dominant (Asturias), image-dominant (Lõuna-Eesti), and balanced (Värmland). Through multilingual dictionary-based text mining applied to more than 326,000 records across twelve lexical categories, results demonstrate that digitisation philosophy shapes both the prevalence and co-occurrence structure of craft vocabulary. Cramér's V values ranging from 0.098 (Asturias) to 0.999 (Lõuna-Eesti) reveal structurally distinct patterns of tangible–intangible co-presence, with direct implications for data accessibility in rural cultural tourism contexts.

Keywords: intangible cultural heritage; Europeana; digital humanities; craftsmanship; rural cultural tourism; digitisation philosophy; text mining; Asturias; Lõuna-Eesti; Värmland.

INTRODUCTION

The digitisation of European cultural heritage is generating documentary corpora of unprecedented scale. Aggregation platforms such as Europeana provide access to tens of millions of records from museums, archives, and libraries across Europe. Yet this accumulation is not neutral: it reflects the institutional priorities of contributing organisations, the formats that technical infrastructures privilege, and the conceptual schemas that determine what is considered catalogable.

The UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage recognises traditional arts and crafts as a central dimension of heritage deserving active protection.[1] This category encompasses craft knowledge, intergenerational transmission processes, the relationships between craft practice and landscape, and the social dimension of artisanal production—forms of

*E-mail: j.uria@uriaxait.com

knowledge that are not reducible to the produced object but constitute cultural practices with their own structure and history.

Faced with this complexity, institutional cataloguing systems tend to privilege material culture: the physical object over the process that produces it; the artefact over the knowledge that generates it; the tangible over the intangible. This structural gap has been noted in the digital heritage literature. Capurro has argued that the selection of objects for digitisation involves curatorial biases that construct specific narratives about European identity, and that there is a tension between the rhetoric of 'diversity in unity' and the technical standardisation that can homogenise local cultural practices.[2] In a network analysis of Europeana data, Capurro and Severo documented that content aggregators lack significant presence from agents linked to civil society and cultural initiatives, generating an institutionalisation bias.[3]

Haslhofer and Isaac, in their study of the Europeana Linked Open Data pilot project, showed that the technical architecture of the system—based on Dublin Core and its extensions—favours the description of physical objects over practices or knowledge.[4] Baca has likewise noted that cultural heritage metadata schemas have been historically designed by and for institutions holding collections of material objects, which structurally conditions their capacity to represent intangible heritage.[5]

To this ontological bias one must add a variable that has received comparatively less scholarly attention: the digitisation philosophy adopted by contributing institutions in each region. Not all regions digitise the same document types or with the same priorities. Some produce text-dominant corpora (archival records, inventories, catalogues); others, image-dominant corpora (photographs, illustrations, graphic material); others combine both types. This variation in the typological composition of the corpus is not merely a technical detail: it may systematically condition how craft heritage is represented, particularly in its intangible dimension.

In this context, the present article proposes the following hypothesis: the predominant digitisation philosophy of a region—oriented towards text, image, or a balance between both types—structurally conditions the representation of intangible craft heritage in Europeana. This has direct implications for data accessibility and exploitation in the domain of rural cultural tourism. Craft heritage, in both its tangible and intangible manifestations, constitutes a central resource for tourism development projects in rural areas,[6] and its presence in major European heritage data infrastructures has practical consequences for the visibility, narrativisation, and valorisation of these territories.

To test this hypothesis, the article compares three regions representing distinct typological composition models: the Principality of Asturias (Spain), with a text-dominant corpus (90.9% TEXT records); the region of Lõuna-Eesti (Estonia), with an image-dominant corpus (90.4% IMAGE); and the county of Värmland (Sweden), with a balanced composition (35.9% TEXT, 64.0% IMAGE). All three regions are part of the CULTURALITY Horizon Europe project, which analyses cultural tourism in rural European territories.[7]

The article is organised as follows. The Methods section describes the corpus, the multilingual lexical dictionaries employed, and the analytical metrics. The following three sections present results for corpus characteristics, tangible

craft heritage, and intangible craft heritage respectively. A subsequent section examines co-occurrence patterns. The Discussion section addresses the implications of findings for digital heritage policy and rural cultural tourism. A further explanation is provided as to how these findings support the working methodology of the CULTURALITY project and the implications that the presented findings have for the project's line of research. The article concludes with a summary of main contributions and limitations.

METHODS

Corpus design

The corpus comprises Europeana metadata obtained through automated region-based queries using the platform's public API.[8] Each query used the region name as the search term, subsequently filtered by country (Spain, Estonia, and Sweden respectively) to eliminate homonymous records. The procedure yielded 68,325 records for Asturias, 184,420 for Lõuna-Eesti, and 74,172 for Värmland, totalling 326,917 analysed records.

Corpus characteristics are summarised in Table 1. Typological composition differs substantially across regions. Asturias exhibits near-total concentration in TEXT records (90.9%), reflecting the weight of documentary archive collections, digitalised newspaper archives, and textual catalogues. Lõuna-Eesti shows the inverse pattern: 90.4% IMAGE records, corresponding to photographic collections, illustrations, and graphic material. Värmland presents a more balanced composition, with IMAGE as the predominant type (64.0%) alongside a significant textual presence (35.9%) and higher overall type diversity (Shannon type diversity index: 0.406, compared to 0.190 for Asturias and 0.196 for Lõuna-Eesti).

Characteristic	Asturias (Spain)	Lõuna-Eesti (Estonia)	Värmland (Sweden)
Total records	68,325	184,420	74,172
Unique titles	65,267	90,668	17,990
Providers	14	9	7
TEXT (%)	90.9	9.6	35.9
IMAGE (%)	9.0	90.4	64.0
Georeferenced (%)	2.2	28.1	54.9
Type diversity	0.190	0.196	0.406
Lexical complexity (MTTR)	0.690	0.776	0.842
H Shannon (norm.)	0.9972	0.9981	0.9999
Cramér's V Tang./Int.	0.098	0.999	0.900

T/I Ratio (global)	1.100	1.415	1.015

Table 1. General characteristics of the Europeana corpus for the three analysed regions. Typological composition differences constitute the structural factor articulating the comparative analysis.

Linguistic analysis was conducted on the description (dcDescription), title (dcTitle) and coverage (dcCoverage) fields of each record. Descriptive texts were filtered by language: Castilian and Asturian for Asturias, Estonian for Lõuna-Eesti, and Swedish for Värmland, thereby reducing bias from content analysis in languages foreign to each region.

Lexical complexity is measured using the Mean Type-Token Ratio (MTTR), a stabilised variant of the classical type-token ratio that corrects the length-sensitivity bias of the standard TTR. The formula is:

$$MTTR = (1/k) \cdot \Sigma (T_j / N) \quad (1)$$

where k is the number of sliding windows of N tokens (words) into which the corpus is segmented, T_j is the number of unique types (distinct forms) in window j, and N is the fixed window size (N = 500 tokens in this study). A value close to 1 indicates maximum lexical variety; a low value reflects high repetition of a limited vocabulary. In this analytical context, *MTTR* provides an estimate of the descriptive richness of textual metadata per region, independently of total corpus size, enabling interpretation of whether greater or lesser prevalence of craft vocabulary reflects lexically rich documentation or the repetition of a small number of recurrent terms.

Lexical dictionaries

The analysis employs multilingual dictionaries structured around two dimensions of craft heritage and six material categories: metal, wood, textile, ceramic, food, and other. The distinction between the tangible dimension (objects, materials, artefacts) and the intangible dimension (processes, knowledge, crafts) is operationalised through differentiated lexical sets in each language. For example, in Castilian the tangible-metal category includes terms such as *hierro*, *acero*, *yunque*, *forja*, *herramienta*; the intangible-metal category includes *forjar*, *herrero*, *fundir*, *templar*, *remachar*. In Swedish, the equivalents would be *järn*, *stål*, *ambolt* and *smida*, *smed*, *härda*, respectively. Complete dictionaries are available on request from the author.[9]

Keyword identification was performed using compiled regular expressions with Unicode word-boundary anchors, capturing morphological variants without generating false positives from partial matches. This dictionary-based approach offers interpretability advantages over more complex methods: each match is traceable to a specific textual occurrence, facilitating manual validation and methodological transparency.[10] Grimmer and Stewart note that dictionary methods are particularly suited when the categories of interest are well defined a priori and the relevant vocabulary is relatively stable.[11]

Term detection in metadata fields is performed through direct regular-expression counting: for each record, the presence of any dictionary term in the analysed text fields is verified using patterns compiled with Unicode word-boundary anchors. No semantic inference is applied at the detection stage: a record generates a hit in a category if and only if it contains at least one of the lexical forms explicitly included in that category's dictionary.

Analytical metrics

Three frequency metrics are employed. Normalised frequency (*NormFreq* = *total hits* / *total records* × 1,000) enables comparison across corpora of different sizes. Prevalence (percentage of records with at least one hit) distinguishes broad diffusion from thematic concentration. Mean intensity (*total hits* / *records with hits*) captures lexical density in records where the category appears.

Thematic diversity is measured by the normalised Shannon entropy index:[12]

$$H = -\Sigma_i \, p_i \ln p_i \quad \text{(normalised to theoretical maximum)} \quad (2)$$

where p_i is the proportion of hits in category i. A value close to 1 indicates uniform distribution across categories; a low value indicates thematic concentration.

The association between the tangible and intangible dimensions is assessed using Cramér's V, a measure derived from the chi-squared statistic, normalised between 0 and 1:[13]

$$V = \sqrt{(\chi^2 / N \cdot \min(r-1, c-1))} \quad (3)$$

where χ^2 is the chi-squared statistic of the tangible/intangible contingency table, N is the number of records, and $\min(r-1, c-1)$ ensures normalisation by degrees of freedom. A value close to 1 indicates that records containing tangible vocabulary also tend to contain intangible vocabulary; a value close to 0 indicates that the two dimensions appear independently.

Co-occurrence between categories is quantified through a symmetric $n \times n$ matrix whose elements M_{ij} indicate the number of records in which categories i and j co-occur. Jaccard indices (M_{ij} / (M_{ii} + M_{jj} − M_{ij})) are used for visualisation, enabling comparison of association strength independently of the absolute frequencies of each category.

RESULTS

Corpus characteristics and digitisation profile

Lexical complexity is highest in Värmland (*MTTR* = 0.842) and Lõuna-Eesti (0.776), compared to Asturias (0.690). This indicates that Värmland and Lõuna-Eesti corpora exhibit greater lexical variety in their textual descriptions, which may facilitate dictionary-based craft vocabulary capture. However, Lõuna-Eesti's corpus contains proportionally far less text than the other two, which relativises this lexical advantage. Geographic coverage also differs markedly: 54.9% of Värmland records include geographic coordinates, compared to 28.1% in Lõuna-Eesti and only 2.2% in Asturias. Figure 1 summarises the typological composition and frequency distributions across all three regions.

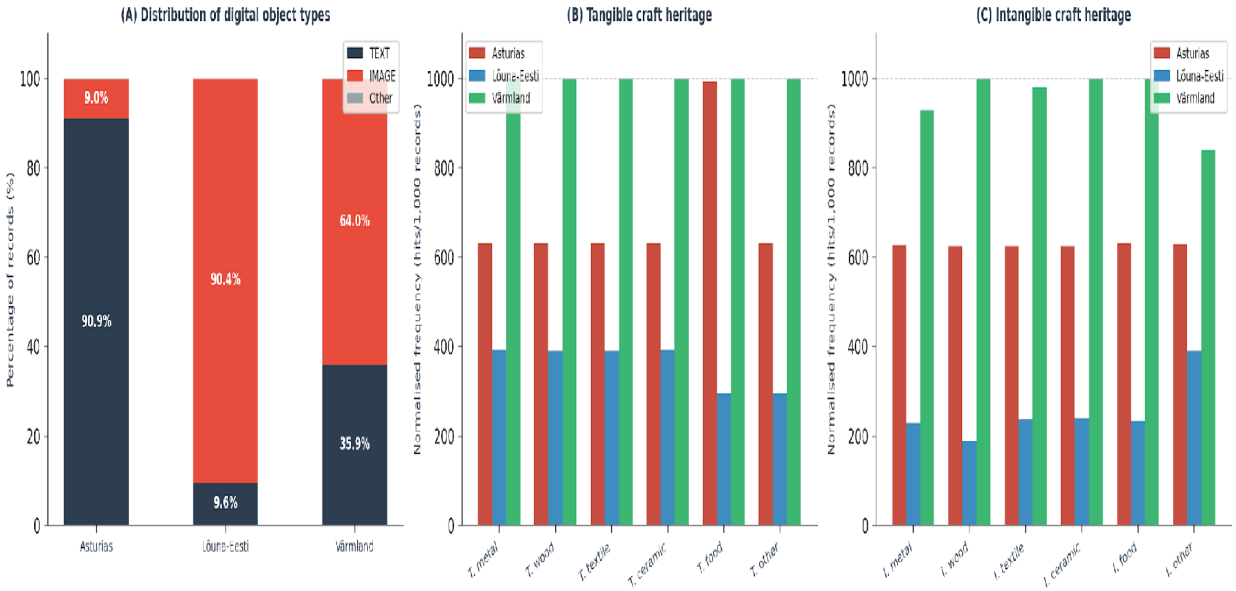


Fig. 1. Digitisation profile and craft vocabulary representation by region. (A) Typological composition of digital object types per region. (B) Normalised frequencies (hits/1,000 records) for tangible craft heritage categories. (C) Normalised frequencies for intangible craft heritage categories. Dashed lines at 1,000 indicate maximum saturation.

Tangible craft heritage representation

Normalised frequencies for tangible craft heritage categories show clearly differentiated patterns by region (Fig. 1B). Värmland achieves values close to 1,000 hits per 1,000 records in practically all categories, indicating near-total saturation of tangible vocabulary in the Swedish corpus. This saturation is consistent with the corpus's high lexical complexity and typological diversity: Värmland records tend to contain descriptions rich in terms covering multiple materials and craft objects.

A significant anomaly is observed in Asturias: the food-tangible category reaches a normalised frequency of 992.1/1,000 (prevalence 99.2%), while remaining categories cluster around 630/1,000 (63% prevalence). This anomaly reflects the weight of Asturian documentary collections with high food-vocabulary content—records related to the Cabrales cheese designation of origin, Asturian cider, regional charcuterie, and other products of the traditional agri-food economy. Importantly, this result does not necessarily reflect greater artisanal richness in food production, but rather the presence of a highly specific vocabulary in the available institutional documentation.

Lõuna-Eesti exhibits intermediate tangible frequencies, with a notable difference between material categories (metal: 392/1,000; wood: 390/1,000; ceramic: 392/1,000) and food (294/1,000) and other (294/1,000) categories. This pattern suggests that the ethnographic photographs composing this corpus are specialised in craft objects made from traditional materials (tools, ceramic utensils, textiles), while food contexts are less represented in the available photographic record.

Intangible craft heritage representation

Analysis of intangible craft vocabulary reveals more pronounced regional differences than those observed for the tangible dimension (Fig. 1C). Värmland maintains near-100% saturation across all intangible categories, with the notable exception of intangible-metal (930.9/1,000; 93.1% prevalence), which is the only case in which an intangible category falls significantly below its tangible counterpart in this corpus.

In Lõuna-Eesti, a marked anomaly is observed in the intangible-wood category: 186.9/1,000, compared to 390.2/1,000 for tangible-wood. This gap—the most pronounced in the entire analysis—suggests that craft processes linked to wood (carpentry, cabinet-making, carving) are significantly underrepresented in Estonian photographic record metadata. Images may represent the wooden object (tangible) without associated textual metadata containing vocabulary about its fabrication processes (intangible). This is consistent with the nature of the photographic object: a photograph of a carved wooden box represents the artefact but rarely describes the carving process in its metadata.

Asturias exhibits intangible frequencies in the range 625–631/1,000 across all categories, without the anomalies observed in tangible-food. This relative uniformity of intangible representation is itself a meaningful finding: Asturian textual documents appear to distribute craft process vocabulary more evenly across categories than craft object vocabulary.

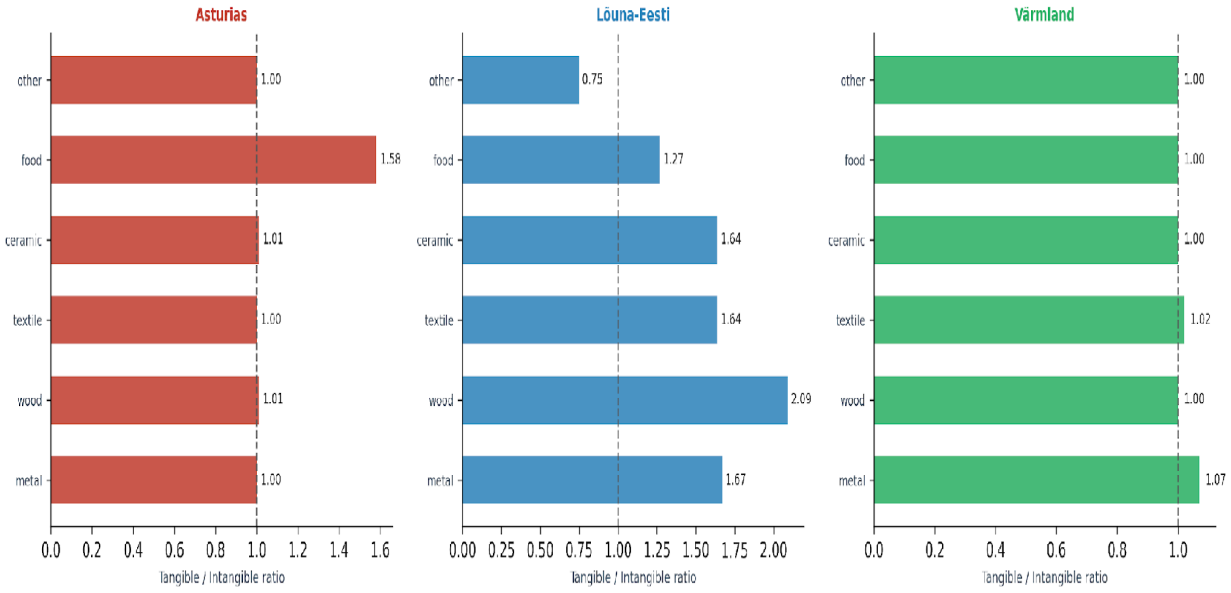


Fig. 2. Tangible/intangible ratio by subcategory and region. Values above 1.0 (dashed line) indicate tangible dominance; 1.0 indicates parity. Värmland presents the greatest equilibrium; Lõuna-Eesti shows the highest tangible bias, particularly in the wood category.

Tangible/intangible ratio and structural association

The ratio between tangible and intangible frequencies per subcategory (Fig. 2) directly visualises the degree to which material heritage exceeds or equals immaterial heritage in each corpus. The structural hypothesis predicts ratios above 1 in all regions, with greater magnitude in image-dominant corpora.

Results are consistent with this hypothesis, though with important nuances. In Lõuna-Eesti, the global ratio of 1.415—the highest of the three corpora—reaches its maximum in the wood category (ratio ≈ 2.09), confirming the gap observed in absolute frequencies. In Värmland, the global ratio of 1.015 and per-subcategory ratios close to 1 demonstrate that this corpus achieves the greatest parity between tangible and intangible dimensions. In Asturias, the global ratio of 1.100 is moderate, although the tangible bias concentrates in the food category (ratio ≈ 1.58) owing to the anomaly already described.

Shannon entropy and Cramér's V provide additional insight into the internal structure of each corpus (Fig. 3). Normalised Shannon entropy is high in all three regions (Asturias: 0.9972; Lõuna-Eesti: 0.9981; Värmland: 0.9999), indicating that the twelve categories are relatively uniformly represented in all cases. Cramér's V—which measures the association between the presence of tangible and intangible vocabulary within the same records—differs radically: Asturias (0.098), Lõuna-Eesti (0.999), and Värmland (0.900).

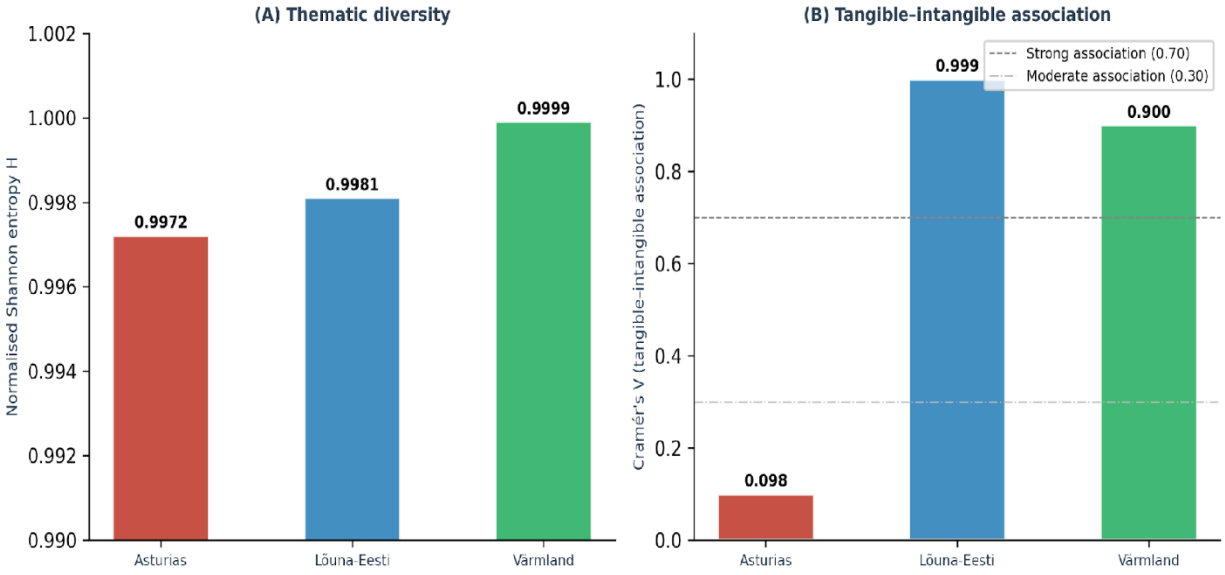


Fig. 3. Thematic diversity and tangible–intangible association by region. (A) Normalised Shannon entropy: thematic diversity is high and broadly equivalent across corpora. (B) Cramér's V: the tangible–intangible structural association varies radically according to digitisation philosophy.

This contrast is the most significant finding of the article. In Asturias, Cramér's V of 0.098 indicates a practically non-existent association between tangible and intangible vocabulary at the record level. An Asturian textual document containing vocabulary about blacksmith's tools (tangible-metal) has no significantly greater probability of containing vocabulary about the forging process (intangible-metal) than a document that does not. Texts describe either the object or the process, with scarce co-presence of both vocabularies within the same record.

In Lõuna-Eesti, V = 0.999 indicates that records containing tangible vocabulary almost invariably also contain intangible vocabulary—reflecting that textual metadata associated with photographs tends to be either very rich (containing terms from both dimensions) or very sparse (containing terms from

neither). In Värmland, V = 0.900 indicates strong association: records in the balanced corpus tend to simultaneously contain vocabulary from both dimensions.

Co-occurrence structure

Co-occurrence matrices, normalised using the Jaccard index, examine the structure of category-level association within each corpus (Fig. 4). The bold lines separating the tangible block (rows and columns 1–6) from the intangible block (7–12) enable visualisation of whether association is greater within each dimension or between them.

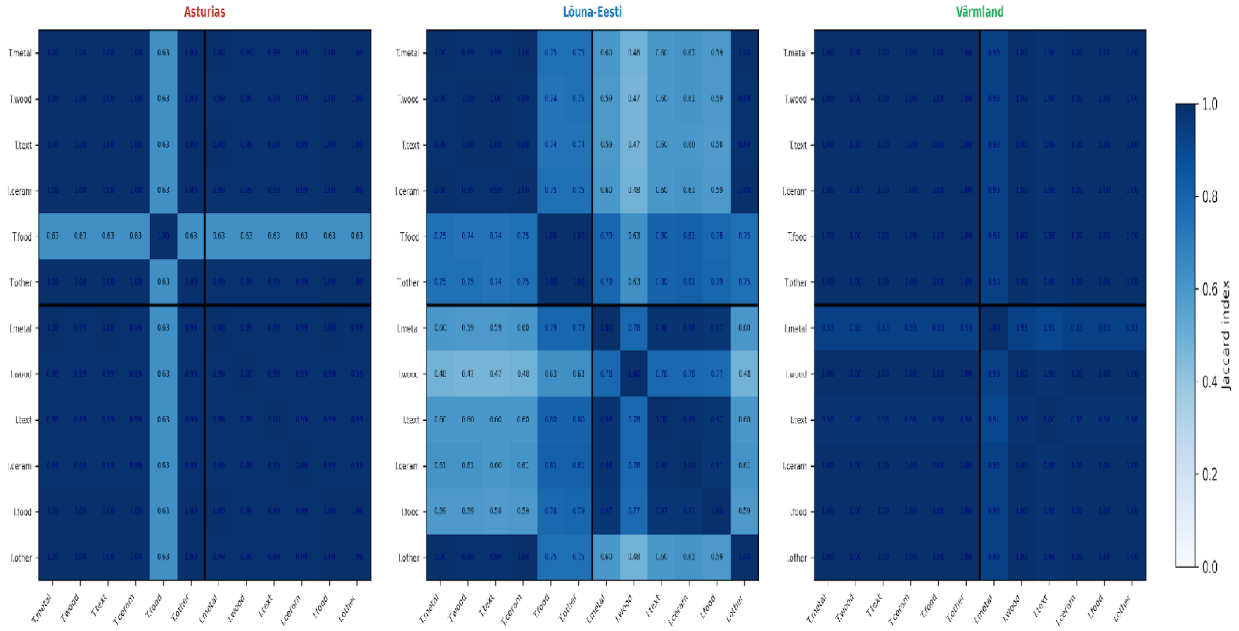


Fig. 4. Co-occurrence matrices (Jaccard index) for Asturias (left), Lõuna-Eesti (centre), and Värmland (right). Bold lines divide the tangible (rows/columns 1–6) and intangible (7–12) blocks. Values close to 1 indicate strong co-occurrence; the Lõuna-Eesti intangible-wood column shows distinctively low cross-block values.

In Asturias, within-block association (tangible–tangible, intangible–intangible) is moderate, while cross-block association (tangible–intangible) is systematically lower. This is consistent with Cramér's V of 0.098: textual documents tend to be thematically specialised in a single dimension. The most notable exception is the tangible-food category, whose high prevalence (99.2%) makes it practically omnipresent, generating higher Jaccard values with all other categories for statistical reasons (a prevalence artefact, not genuine thematic association).

In Lõuna-Eesti, the most pronounced gap corresponds to the intangible-wood category, whose Jaccard index with respect to almost all other categories is the lowest in the matrix. This confirms the singularity of wood-intangible already identified in the frequency analysis: craft processes relating to wood are especially underrepresented in Estonian photographic metadata. The most plausible explanation is that textual metadata associated with photographs of wooden objects in Lõuna-Eesti archives do not systematically include the vocabulary of the carpenter's or carver's craft.

In Värmland, the matrix shows Jaccard values close to 1 across almost all cells, reflecting corpus saturation. The most notable exception is the intangible-metal pair with any other category, showing slightly lower values (around 0.93–0.95), suggesting that even in the best-balanced corpus, metalwork vocabulary (the blacksmith's craft, forging processes) is most likely to appear in specialised records without additional category vocabulary.

The comparative profile across all three regions is summarised in the radar chart (Fig. 5), which consolidates the normalised representation patterns for all twelve categories. The food-tangible peak in Asturias, the intangible-wood deficit in Lõuna-Eesti, and the near-complete saturation of Värmland are simultaneously visible in a single visualisation.

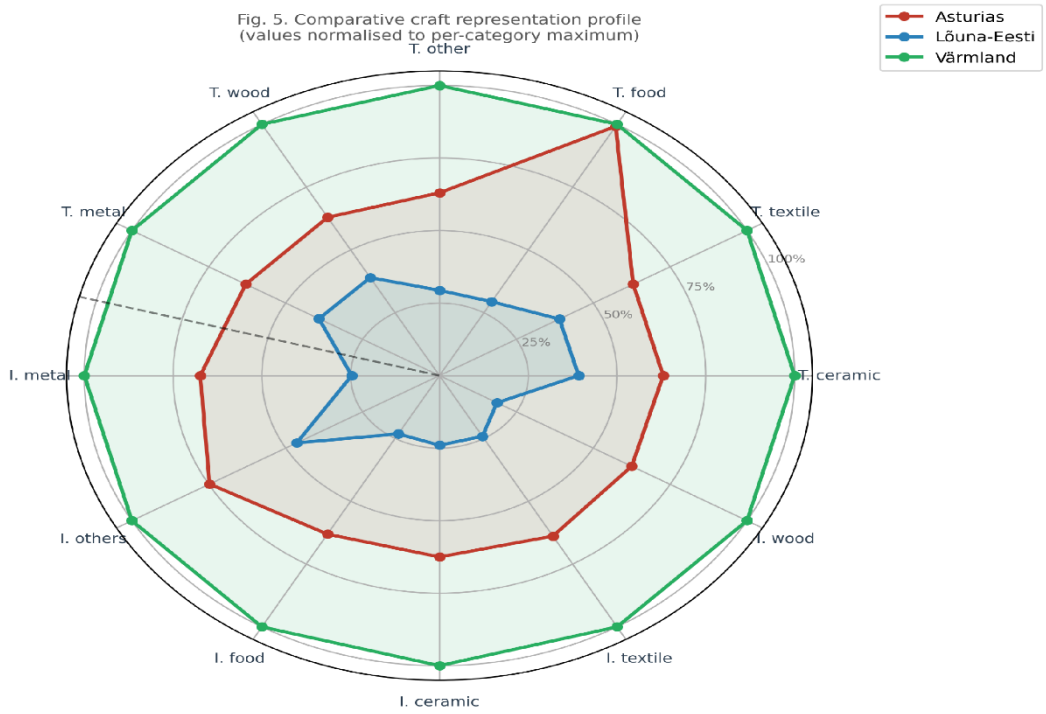


Fig. 5. Comparative craft representation profile (values normalised to per-category maximum) for the three regions. Värmland presents the most complete and balanced profile; Asturias shows the food-tangible asymmetry; Lõuna-Eesti evidences the intangible-wood gap.

Discussion

The results presented allow the working hypothesis to be tested with robust quantitative evidence: the prevailing digitisation philosophy of a region structurally conditions the representation of craft heritage in Europeana. The three corpora analysed—text-dominant (Asturias), image-dominant (Lõuna-Eesti), and balanced (Värmland)—exhibit radically distinct representation profiles despite apparently similar thematic diversity (Shannon entropy close to 1 in all three cases).

The most striking finding is the contrast in Cramér's V: from 0.098 in Asturias to 0.999 in Lõuna-Eesti. At first glance, the high value for Lõuna-Eesti might appear to favour intangible representation. The correct interpretation, however, is precisely the opposite: in the Estonian image corpus, the association between tangible and intangible is high because both dimensions collapse onto the

same binary pattern—a record either has textual metadata or it does not, and when it does, it contains vocabulary from both dimensions. The real gap in Lõuna-Eesti is manifest in the tangible/intangible ratio (1.415) and the intangible-wood anomaly (ratio ≈ 2.09): the photographic system can capture the object but not the craft knowledge. This implies that the ontological design of Europeana—centred on the described cultural heritage object (edm:ProvidedCHO)—limits the degree to which cultural practices can be documented. The intangible dimension only appears indirectly: as a descriptive attribute of the object, not as an autonomous documentary entity.

In Asturias, the situation is different but equally problematic. Textual documents are capable of describing craft processes (intangible) relatively independently of the objects they produce (tangible), which explains the low Cramér's V. But this does not equate to richer intangible representation: it simply indicates that archival texts do not commonly combine both dimensions within the same record. The blacksmith's knowledge and the blacksmith's tool appear in different documents, not integrated.

Värmland presents the most favourable profile: high prevalence across all categories, a tangible/intangible ratio close to 1, and an elevated Cramér's V reflecting the co-presence of both dimensions within the same records. However, this advantage is mediated by corpus saturation: when virtually all records contain terms from all categories, the discriminatory power of lexical analysis is reduced. Värmland's corpus may not reflect genuinely richer documentation of intangible craft heritage, but rather the general richness of its textual descriptions, which tend to incidentally contain broad vocabulary covering multiple dimensions.

These observations connect with the broader debate on biases in digital heritage platforms. The difference between the three corpora does not lie in Värmland documenting intangible heritage better, but in its contributing institutions employing richer textual descriptions that collaterally capture more process vocabulary. To go beyond the preliminary insights of text mining requires more sophisticated computational methods.

The most practically relevant implication for rural cultural tourism is the following. Heritage data platforms such as Europeana are increasingly used as content sources for the development of cultural tourism narratives. Craft practices and craft knowledge are central components of the territorial identity of rural spaces and a significant tourism resource. Yet Europeana data preferentially reflect material culture: the craft object, not the production process; the artefact, not the embodied knowledge. In regions with image-dominant corpora (such as Lõuna-Eesti), this limitation is more pronounced; in regions with balanced corpora (such as Värmland), it is less severe but equally present; in regions with text corpora (such as Asturias), the intangible appears but in fragmented form, without integration with the tangible.

This fragmentation has consequences for the narrative construction of craft heritage in cultural tourism contexts. A tourism platform aggregating Europeana data may find images of Asturian ceramics, inventories of potters, or photographs of Estonian looms, but will rarely find structured data connecting the object with the process, the artefact with the craft knowledge, the piece with the tradition. The tourism experience that can be constructed from these data is necessarily incomplete.

Faced with this limitation, community-based craft heritage documentation initiatives—such as the intangible heritage inventories developed within the framework of the UNESCO Convention[2] or participatory archive projects—offer alternative models that merit greater integration with European-scale heritage data infrastructures. The CULTURALITY project's strategy of developing virtual spaces for visualising craft heritage—specifically through the VERA (Virtual European Rural Artisan) Platform[16]—demonstrates that institutional digitisation archives represent one, but not the only or necessarily optimal, pathway to making craft heritage narratively accessible.

Part of Work Package 2 of CULTURALITY involves collecting field data, which consists of conducting interviews to obtain documentary material and video recordings of the interviews themselves. The ontology of the database design places significant emphasis on the construction of a narrative of cultural practices. This implies that whilst the materials provide information about the objects of production, the information is documented in relation to the process; thus, in this case, the content to be mined using techniques such as those described in this article would not involve a description but rather the collection of materials as a whole.

On the one hand, the metadata from the fieldwork itself is richer and more diverse in nature, as the very description of the cultural practice and the experience of the interviewee already imply the existence of a narrative and a broader cultural context. It is this context that can be analysed using artificial intelligence techniques. On the other hand, the material itself involves multiple dimensions of analysis of the audiovisual content. The videos imply a contextualisation and a narrative implicit in the very narratives of the people who produce the craft products.

One of the findings of the CULTURALITY project is that, in general, the target audience for the promotion of brands based on cultural tourism has a clear preference for the intangible aspects of heritage, identifying key elements such as identity, authenticity and community as the distinguishing features that lend cultural—rather than economic—value to cultural practices. These findings will be discussed in detail in subsequent publications.

Thus, this study has put forward a working hypothesis, the verification of which—through the results obtained via large-scale data mining of institutional heritage databases—establishes the starting point for presenting findings on the potential of cultural heritage linked to the cultural practices of craftsmanship as a catalyst for cultural tourism in remote rural areas.

CONCLUSIONS

This article has presented quantitative evidence supporting an affirmative response to the hypothesis proposed: the predominant digitisation philosophy of a region structurally conditions the representation of intangible craft heritage in Europeana. The three analysed corpora—text-dominant (Asturias), image-dominant (Lõuna-Eesti), and balanced (Värmland)—display radically distinct representation profiles despite apparently similar thematic diversity.

Four main findings emerge. First, Värmland achieves the greatest parity between tangible and intangible dimensions (global ratio 1.015; Cramér's V = 0.900), but at the cost of lexical saturation that reduces the discriminatory power

of the analysis. Second, Lõuna-Eesti exhibits the highest tangible bias (ratio 1.415) and a specific anomaly in the intangible-wood category (ratio ≈ 2.09), consistent with the structural limits of photographic metadata for capturing processes. Third, Asturias shows the greatest independence between dimensions (Cramér's V = 0.098), reflecting the textual fragmentation of craft heritage in its documentary archive. Fourth, in all cases, the presence of intangible craft vocabulary in Europeana is a collateral effect of the general descriptive richness of metadata, not the result of a conscious strategy for intangible heritage documentation.

From the standpoint of limitations, the lexical dictionary method is bounded by the included terms: a low count indicates absence in the captured vocabulary, not cultural absence. Europeana corpora reflects the digitisation priorities of contributing institutions, not a representative sample of European craft heritage. The saturation of the Värmland corpus limits the interpretability of certain results. These considerations do not invalidate the conclusions but must be borne in mind when extrapolating findings.

The implications for digital heritage policy are clear. Expanding Europeana metadata schemas to explicitly incorporate the intangible dimension of craft heritage is a methodological necessity with practical consequences for rural cultural tourism. Establishing workflows between institutional cataloguers and oral and craft heritage archives, developing documentation formats specific to craft processes, and integrating community-led intangible heritage inventory initiatives into European data infrastructures are development directions for which the results presented here provide empirical grounding.

In sum, Europeana data on craft heritage preferentially document the material culture of the analysed territories. Craft knowledge—the know-how of the artisan—remains, to a large extent, invisible heritage within the data infrastructures that purport to represent the cultural richness of rural European regions.

FUNDING AND PROJECT FRAMEWORK

This work was carried out within the framework of the CULTURALITY project (Cultural Tourism and Identity in Rural Regions), funded by the European Union's Horizon Europe programme under Grant Agreement No. 101094016. CULTURALITY is a European research consortium analysing the role of cultural tourism in the revitalisation of rural territories, with particular attention to craft traditions, intangible heritage, and regional identity as resources for sustainable development. The regions of Asturias, Lõuna-Eesti, and Värmland form part of the project's analytical scope, together with other European regions. Further information at: https://culturalityproject.eu/.

The views expressed in this article are the sole responsibility of the author and do not necessarily represent those of the CULTURALITY consortium or the European Commission. The European Commission assumes no responsibility for any use that may be made of the information contained in this work.

NOTES

[1] UNESCO, *Convention for the Safeguarding of the Intangible Cultural Heritage* (Paris: UNESCO, 2003), https://ich.unesco.org/en/convention.
[2] Carlotta Capurro, "Digitality as a Cultural Policy Instrument: Europeana and the Europeanisation of Digital Heritage," in *Critical Heritage Studies and the Futures of Europe* (2023), 223–42, https://www.jstor.org/stable/jj.3385987.23.
[3] Carlotta Capurro and Marta Severo, "Mapping European Digital Heritage Politics: An Empirical Study of Europeana as a Web-Based Network," *Heritage & Society* 19, no. 1 (2026): 136–56, https://doi.org/10.1080/2159032X.2023.2266801.
[4] Bernhard Haslhofer and Antoine Isaac, "data.europeana.eu: The Europeana Linked Open Data Pilot," in *Proceedings of the International Conference on Dublin Core and Metadata Applications* (The Hague: Dublin Core Metadata Initiative, 2011), 94–104.
[5] Murtha Baca, ed., *Introduction to Metadata*, 3rd ed. (Los Angeles: Getty Publications, 2016), 1–24.
[6] Krešimir Starčević, Željko Vojinović, and Jerko Glavaš, "Intangible Cultural Heritage and Their Influence on Financial Results in Rural Tourism," *Economics of Agriculture* 69, no. 2 (2022): 483–96.
[7] CULTURALITY – Cultural Tourism and Identity in Rural Regions, Horizon Europe project no. 101094016, https://culturalityproject.eu/.
[8] The query code is available on request from the author.
[9] The complete multilingual dictionaries are available on request from the author.
[10] Justin Grimmer and Brandon M. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis* 21, no. 3 (2013): 267–97, https://doi.org/10.1093/pan/mps028.
[11] Ibid., 270–272.
[12] Claude E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, no. 3 (1948): 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
[13] Harald Cramér, *Mathematical Methods of Statistics* (Princeton: Princeton University Press, 1946), 282; Karl Pearson, "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling," *Philosophical Magazine* 50, no. 302 (1900): 157–75, https://doi.org/10.1080/14786440009463897.
[14] VERA (Virtual European Rural Artisan) Platform, developed within CULTURALITY Work Package 7, https://culturality.museum/.

Reception date: March 24, 2026
Revision date: September 21, 2026
Acceptance date: October 13, 2026