

---

---

# Using Topic Modelling to Reassess Heritage Values from a People-centred Perspective: Applications from the North of England

Martina Tenzer  & John Schofield 

*The historic environment—comprising a palimpsest of landscapes, buildings and objects—carries meaning and plays a crucial role in giving people a sense of place, identity and belonging. It represents a repository of ever-accumulating collective and individually held values—shared perceptions, experiences, life histories, beliefs and traditions. These social or private values are mostly ascribed by people to familiar places within this environment based on the ontological security which this everyday heritage provides. However, these values are notoriously hard to capture and categorize. This makes it difficult to incorporate them into heritage-management strategies, which typically rely on objective, fact-based datasets. In this paper, we present a new methodology to capture those elusive values, by combining Topic Modelling with the principles of Grounded Theory. Results show that our novel approach is viable and replicable and that these important values can be effectively and meaningfully integrated, thus creating more inclusive approaches to heritage management than exist currently.*

## Introduction

The historic environment is a generic and inclusive term that encapsulates the landscape and all places within it that are considered to be heritage. The historic environment is therefore a palimpsest, constantly evolving through the dual processes of change and creation (Bradley *et al.* 2004). As such, this environment carries meaning and plays a crucial role in developing a sense of place, identity and belonging for its occupants (Avrami *et al.* 2000; Council of Europe 2000; 2005; Gibson & Pendlebury 2016; Ireland & Schofield 2015; Jones 2017; Pearson 1995; Schofield 2014; Stephenson 2008; Tuan 1990; West 2010). Furthermore, it is a repository of collective and individually held values. These values create communities ('heritage communities', after Council of Europe 2005) of shared perceptions, experiences, life histories, beliefs and traditions. Values are therefore routinely ascribed by people to their familiar landscapes, neighbourhoods and places

based on the ontological security of this everyday heritage in daily life and routine (Grenville 2007). People and the landscapes they occupy are thus intimately enmeshed and meaningful places are deeply embedded in people's psyche, as demonstrated in recent research by Gatersleben *et al.* (2020), who used MRI scans to identify the activation of brain areas associated with emotional responses to such meaningful places in ways not found for either meaningful objects or neutral places.

People create a wealth of local knowledge and expertise through their everyday social life. This information is vital for understanding what makes places meaningful and valued. It should also be crucial for the management of those places to acknowledge the varied forms of perception vital for understanding what makes a place important to people. However, these collective and individually held values are diverse and hard to capture, making them difficult to incorporate into heritage management strategies that are historically designed around

objective factual data,<sup>1</sup> typically based on datasets, such as the Historic Environment Records, finds databases and historic maps. The shortcomings we describe in the following section provide the justification for our approach.

### Reviewing heritage values

A paradigm shift since the 1960s—the ‘cultural turn’—had far-reaching implications for understanding the importance of values in the heritage sector and their creation through the meaning-making of communities and individuals (Tuan 1980). The nature of values was no longer seen as intrinsic in the fabric of the material world: buildings, landscapes, sites, or things (Pearson 1995, 308), but as a construct and the result of negotiation within democratic societies, communities and individuals, ‘ascribed to heritage by society at large’ (Avrami 2009, 179; see also Jones 2017, 21). Views on the role of heritage professionals and non-experts, and the concept of values in general, changed significantly—the heritage expert no longer had the monopoly of authority in defining heritage significance (Bonnell & Hunt 1999; Cosgrove 2004, 57; Gibson & Pendlebury 2016, 1–2).

However, while there is a consensus on the importance of recognizing people’s perceptions and local knowledge for a more sustainable way to manage and think about heritage, practical solutions for the integration of people’s voices in the decision-making process have been slow to develop and become integrated into heritage management. Heritage values are a particularly good example of these shortcomings. Ascribed values, when identified and determined by heritage specialists, can help to assess and determine the measures of management, inform statutory heritage protection decision making and provide support appropriate for specific parts of the historic environment. However, value categorization has been in dispute ever since the recognition of the benefits of positioning locally held viewpoints on value and significance alongside those of specialists, not least in terms of the tension created between the benefits of such an approach (e.g. its inclusivity) but also the significant challenges it entails (e.g. its practicality).

The suggestion to incorporate locally held values into heritage management systems was first promoted for heritage managers in the Amsterdam Charter (ICOMOS 1975), reinforcing the existing notion of public consultation and inclusion in heritage decision-making. The Burra Charter (ICOMOS 1979) did not mention this point before the 1999

version (now ICOMOS 2013). However, Smith’s (2006) analysis of international conventions and charters, such as the Athens and Venice Charters, but also the Burra Charter, showed continued adherence to the principles of the Authorized Heritage Discourse (AHD), defined by Smith (2006) as a dominant concept of heritage management that is ‘reducing the authority of non-experts’ and ‘appeals to the moral authority of expertise’. While the Burra Charter contains some progressive ideas about inclusion, its authoritative language still nonetheless strengthens the position of experts, albeit indirectly.

Public perception and participation was then integrated in the European Landscape or Florence Convention (Council of Europe 2000), the Faro Convention on the Value of Cultural Heritage for Society (Council of Europe 2005), and in Historic England’s *Conservation Principles* document (English Heritage 2008). All these policy or guidance documents called into question the comprehensiveness and inclusivity of using only those objective, scientific value categories which provide the current framework (Supplementary Material 1). They challenged heritage managers to consider how much local knowledge and expertise is included in heritage evaluations and how fluid they are, to take account of the changing social fabric of an area over time, or people’s changing opinions. Alongside Historic England, another national agency—Natural England—considers the natural and cultural value of landscapes based on the perception of people, aligning closely with the *European Landscape Convention*, referred to earlier. However, while current studies use ‘proxies’ or ‘secondary measures’ to identify people’s perception (for example, tourist numbers, numbers of footpaths and car parks as ‘opportunities for health walks’, or ‘measures of accessibility’: Natural England 2015, 10), no time-efficient and effective method for the collection of qualitative data has been developed, mirroring the situation for cultural heritage.<sup>2</sup>

With these developments in mind, we ask: is the current value system on which experts assess heritage fit for purpose? And is the valuation framework for conservation, arguably exemplified by that currently offered by Historic England (English Heritage 2008), sufficiently forward-thinking and up to the task in a fast-changing world? This question of changing perceptions aligns with the growing realization that the term ‘heritage’ itself is fluid and dynamic: a social process of creating meaning and memory, communities and identities. Value categories, it was argued, should also be sufficiently flexible to adapt to societal change and environmental

pressure (Byrne 2008; Dalglish & Leslie 2016; Harrison 2010; Jones 2017; Jones & Leech 2015; Smith 2006).

Against this background, we explore and critique the current value categories described in *Conservation Principles*. As a significant departure from current applications, we propose an innovative approach which involves identifying those hard-to-obtain but important values held by local people and their communities. This approach will address the lack of inclusion of people's perceptions and the dynamic quality of social values<sup>3</sup> as an ongoing challenge for heritage and landscape management.<sup>4</sup> We use Topic Modelling to identify latent or emerging value themes from a public survey, providing the individual stories (the intangible element) of 'everyday heritage'. This offers an open-minded approach to qualitative data. We propose Natural Language Processing (NLP) and Machine Learning (ML) with a focus on Topic Modelling as a method, following basic principles of Grounded Theory for improving the system of valuation in heritage management (Charmaz 2006; Creswell 2017, 276). Following Grounded Theory means a first investigation of the data free of the researcher bias<sup>5</sup>—with no predefined codes or assumptions. The themes latent in the data are explored as they emerge and in a later phase of the study are structured into a framework of topics. The results are correlated with the current framework of values as applied by Historic England, suggesting how new categories could help to address the changing expectations, needs and demands of the public, partly developed as a reaction to a changing world after the COVID-19 pandemic and the increasing pressures on natural and cultural landscapes that have emerged as a result (Ginzarly & Srour 2022; Historic England 2022; UNESCO 2021).

## Data sources and methodology

### *Overview of the method*

To explore ways to capture social values in a form that can be used within the planning process and in heritage management decision-making requires a method that is efficient and scalable. Here we use Topic Modelling to present place attachment in a format that can be used for categorizing individually held values in cultural landscapes. We will first describe the method before going on to deliver a proof of concept. While this novel methodology is based around two areas in the UK, it is transferable to any context where local authorities have capacity to routinely use survey data or online portals to

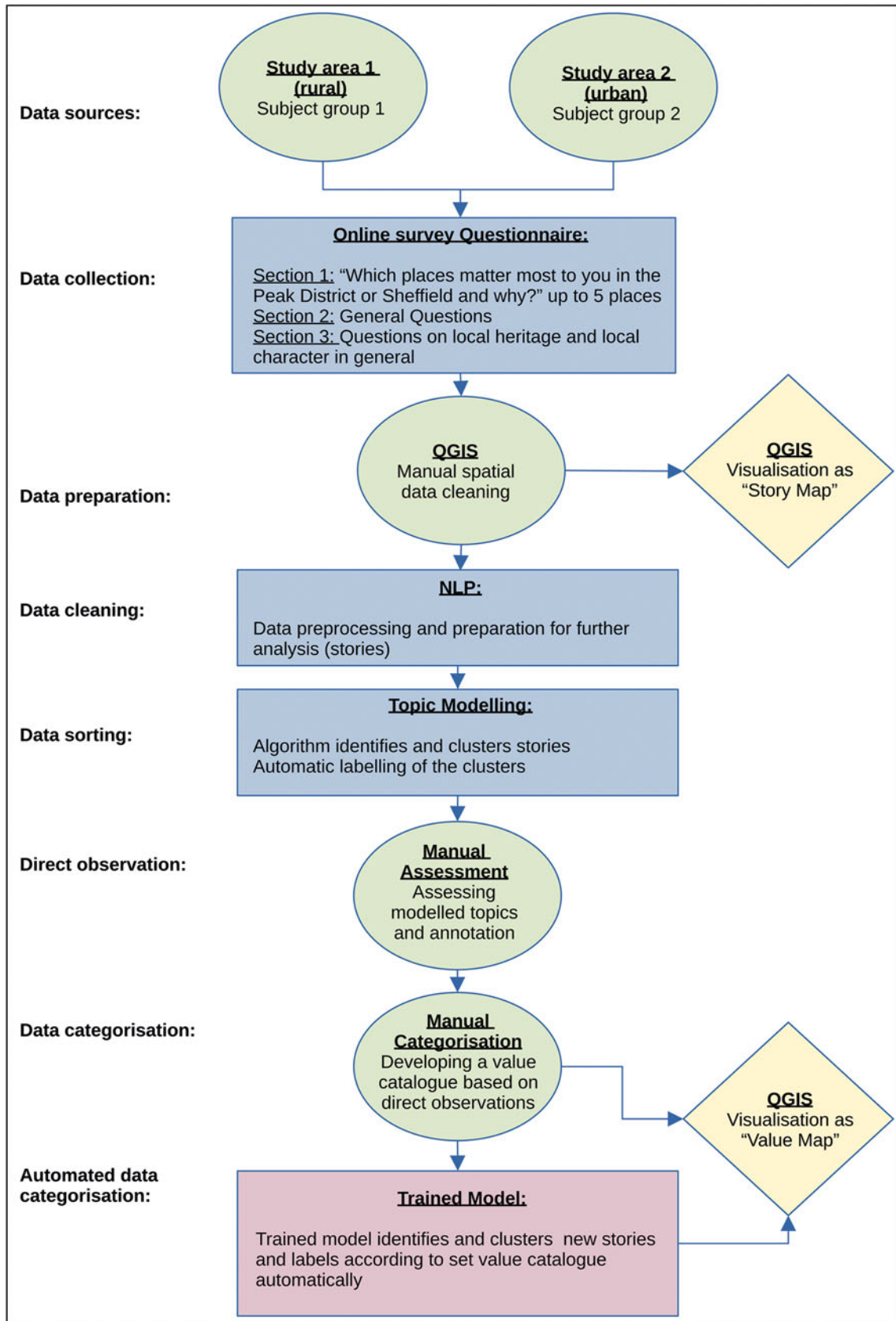
record people's interactions with or feelings about a place.

The flowchart in Figure 1 details the process applied in this research. Residents of two study areas were asked to provide five favourite places that matter most to them and 'stories' about their personal connection to these places. The stories were then fed into a Topic Modelling algorithm to get a first, bias-reduced insight into the latent themes within the data and preliminary topic labels summarizing the topic clusters. Direct observation then allowed evaluation of the modelling result and an assessment of the usefulness of this approach. The annotated data were subsequently categorized based on the value catalogue provided by Historic England's *Conservation Principles*. Visualization in GIS can be created at two stages of this process: in the form of a 'story map' as the basis for communication between local authorities and residents; and in the form of a 'value map' following the process of categorization, with the potential to provide background information for planning purposes.

### *Study areas*

Two study areas were identified for this project: the City of Sheffield and the Peak District National Park, both located in the north of England (Fig. 2). These locations were chosen to offer insight into the enmeshed relationship between people and places amongst a combination of urban and rural communities. Additional methods include in-depth interviews and a social media analysis (Tenzer 2022; Tenzer & Schofield n.d.). While the district of Sheffield overlaps with the National Park, the two landscapes are in many respects different and distinctive. Residents of both areas are closely connected to the landscapes of both study areas through leisure activities, work and shopping, cultural activities and educational organizations. This integration was considered an advantage when selecting the two study areas.

The Peak District National Park—the first designated National Park—covers an area of 1438 sq. km. It is a place of residence for approximately 38,000 people at the time of writing (PDNPA n.d.). The Park includes various archaeological and historical sites from the Palaeolithic to recent periods, including prehistoric burial mounds, henges and stone circles, medieval field systems and settlements and post-medieval industrial sites, including stone quarries. At the point of writing, the Peak District National Park has 2900 listed buildings, 109 conservation areas and 450 scheduled monuments, all of which involve various degrees of statutory protection (PDNPA n.d.).





Sheffield lies to the west of the Peak District National Park. It is dominated by seven hills and two universities, covering an area of 367 sq. km with approximately 556,500 residents. It is best known for its industrial heritage. Today, both visitors and residents favour the town for its multicultural character, the wide offer of cultural events and the quality of an outdoor city with its vicinity to the National Park. There are currently 1200 listed buildings, 38 conservation areas and 43 scheduled monuments in the city.<sup>6</sup>

#### Survey method

Residents of the Peak District National Park (PDNP) and the City of Sheffield were first invited to participate in an online survey. The online questionnaire was published through the channels of both local authorities, the Peak District National Park Authority (PDNPA) and Sheffield City Council, comprising their websites, social media channels (i.e. Facebook, Twitter, Instagram, LinkedIn) and a specific mailing list of over 4500 recipients in Sheffield. Participants therefore included residents of the two study areas exclusively,<sup>7</sup> who self-identified as residents of either study area. In total 476 responses were received. Forty-eight participants identified themselves as residents of the PDNP and 386 as residents of Sheffield; 42 participants did not answer this question.

The survey questionnaire consisted of three sections: general information (such as place of residence, age, ethnicity, and level of education); the question which places matter most to the respondent, and why; and questions on their perception of the local heritage and landscapes (see Supplementary Material 3). Questions in sections 1 and 3 were multiple-choice. To allow the most flexibility and subjectivity in the second section, respondents were given the opportunity to enter a location on a map and their 'story'—the personal reason for a connection to the place—as a free-text entry of up to 300 words. To identify more places of individual importance, the respondents were asked to enter up to five places in one questionnaire. Not all respondents provided five places as requested. The average was just over one place/response per respondent, resulting in 547 places, of which 509 provided a story, experience or perception of heritage, landscapes or history relating to one of the two study areas.

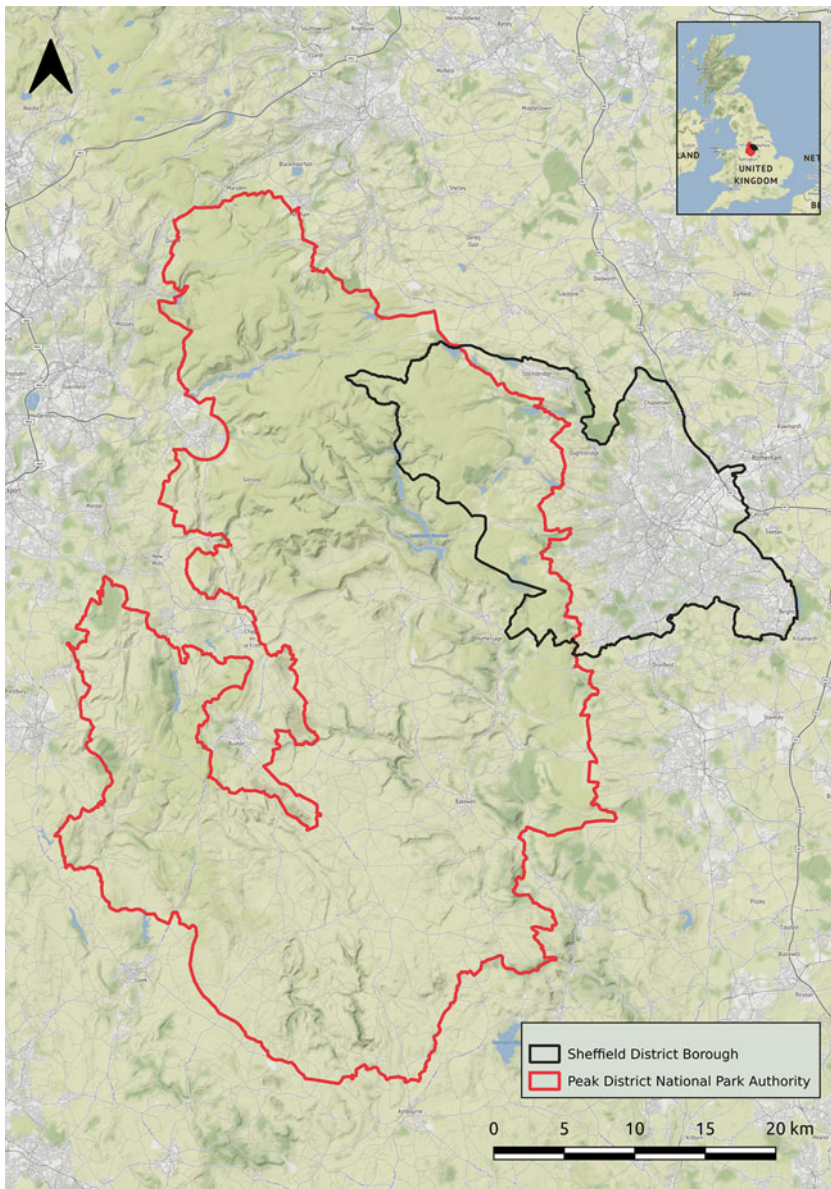
The survey responses were downloaded and imported into a Geographic Information System (GIS).<sup>8</sup> The dataset was then cleaned. This included: eliminating places not located within the study areas and repositioning locations set in the wrong place (determined by written input in the location field). The participants provided 174 favourite places in the PDNP and 298 in Sheffield. The intersection of the study areas included 60 locations. Eleven locations close to but outside the study area limits to the east of the City of Sheffield were included as the locations were close to the boundary, i.e. Buxton (see Figure 3).

The online questionnaire for this research followed the principle of offering a low-cost/cost-free, practical solution for survey design by using the Qualtrics software<sup>9</sup> and an embedded Google Maps map. This questionnaire allowed the participants to use a familiar map interface to locate and pin a location, which automatically provided geospatial coordinates for the GIS map analysis to create georeferenced stories. Where participants were not able to locate the place on the map, they could also enter locations as free text. In this case a Named Entity Recognition process was used as detailed in Tenzer (2022). The data were then ready to be preprocessed for analysis with NLP and TM.

#### Natural Language Processing and Topic Modelling

NLP and TM are emerging methods for data analysis that are particularly relevant to qualitative research. Heritage and landscape studies have yet fully to appreciate and integrate the opportunities offered by these tools (but see Bordoni *et al.* 2016; Condorelli *et al.* 2020; Fiorucci *et al.* 2020; Matrone *et al.* 2020; Verschoof-van der Vaart *et al.* 2020). The Council of Europe actively encourages the use of Artificial Intelligence (AI) in all sectors, including heritage (Traviglia 2022), as a result of increasing interest in Text Mining and Machine Learning/Topic Modelling for computational language analysis (Goerz & Scholz 2010; Sassolini & Cinini 2010; Sporleder 2010). Our research applies AI in line with these projects, using NLP algorithms, to model categories of values based on perceptions, experiences, concerns and visions of residents of the two study areas, as described in the 'stories' provided by the online surveys.

**Figure 1** (opposite). Detailed methodology developed in this research. The aim is to create the topics based on the 'stories' of survey participants and correlate these with value categories as set out in *Historic England's Conservation Principles* (English Heritage 2008), extending the current value catalogue. (Green: manual process; blue: computed process; red: outlook; yellow: visualization.)



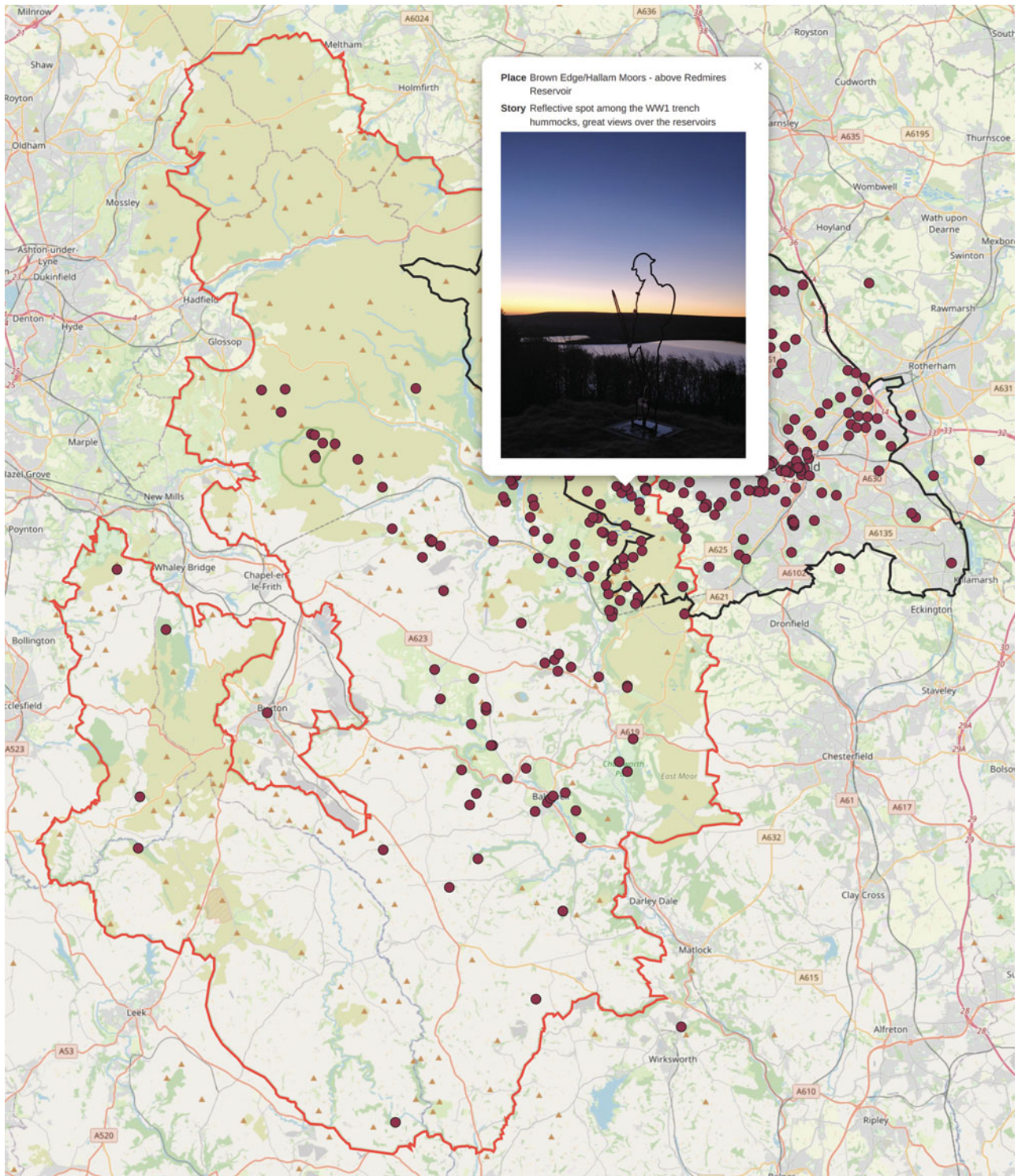
**Figure 2.** Study areas: Peak District National Park and Sheffield city. Residents of these areas were invited to participate in a survey and provide up to five favourite places within the study areas and their ‘stories’ of personal connection. (Map created in QGIS; data contain OS data © Crown copyright and database right 2022. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under OdbL.)

TM is a time-efficient method to analyse qualitative data and has been tested for its capabilities elsewhere (e.g. Abram *et al.* 2020; Cai *et al.* 2021; Franzosi *et al.* 2022; Ginzarly & Srour 2022). More importantly, this method of text analysis allows themes and topics emerging from or ‘latent’ within the data to be captured without preconceived codes. This approach aligns with the underpinning elements of Grounded Theory (e.g. Charmaz 2006; Odacioglu & Zhang 2022). Grounded Theory is usually applied ‘when little is known about the phenomenon’ (Chun Tie *et al.* 2019), or in the case of this research, where traditional thinking dominated by experts (often referred to in the context of an Authorized Heritage Discourse: see Schofield 2014; Smith 2006; Waterton *et al.* 2006)

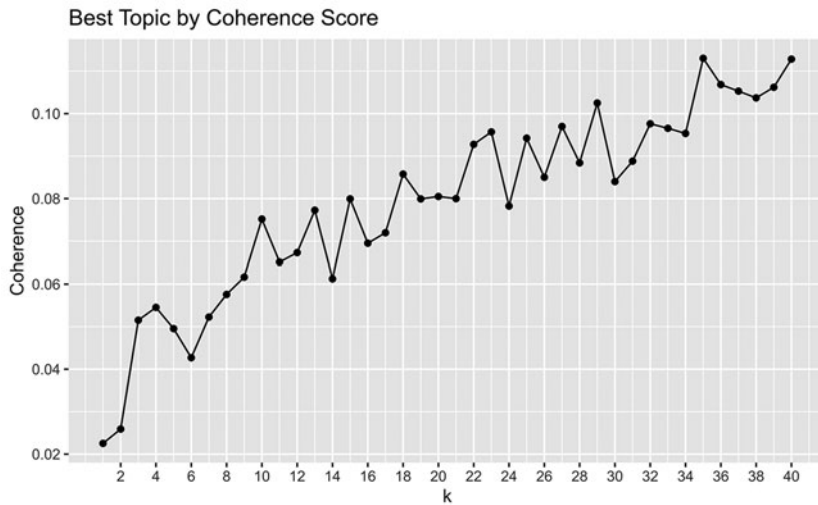
is preferably avoided, exploring the connection between people and places, perception and place-making, in a narrow and individual sense, without the bias of preconceived expert knowledge.

The following steps are part of the analysis shown in the flowchart in Figure 1. To undertake this analysis, the survey data were downloaded. Coordinates, place names, stories and photographs were extracted from the survey dataset—separating these from the demographic data—and then cleaned and preprocessed with *textmineR*, eliminating places not located within the study area and relocating pins set in error. The resulting data set was then fed into the NLP toolchain using R and Python routines in RStudio.





**Figure 3.** Interactive 'story map', showing the favourite places provided by residents of the two study areas who participated in an online survey. The participants were asked to provide up to five places that matter most to them and give an up to 300-word description of the reasons for the connection to these places. (Map created in QGIS/Leaflet, © OpenStreetMap contributors; data contain OS data © Crown copyright and database right 2022.)



**Figure 4.** The Topic coherence gives the degree to which documents in one cluster are closely related. This depends on the size of the cluster and the number of topics chosen. For example, if we subdivide the documents into four different topics, the documents have less coherence—similar semantics and supporting the statements of each other—than at the point of 35 clusters, where the number of documents in one cluster is smaller but the relationship between the documents higher. At 35 topics the coherence is highest, flattening out with more subdivisions. Therefore, 35 topics will provide the best first insight into the latent topics within the documents (tqx94 2022).

The five steps of the Topic Modelling process are as follows:

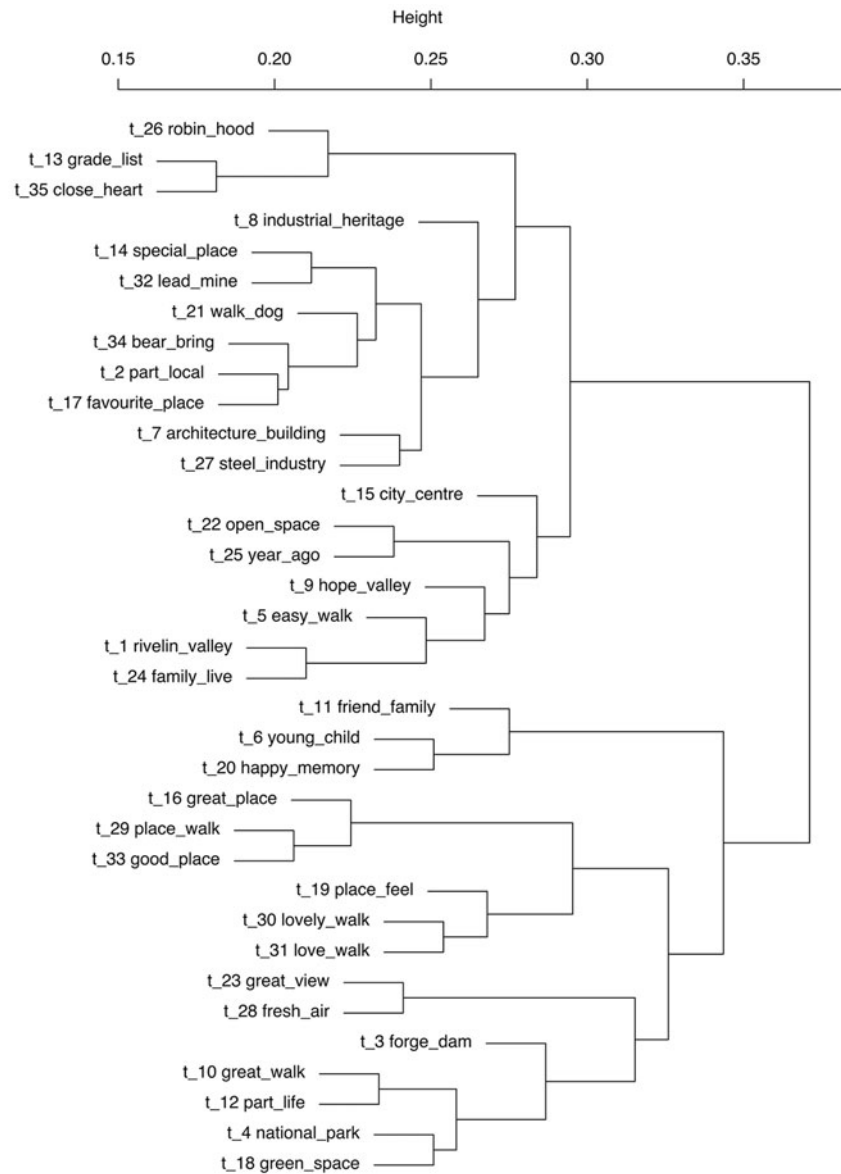
1. Data wrangling: cleaning text, lemmatizing, stemming, removing stop words, creating a DocumentTermMatrix.
2. Choosing a model algorithm: the R package *textmineR* provides an innovative method of topic modelling and labelling (currently under development at George Mason University, USA, by Tommy Jones: see Jones 2021; Jones *et al.* 2021). This unsupervised machine-learning model uses unlabelled data and implements the Latent Dirichlet Allocation algorithm and Gibbs Sampling (Blei *et al.* 2003). It creates word clusters and identifies topic labels based on the probabilistic distribution of words over topics and topics over documents. This means that each topic is a combination of keywords which have a specific importance (weight) contributing to the weightage of each topic.
3. Running the model and adjusting parameters: most of the parameters were left to default settings, while the number of topics and iterations was optimized. The model iterates *n*-times over the data and provides information on the topic coherence level at *k* topics (Fig. 4). This means that the model in our case attempts 4000 times to make more sense of the relationship between the documents and to create more meaningful clusters where documents relate more closely to each other.
4. Choosing the optimal number of topics: coherence (Fig. 4) measures the degree to which the documents (stories) in a topic show high semantic similarities and support each other in their

statements. This gives us the model we want to choose for further analysis. In our study, 40 models were created, and the best coherence is provided at 35 topics (tqx94 2022). After 35 topics, the coherence score flattens out, meaning that a greater subdivision of the data would not provide any more coherent and meaningful topic groups. The coherence is not very high in our case and more data would help to fine-tune the model. Nonetheless, this gives a good starting point for the manual analysis.

5. Topic labelling: labels are based on the most frequently appearing key terms in the documents/stories of each cluster (bi-grams: two closely associated words in a text) (Fig. 5) (tqx94 2022). Labels created in this process are not meaningful titles, but give a good indication of the dominant theme in each cluster.

Using this unsupervised learning method prevents the introduction of researcher bias in the next step of data analysis, which is comparable to the coding phase of text in other approaches, such as NVivo, as the algorithm has completed the clustering independently and created a pre-labelled dataset. This approach follows to the tenets of Grounded Theory (Charmaz 2006). Our attempt to capture the themes latent in or emerging from the data themselves opposes the commonly used and widely criticized approach of applying codes or categories and trying to fit the data into these codes, introducing bias and issues with reliability and validity (Banks *et al.* 2018; Welsh 2002). The approach of creating two bi-gram label variants per topic has proved to help identify the themes more accurately (Label 1 and Label 2)





**Figure 5.** The Cluster Dendrogram shows the 35 topic clusters with their respective labels chosen by the modelling process. The levels (Height) show the similarities of topics. Topics in the lower level of the diagram on branches close together show topic clusters that have themes closely related to each other. For example, topic 10: ‘great\_walk’ and topic 12: ‘part\_life’ consist of documents or ‘stories’ with content that support each other and can be summarized under the respective label. Similarly, topic 4: ‘national\_park’ and topic 18: ‘green\_space’ consist of documents with a similar theme. At a higher level, all four clusters are related to each other—to a lesser degree than at a lower level but more closely related than, for example, topic 3: ‘forge\_dam’, again, a level higher up.

(see Table 1). Following the Topic Modelling process, the dataset was exported and the labelling by the algorithm was manually assessed in a direct approach to observe the coherence of the topics (Lau *et al.* 2014). In the same step, the most dominant topics were identified manually and compared to the suggestions made by the algorithm.

A manual evaluation has shown that some of the topics are related to heritage, history and archaeology (Robin\_Hood, industrial\_heritage, steel\_industry, list\_building), while others are related to private life and community (friend\_family, happy\_memory) or aesthetic values (great\_view, natural\_beauty). However, topic labelling still needs human input and dataset structuring for meaningful labelling

and categorization of the data. The application of the algorithm does not fully replace manual assessment and fine-tuning of coding (Cai *et al.* 2021; Chang *et al.* 2009; Leeson *et al.* 2019). In our data model, the manual categorization was subsequently carried out to understand on what basis the documents were clustered (top terms) and if the clustering proved meaningful.

#### *Value categories and perception correlation*

In the third step, the data were manually coded and the quality of the topic clusters and labels was assessed. To correlate these emerging values based on the experiences and perceptions of people with the value categories set by HE (Historic England

**Table 1.** Following the Topic Modelling process, the optimal number of topics (35) was chosen to create labels for the topic clusters. The labels are based on the most frequent key terms in the documents/stories of each topic cluster. Labels vary in their quality, which makes a manual evaluation necessary.

Topic	Label 1	Label 2
1	rivelin_valley	bear_bring
2	part_local	start_point
3	forge_dam	play_area
4	national_park	car_park
5	easy_walk	endcliffe_park
6	young_child	child_grandchild
7	architecture_building	back_yard
8	industrial_heritage	industrial_history
9	hope_valley	walk_edge
10	great_walk	natural_beauty
11	friend_family	meet_friend
12	part_life	amaze_view
13	grade_list	list_build
14	special_place	post_office
15	city_centre	close_city
16	great_place	place_visit
17	favourite_place	enjoy_walk
18	green_space	rich_history
19	place_feel	walk_home
20	happy_memory	lot_memory
21	walk_dog	cricket_pitch
22	open_space	botanical_garden
23	great_view	easily_accessible
24	family_live	family_tree
25	year_ago	live_year
26	robin_hood	beautiful_build
27	steel_industry	portland_work
28	fresh_air	good_view
29	place_walk	good_place
30	lovely_walk	place_walk
31	love_walk	walk_area
32	lead_mine	geological_historical
33	good_place	huge_amount
34	bear_bring	bakewell_pudding
35	close_heart	area_close

since 2015, formerly English Heritage) (English Heritage 2008), each document was allocated to one of the following value categories (subcategories): evidential value, historical value (associative, illustrative), aesthetic value (design), communal value

(social (renamed to 'private'<sup>10</sup>), commemorative, spiritual). The subcategories of 'Communal' values were developed to be more nuanced regarding the variety of aspects of stories, accommodating elements that did not fit into the HE categories. Also, a new category capturing the concept of nature in people's perception of landscape values was integrated to address the increasing awareness of valuing the environment in view of climate change and biodiversity loss (Table 2). This was also necessary to overcome the diametrical-opposite division developed in value systems over the past decades, dividing cultural/natural landscapes, tangible/intangible elements, and learning/mental health approaches to heritage—a subdivision rarely evident in the daily experience of the environment.

## Results

In this section we present the results of the modelling process, direct observations and manual annotation for correlation with the values defined in Historic England's *Conservation Principles*. We do this by first introducing the results of the TM. We then describe the manual observation of usefulness of the modelled topics and labels. Finally, we summarize the results of the correlation and describe the development of a more nuanced division of communal values as identified in the survey data.

Results show that people's perceptions correlate with some of the value categories of HE, showing that the expert definition of heritage values is capable of capturing parts of the individually held values. Our bottom-up approach can be aligned with the expert-led approach to find a common ground for heritage value categorization.

### Modelling categories

As suggested in Figure 4 and illustrated in Figures 5 and 6, the modelling computed 35 clusters as an optimal topic number. The sizes of the clusters varied from five documents/stories (Topic 35) to 33 (Topic 6).

The TM approach provided mixed results. Some documents did not provide a story and were not included in the clustering ( $n = 37$ ). Other documents provided stories that were too short (just one or a few words), which did not allow the algorithm to cluster in a meaningful way. These documents were labelled 0 ( $n = 20$ ). However, during the manual process, these documents could still be allocated to one of the categories. For example, 'magical place' was allocated to the subcategory 'Spiritual' value;<sup>11</sup> the location 'Surprise View' with the description: 'In a

**Table 2.** Historic England's value categories as set out in the *Conservation Principles* (English Heritage 2008). Additional nuance to these categories is proposed based on the underlying themes identified in survey data. The additional values (shaded cells) address societal trends and the changing and dynamic demands and needs of residents in the Peak District National Park and the city of Sheffield following the COVID-19 pandemic.

Categories Historic England	Subcategories	Description
Evidential Value		Evidential value derives from the potential of a place to yield evidence about past human activity.
Historical Value	Illustrative, Associative	Historical value derives from the ways in which past people, events and aspects of life can be connected through a place to the present.
Aesthetic Value	Design	Aesthetic value derives from the ways in which people draw sensory and intellectual stimulation from a place.
Communal Value	Private	Private (compared to public) heritage values are commonly overlooked as 'family history' with no place in the generalization of heritage for the common good.
	Commemorative, Spiritual	Communal value derives from the meanings of a place for the people who relate to it, or for whom it figures in their collective experience or memory.
	Arts & Culture	Aspects of landscapes and environments that afford a communal experience of culture, entertainment, or learning.
Green space Value	Health	Nature values derive from the quality of green spaces, biodiversity, wildlife. Health value derives from the qualities and opportunities of a place or landscape that provides space for outdoor activities, in particular developed during the COVID-19 pandemic.

word, "breathtaking" containing the keyword 'view' in the location name was associated with 'Aesthetic' value; the location 'Kinder Ravine' provided the description 'Scrambling', which was annotated with the code 'outdoor activity' and associated with the type 'Green Space' and the subcategory 'Health'. In total, 510 documents provided sufficient textual data for the Topic Modelling algorithm to process.

The top-terms list created during the TM process clustered the most frequent words of each document, which provides the basis for the clustering decision and the labelling of each topic cluster. This list was used to decide on the most important keywords of each story or document. The resulting contracted list was subsequently condensed into a single one-word code. This code was used to allocate the document to one of the four value categories included and defined in Historic England's *Conservation Principles* (English Heritage 2008) (historical, evidential, aesthetic, communal<sup>12</sup>) (Table 2).

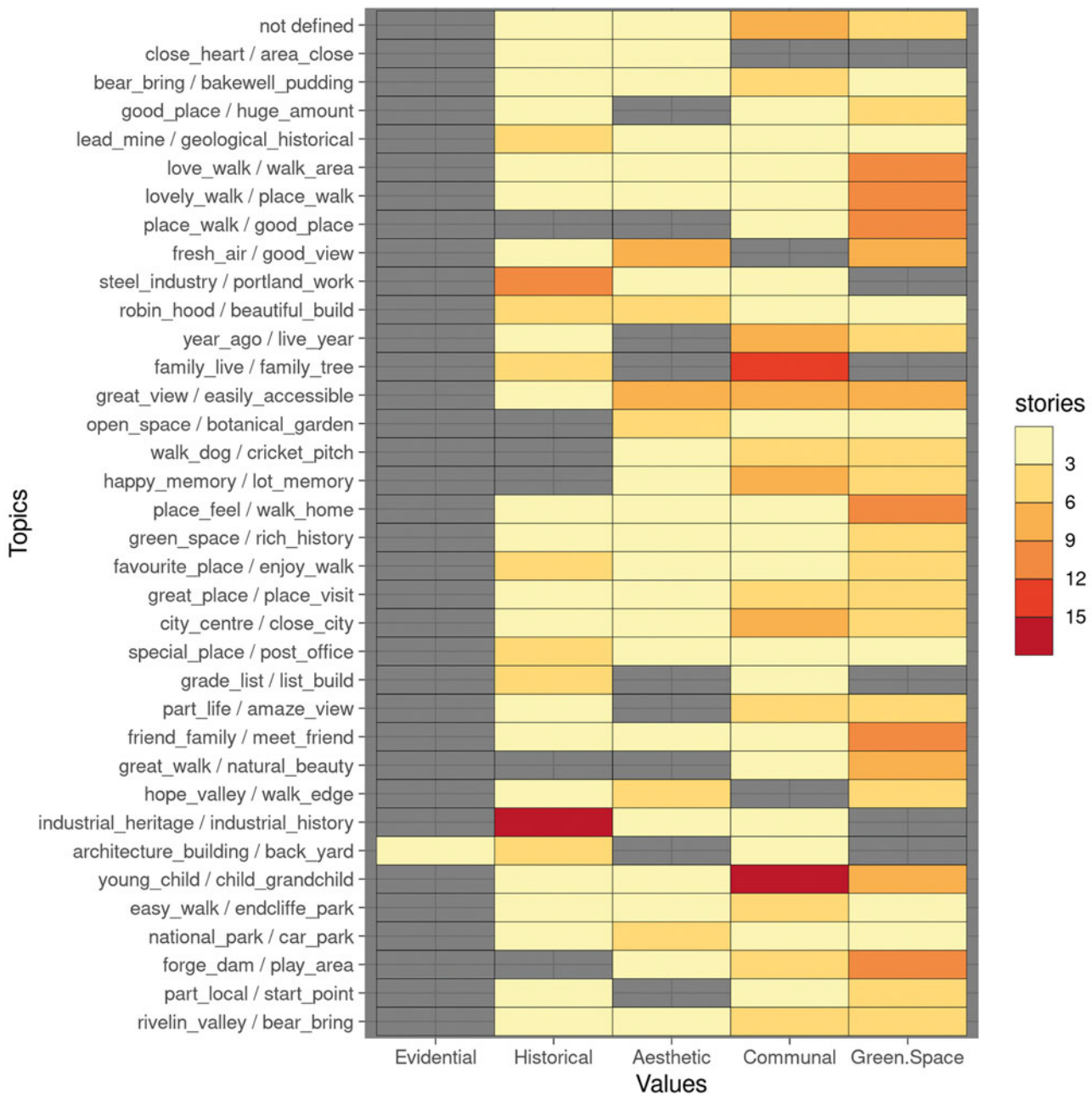
#### Value category development

Automated Topic Modelling provides valuable first insight into the latent themes in the survey data. The stories showed clear trends and commonalities (Table 1). However, it was necessary to assess and refine the categories manually. It became clear that some of the documents could not be fitted into this predefined set of values defined in *Conservation Principles*; for example, topics labelled with 'calm\_place', 'great\_walk', or 'nature\_reserve'. A heatmap

shows that survey respondents based their place attachment predominantly on communal, green space and health aspects (Figs 7 & 8) with green space and health dominating the relationship between people and places. A more nuanced approach to defining heritage values might extend the existing framework by further subdividing the category of communal values (private, spiritual, commemorative) to include 'Arts & Culture' (Fig. 7; Table 2), to capture the intangible qualities of the cultural aspect of areas in Sheffield, in particular. Furthermore, a new (sub)category of 'Green Space' would help to overcome the artificial divide between cultural and natural heritage which people rarely recognize (Byrne & Ween 2015; Harrison 2020; Latour 1993). This category would capture the values people place on wildlife and ecology, for example, while addressing the increasing awareness of environmental pressures on natural resources. Another subcategory of 'Green Space' value could be defined as 'Health' value (both physical and mental), which encompasses, for example, people's COVID-19 experiences and the different approaches to landscapes that have emerged as a result of the lockdowns that were imposed at this time alongside relaxation and the calming qualities of woods in Sheffield's parks that relate to public benefit.

We can illustrate this argument with some examples. First, the document cluster of Topic 19 consisted of 18 documents. Label 1 was given as 'place\_feel' and label 2 as 'walk\_home'. A list of the most

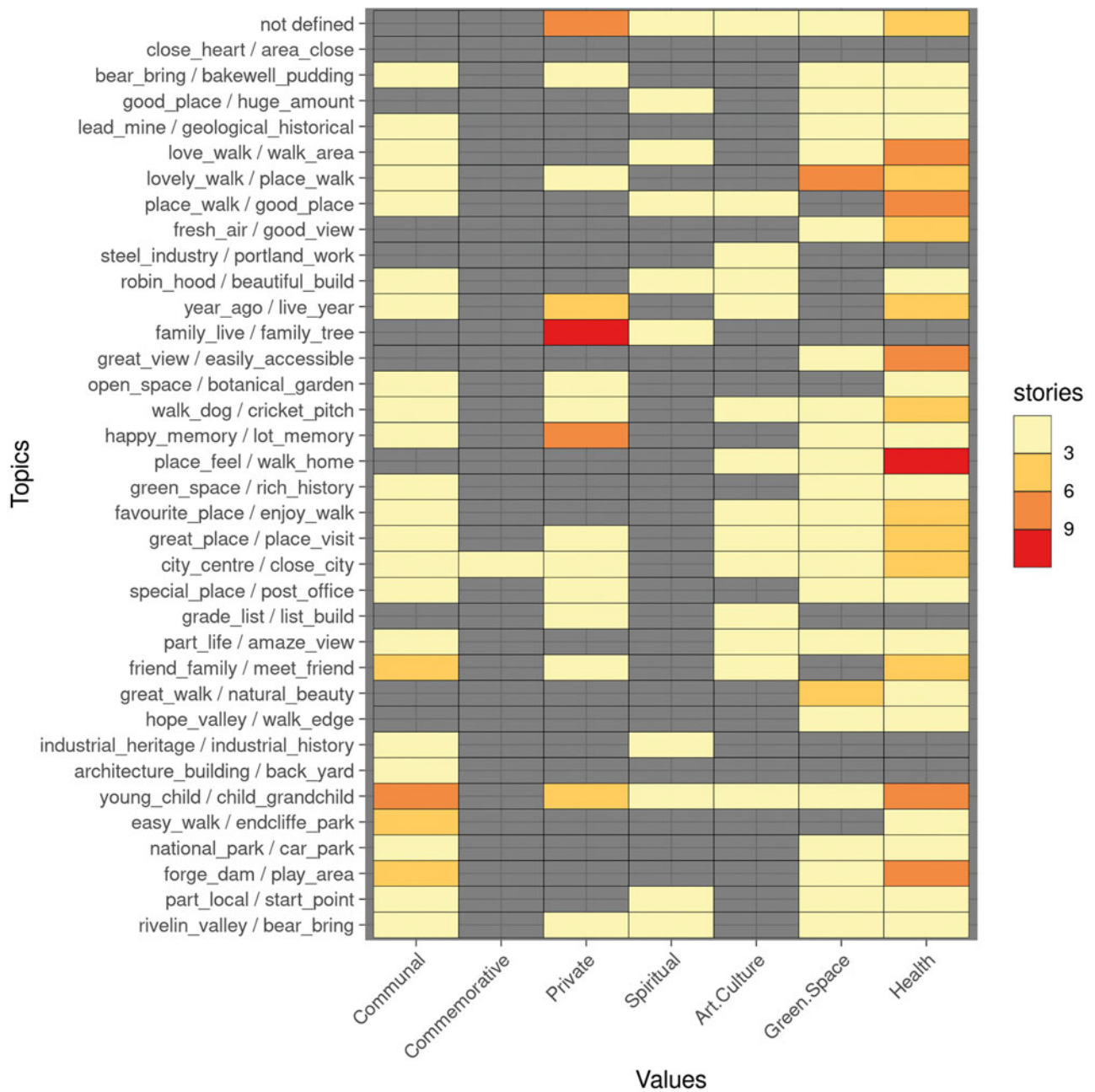




**Figure 6.** Result of manual evaluation of the topic modelling process. Document count over topics, showing the number of documents (stories) allocated to the corresponding value categories as defined by Historic England. High counts (red) show the dominant value category in a topic cluster. Rows with yellow cells (low counts) show incoherent topics with high variation in values. The proposed category for 'Green Space' value, including 'Health' value, dominates in the assessment.

dominant words in the stories showed that 'lock-down' was mentioned in five documents. Ten documents were labelled with the broad subcategory 'Health'. These documents mentioned 'access' to 'nature' or 'green spaces', 'peace', 'walking', 'calming' and 'relax'. Three documents were allocated to 'Aesthetic' and 'Green Space'. The documents

which were not allocated to the aforementioned categories were categorized as 'Aesthetic (Design)' (n = 3), 'Historic (Associative)' (n = 2), 'Natural Resort' (n = 2), and 'Arts & Culture' (n = 1). Topic cluster 24 was labelled by the algorithm as 'family\_live' and 'family\_tree'. Of the 18 documents in this cluster, 12 were manually categorized as relating to



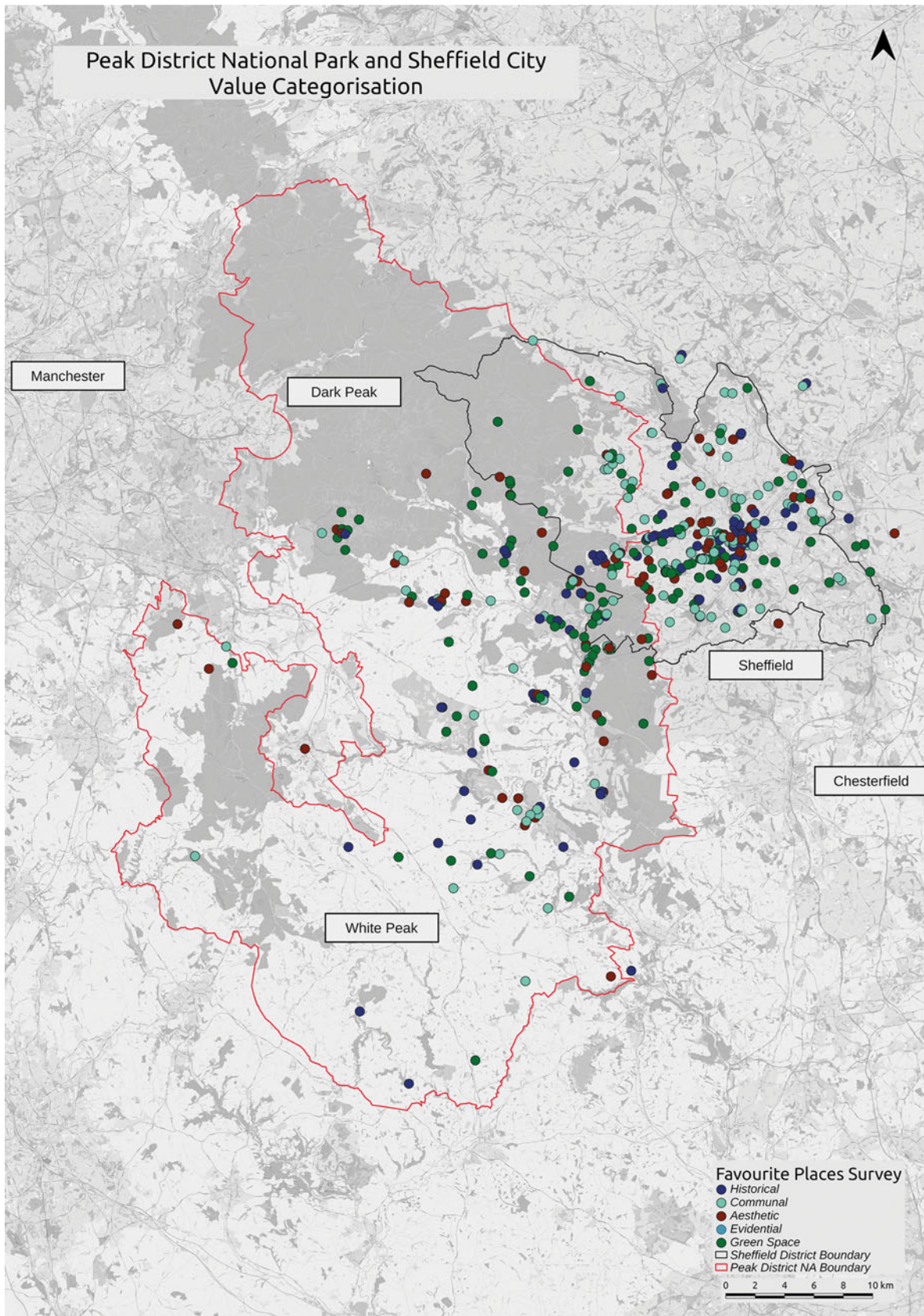
**Figure 7.** The category of communal values can be subdivided into private, spiritual and commemorative values, as set out in Historic England’s Conservation Principles. Themes emerging from the survey data show a trend to connect places based on green space, health and arts & culture. Yellow colours in the heatmap represent low numbers of stories, while darker red colours represent higher numbers of stories categorized in the respective categories. The graph shows that survey respondents based individual place attachment predominantly on aspects of communal, green space and health aspects.

‘Private’ value and two to ‘Spiritual’ value. This exemplifies another cluster of high coherence and close association with the modelled labels.

Cluster 27 comprised 12 documents labelled as ‘steel\_industry’ and ‘portland\_works’; the manual

assessment in this case showed that nine documents could be categorized as ‘Historical’ value, with the distinction of ‘Associative’ and ‘Illustrative’.

Figure 6 shows the distribution of documents (stories) over the developed catalogue of heritage





values, as the result of TM and manual assessment/refined categorization. The colour scale indicates the number of documents for a topic cluster within each cluster. Dark red cells indicate a high number of documents in one category and, therefore, a good correlation of document content in the Topic Modelling process. Rows with yellow cells show a great variety of value categories in one cluster and low coherence in the respective cluster overall. Furthermore, the graph shows the value categories most dominant in the perception of survey responses. 'Communal' values were mentioned 142 times, 'Historical' values 104 times, 'Aesthetic' values 79 times and 'Green Space' qualities 175 times (of these, 119 documents were categorized as 'Health'). Evidential value could only be identified in one document based on the description of the connection to a place. The topic allocation in Figure 5 can be correlated with Table 2. For example, Figure 6 shows Topic 8 to be most dominant, categorized as 'Historical' with labels in Table 2 defined as 'industrial\_heritage' and 'industrial\_history'. This example shows a case where manual annotation confirmed the outcome of the topic modelling process. The dominant topics and categories in the proposed categories 'Cultural' value and the subdivision of the introduced category of 'Green Space' values with 'Health' can be seen in Figure 7. Apart from the category 'Communal' values, 'Green Space' and 'Health' feature most strongly in public perceptions.

In general, topics labelled as 'young\_child' (Topic 6), 'friend\_family' (Topic 11), 'happy\_memory' (Topic 20), 'family\_live' (Topic 24), 'close\_heart' (Topic 35) were most dominantly associated with 'Communal' values (Spiritual, Private), while labels such as 'place\_walk' (Topic 30), 'love\_walk' (Topic 31), 'place\_feel' (Topic 19), were associated with 'Health'. Documents of the category 'Historical' value were labelled 'lead\_mine' (Topic 32), 'steel\_industry' (Topic 27), 'grade\_list' (Topic 13), 'industrial\_heritage' (Topic 8), 'architecture\_building' (Topic 7).

Examples of stories provided by survey participants which qualified for the different categories are provided in Supplementary Material 2. For example, category 'Green Space' with the subcategory 'Health' frequently referred to 'lockdown' and the COVID-19 pandemic. Unsurprisingly, these areas were primarily located in the city of Sheffield, where people

had to find open spaces for appropriate social distancing. A theme detected and correctly labelled by the algorithm was the practice of 'ash scatters' allocated to 'Communal'—more precisely 'Spiritual' value—primarily associated with specific open-space and viewpoints in the Peak District National Park. Another subcategory of 'Communal' value is shown here as subcategory 'Arts & Culture'.

These examples show that Topic Modelling found coherent topic clusters and labels suitable for a first insight and open-minded approach to the data with no preconceived assumptions and codes. The clustering can be useful for a first analysis as a basis for the preliminary coding in the second step of qualitative analysis and, subsequently, the allocation to specific heritage value typologies/categories.

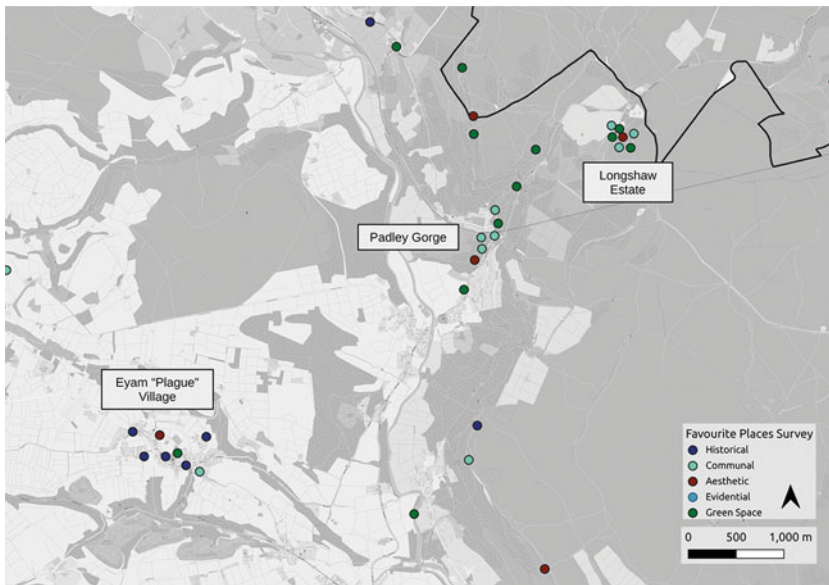
#### Category visualization

QGIS was used throughout this project. The visualization of favourite places in both study areas can be seen in Figure 8. The map is presented as an interactive webmap where the points on the map can be selected and information on the personal connection (stories) provided in the form of a pop-up window. Photographs provided by the participants illustrate the places as an extra layer. Individual values can be selected and presented on the map, for example, only showing locations labelled with an aesthetic value or values for the calming qualities of the landscape. Figure 8 provides a background overview of the categories as set out by HE's *Conservation Principles* and the proposed additional categories as a result of this study.

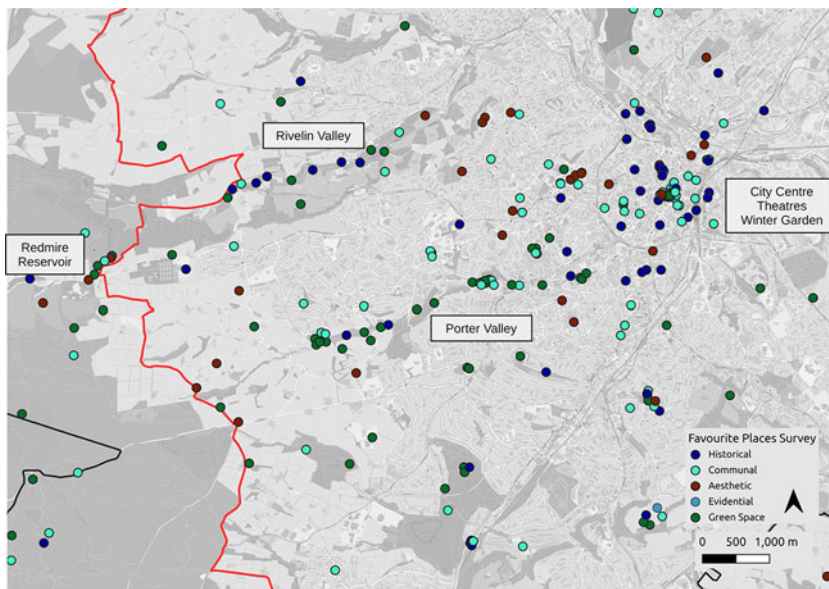
The detailed maps give an indication of the values given to places as identified in this study. The values emerged from the stories based on personal attachment to these places. Figure 9 focuses on a central part of the PDNP with the village of Eyam, which is known as the 'plague village'. An outbreak of bubonic plague in 1665 forced the residents of Eyam to isolate themselves under huge hardship to save the surrounding villages from a spread of the disease.<sup>13</sup> 'Historical' value dominates this location (n = 5).

Notable for the river valleys, leading from the PDNP into Sheffield, is the difference of value-based connection as shown in Figure 10. The Rivelin Valley is predominantly favoured for its 'Historical' value

**Figure 8** (opposite). Mapped overview of categories across the study areas. Distinct areas show clusters. (Map created in QGIS; data contain OS data © Crown copyright and database right 2022. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL).



**Figure 9.** Detail of the overview map (Fig. 8). The plague village Eyam (lower left) is predominantly valued by survey participants for the 'Historical' value ( $n = 5$ ). Padley Gorge (centre), one of the temperate rainforests of Britain (Shrubsole 2022), is valued for the 'Green Space' qualities ( $n = 4$ ) and 'Communal' values ( $n = 4$ ), similar to the National Trust's Longshaw Estate (upper right) with recognition given predominantly to 'Green Space' value ( $n = 3$ ) and 'Communal' value ( $n = 3$ ). (Map created in QGIS, © OpenStreetMap contributors; data contain OS data © Crown copyright and database right 2022.)



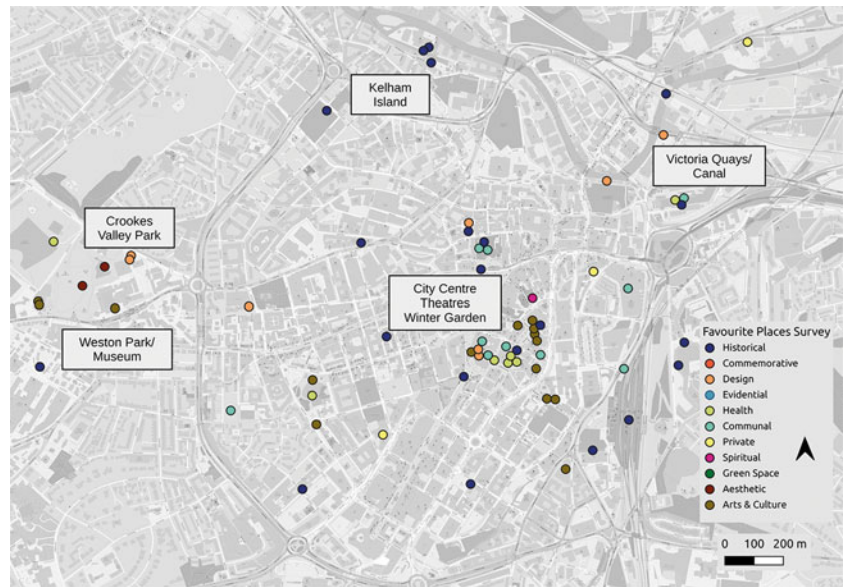
**Figure 10.** Detail of the overview map (Fig. 8). Notable is the difference between the public perception of the two valleys: Rivelin Valley (upper centre) and the Porter Valley (centre). Rivelin Valley, with its deep early industrial history present in the various sites of 'Wheels' along the river, is predominantly valued for its 'Historical' value ( $n = 6$ ), while the Porter Valley in the heart of the city is predominantly valued for its 'Green Space' value ( $n = 14$ ). (Map created in QGIS, © OpenStreetMap contributors; data contain OS data © Crown copyright and database right 2022.)

which refers to the up to 20 watermills along the watercourse appearing from around 1600.<sup>14</sup> In contrast, the Porter Valley to the south is predominantly favoured for its 'Green Space' value and health benefits as a green space in the city centre. A detailed view on the city centre, visualizing categories and subcategories in Figure 11, provides insight into the values of residents associated predominantly with 'Arts & Culture' and 'Historical' value (each  $n = 12$ ).

The mapping of values that has emerged from the survey shows that individual, subjective stories

have the potential to form an overall value pattern at landscape scale. The value distribution presents the varying and manifold valuation of places, showing how one place can have more than one meaning for the public. Nevertheless, despite a multiplicity of meanings and significances, current techniques allow the assessment, integration and visualization of public perception and place attachment in a form that can provide vital background for planning, alongside (notably and for example) Historic Landscape Characterization or HLC.

**Figure 11.** Detail of the overview map (Fig. 8). The legend provides a more fine-grained categorization. 'Historical' ( $n = 12$ ) and 'Arts & Culture' values ( $n = 12$ ) dominate. (Map created in QGIS, © OpenStreetMap contributors; data contain OS data © Crown copyright and database right 2022.)



### Limitations, biases and advantages of AI

It should be pointed out that qualitative research introduces bias at various stages of the research process. Our data collection introduced a limitation and bias as a convenience sample through the means of publication and completion of the survey. We fully acknowledge that marginalized groups have not been specifically explored and would certainly be a study that could follow up, be complementary to and contrast the findings of this paper (see Endnote 4). Both approaches—TM and Manual Observation/Annotation—introduce biases into the process of data analysis. The only automated process that is not influenced by the researcher is the phase where the TM model runs (the computation of the models). TM is based on a model algorithm, which is sometimes treated as a black box because of the complex mathematical ground on which the algorithms and statistical methods are based. However, to make the process as clear and transparent as possible, the choice of model algorithm and the definition of parameters must be documented. Appropriate models can be chosen depending on the analysis and data. Model parameters, such as iteration or number of topics, can be optimized and adapted to control the process.

The final assessment and analysis for the training data<sup>15</sup> is manually observed by the researcher to create a model based on the training data, which can then be applied to new data for an optimization of Topic Modelling in the specific field—in this case, the categorization of heritage values in cultural landscapes. This introduces the human factor, which in turn introduces researcher bias into the

methodology, similar to other qualitative research methods (i.e. NVivo or manual coding). Transparency and reproducibility can be achieved by rigorous documentation of the process.

By way of advantage, the use of TM allows topics and themes to be identified that emerge from or are latent within the data without having a preconceived set of codes. This approach eliminates the risk of researcher bias towards the topics introduced in the data analysis by preconceived codes. This constitutes a significant change in the way qualitative data can speak for themselves and can reveal patterns within the data, before the topics are then further analysed manually in the process of categorization.

AI offers the opportunity to create a reproducible, repeatable and automated workflow of processes, which would not be possible with various individual human assessors. The routine set-up for the analysis works on the same parameters in every iteration of the process. This makes it more reliable for the categorization of data provided by future surveys to review people-centred values on a rolling basis.

### Discussion

Our methodology builds on approaches tested in previous studies: on research projects focusing on a people-place connection (Dalglish & Leslie 2016; Natural England 2015; Primdahl & Kristensen 2016), and on approaches of other disciplines focusing on computer linguistics that provide methodologies for an application in heritage management (Goerz & Scholz 2010; Sassolini & Cinini 2010; Sporleder 2010). AI has been successfully applied in



Heritage Management, but not on a landscape scale (Bordoni *et al.* 2016; Condorelli *et al.* 2020; Fiorucci *et al.* 2020; Matrone *et al.* 2020). Our research methodology therefore combines the approach from disciplines such as data science, geography, archaeology and urban planning on a landscape scale. Using Topic Modelling and a Grounded Theory approach to analyse and integrate public survey data as spatial representation into the planning process offers a way to overcome the challenges qualitative data can often represent in terms of practical application.

Themes emerging from stories of place attachment provided by survey participants in the Peak District National Park and the City of Sheffield align well with some of the existing heritage value categorizations set out by Historic England's *Conservation Principles* (Historical, Evidential, Aesthetic, Communal) (65 per cent). Notable, however, is the lack of 'Evidential' value ( $n=1$ ) in the perceptions of people with just one case, perhaps due to the lack of documentation on this value subject in a publicly accessible format (e.g. information boards, accessible documentation). However, 'Communal' value, usually not prioritized among other values in the significance assessment process, represents almost one-third of the traditional value categories. Within the City of Sheffield, the valuing of the cultural element of the city centre was notable, for which we propose an additional subcategory of 'Arts & Culture' value (which comprised 15 per cent of all 'Communal' value). It also became clear that 35 per cent of all personal connections to place were based on experiences during the COVID-19 pandemic and with a strong orientation toward the use of landscape for mental and physical health benefits (Fagerholm *et al.* 2022). Therefore, we propose the introduction of the category 'Green Space' value to bridge the artificial division between natural and cultural heritage. Some 71 per cent of the data in this category showed a close relationship between nature and health.

These results show that single, individual opinions collected in a public online survey can map onto the landscape as heritage values based on people's individual, personal and subjective connections to place. The argument that individual opinions or the public view cannot be considered in the valuation process of landscapes (because it is too difficult, or the data are too diverse to be meaningful) is therefore no longer tenable. We will return to this point in the conclusion. Individual opinions—when aggregated—can create a generalizable pattern of heritage valuation. The detailed view of the two river valleys in Sheffield exemplifies how personal stories can form a coherent pattern of such values. We present

this example, therefore, as proof of concept for a method that is suitable for integrating people-centred landscape perception and values within a framework for assessing landscapes within planning and decision-making processes. We also argue for this grounded approach providing the basis for generating entirely new categories or sub-categories of values, rather than trying to force observed values into a predefined framework. While we suggested a bottom-up approach to obtain insight into the perceptions of people's understanding and valuing of the heritage in their everyday environment, the information should be understood as an additional layer of background information in the framework created by Historic England. This dataset will enable us to make decisions shaping people's neighbourhoods on a wider base of information including the crucial dataset of 'insider knowledge'. While this study provides a snapshot in time, complementary studies can contribute to a completion of the picture and further automation of the process can allow data collection and analysis on a rolling basis for up-to-date datasets (Tenzer 2022; Tenzer & Schofield *n.d.*).

Our research has highlighted potential areas for future work, for example, the automation of the topic clustering and labelling process, including the application of the proposed value categories in a supervised topic modelling approach. This is necessary as the unsupervised model only works with words used in the documents (stories) and not the English Heritage (2008) value categories. The topics are only sometimes sensible and coherent (meaning high coherence and meaningful clustering). In the future, the datasets created from this work could be used to train supervised deep-learning models that can then be used for fully automated labelling and the classification of users' responses.

Furthermore, there is scope to integrate a more refined approach to survey participants and the inclusion of visitors or demographically to explore variance in perception. A more fine-grained study could explore marginalized communities and the integration of heritage from a different perspective. Also, a focus on the visualization of qualitative data could develop a more nuanced representation of value categories in GIS. Finally, developing the theory of perception and value production in different contexts and communities could increase the understanding of what makes landscapes and people's everyday heritage vital for developing a sense of place, identity and belonging.

AI technologies are set to revolutionize the opportunities to understand and present datasets from heritage and archaeology. New tools and methods are currently being developed by a loose community of digital archaeologists, computer linguists

and data scientists that, together, have yet to find a more inter- and trans-disciplinary approach to the question of values. However, the trends to cooperate are promising and will bring into the future the methods of various disciplines concerned with the past and with landscape.

## Conclusion

This study has developed a novel method that allows the integration of people's connections with place into the assessment framework of landscape and heritage management. Furthermore, this research has introduced Topic Modelling combined with a Grounded Theory approach with a bias-reduced, time-efficient and repeatable method to interpret and categorize people-centred values of everyday heritage. This innovative approach to qualitative data collection for heritage planning can give local authorities and heritage organizations the opportunity to embed locally held values of heritage within landscape management processes.

The meanings and values people place in the landscapes of their everyday lives are more varied and personal than widely considered in heritage and landscape assessment. Individually held, subjective values form a category for heritage assessment that should not be underestimated for its capability to shape identity, create deep bonds and positively impact place-making [in terms of] more than just anecdotal evidence (Modesto & Waterton 2020). However, this is not about individual opinions; rather it is about gaining a deeper understanding of what drives the development of a sense of place, belonging and identity, and how this can change over time. It is also about recognizing the significance of locally held views and values in creating a more inclusive approach to heritage management than that which exists currently, at least in the Anglophone world. This study shows how diverse meaning and valuing is within communities, but that it can form a distinctive pattern across the landscapes and that this pattern can be both captured and accommodated within the planning process and in heritage management strategies.

## Notes

1. We focus on the presentation of the new method in this paper. A comparison with other qualitative approaches lies beyond the scope of this study (for this, see Jones 2017; Jones & Leech 2015; Nardi 2014).
2. They are also recognizing how problematic it is to distinguish the two (Harrison 2015).
3. To avoid a confusion of the widely used term 'social value' as identified by the Burra Charter (also after Johnston 1992), we decided to avoid this term as a subcategory for the Communal category as defined by English Heritage (2008). The category has been renamed as 'private value' including the family history and personal connections of individuals (see also Modesto & Waterton 2020).
4. It has to be noted that the data source and method of data collection represent a convenience sample and are limited and biased by various factors inherent in this approach, i.e. online accessibility, IT literacy, general interest in heritage, interest in participating in local heritage, exclusion from access to the countryside and heritage, etc. The survey showed that 54.2 per cent of participants were in the age brackets of 56–75 and a further 20 per cent in the age bracket of 46–55. 91.6 per cent identified themselves as White British which reveals a limitation of the result towards the dominant resident group. 54 per cent were female participants as opposed to 43 per cent male, 1 per cent non-binary and 2 per cent who answered 'preferred not to say'.
5. As a researcher bias, we define the preconceived assumptions that are noted at the beginning of a qualitative research process to ensure that this bias is clarified throughout the process. However, other sources of bias are introduced in other stages of the research similar to other qualitative methodologies.
6. <https://www.sheffield.gov.uk/>
7. This was reinforced in the publication text on social media and as one of the first survey questions (see Supplementary Material 3).
8. In this project we use QGIS, a free, open-source platform <https://www.qgis.org/en/site/>
9. <https://www.qualtrics.com/blog/qualtrics-survey-software-free/>
10. For 'private values' see endnote 3.
11. We recognize that this blunt interpretation could introduce interpretative limitations and that labelling a place described as 'magical' as 'spiritual' may not align directly with conventional definitions of spirituality. However, categorization demands a degree of flexibility and decisions that may not be free of limitations.
12. 'Communal value' represents the shared value of communities, which is based on shared histories, beliefs, or myths. While these are established independently from the individual, an accumulation of individual values can form a different level of shared value when present in a wider group of the community and are, therefore, a different facet to the communal value, such as commemorative or spiritual.
13. <https://www.eyamvillage.org.uk/>
14. <https://www.joindupheritagesheffield.org.uk/groups/rivelin-valley-conservation-group>
15. Part of the survey dataset that is set apart to train the model for an automatic categorization of new data; test data is the remaining part of the survey dataset that is used to assess the performance of the model.

## Acknowledgement

This project was carried out as a part of an AHRC/UKRI WRoCAH-funded PhD project at the University of York. Grant reference number: AH/R012733/1. The authors want to thank Anna Badcock (PDNPA) and Zac Nellist (SYAS) for their support for this project and help to publish the survey through organizational channels.

## Supplementary materials

Supplementary materials may be found at <https://doi.org/10.1017/S0959774323000203>

Martina Tenzer  
Department of Archaeology  
University of York  
King's Manor  
Exhibition Square  
York YO1 7EP  
UK  
Email: [martina.tenzer@york.ac.uk](mailto:martina.tenzer@york.ac.uk)

John Schofield  
Department of Archaeology  
University of York  
King's Manor  
Exhibition Square  
York YO1 7EP  
UK  
Email: [john.schofield@york.ac.uk](mailto:john.schofield@york.ac.uk)

## References

- Abram, M.D., K.T. Mancini & R.D. Parker, 2020. Methods to integrate Natural Language Processing into qualitative research. *International Journal of Qualitative Methods* 19, 1609406920984608.
- Avrami, E., 2009. Heritage, values and sustainability, in *Conservation: Principles, dilemmas and uncomfortable truths*, eds A. Richmond & A.L. Bracker. Oxford: Butterworth-Heinemann, 177–83.
- Avrami, E., R. Mason & M. de la Torre, 2000. *Values and Heritage Conservation*. Los Angeles (CA): Getty Conservation Institute.
- Banks, G.C., H.M. Woznyj, R.S. Wesslen & R.L. Ross, 2018. A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology* 33(4), 445–59.
- Blei, D.M., A.Y. Ng & M.I. Jordan, 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3(null), 993–1022.
- Bonnell, V.E. & L.A. Hunt, 1999. Introduction, in *Beyond the Cultural Turn: New directions in the study of Society and culture*, eds V.E. Bonnell & L.A. Hunt. Berkeley/London: University of California Press, 1–32.
- Bordoni, L., F. Mele & A. Sorgente, 2016. *Artificial Intelligence for Cultural Heritage*. Cambridge: Cambridge Scholars Publishing.
- Bradley, A., V. Buchli, G. Fairclough, D. Hicks, J. Miller & J. Schofield, 2004. *Change and Creation: Historic landscape character 1950–2000*. London: English Heritage.
- Byrne, D., 2008. Heritage as social action, in *The Heritage Reader*, eds G. Fairclough, R. Harrison, J. Jameson Jr & J. Schofield. Abingdon: Routledge, 149–73.
- Byrne, D. & G.B. Ween, 2015. Bridging cultural and natural heritage, in *Global Heritage: A reader*, ed. L. Meskell. Chichester: Wiley Blackwell, 94–111.
- Cai, Z., A. Siebert-Evenstone, B. Eagan & D.W. Shaffer, 2021. Using Topic Modeling for code discovery in large scale text data, in *Advances in Quantitative Ethnography*, eds A.R. Ruis & S.B. Lee. Cham: Springer International, 18–31.
- Chang, J., S. Gerrish, C. Wang, J. Boyd-graber & D. Blei, 2009. Reading tea leaves: how humans interpret topic models, in *Advances in Neural Information Processing Systems 22*, eds Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams & A. Culotta. Red Hook (NY): Curran Associates, Inc.
- Charmaz, K., 2006. *Constructing Grounded Theory: A practical guide through qualitative analysis*. Los Angeles/London: Sage.
- Chun Tie, Y., M. Birks & K. Francis, 2019. Grounded theory research: a design framework for novice researchers. *SAGE Open Medicine* 7, 2050312118822927.
- Condorelli, F., F. Rinaudo, F. Salvatore & S. Tagliaventi, 2020. A neural networks approach to detecting lost heritage in historical video. *ISPRS International Journal of Geo-Information* 9(5), 297.
- Cosgrove, D., 2004. Landscape and Landschaft. *Geman Historical Institute Bulletin* 35(Fall), 57–71.
- Council of Europe, 2000. *European Landscape Convention (Florence Convention)*. (European Treaty series 176.) Strasbourg: Council of Europe.
- Council of Europe, 2005. *Council of Europe Framework on the Value of Cultural Heritage for Society: Explanatory Report*. Faro: Council of Europe.
- Creswell, J.W., 2017. *Research Design: Qualitative, quantitative, and mixed methods approaches*. London: Sage.
- Dalglis, C. & A. Leslie, 2016. A question of what matters: landscape characterisation as a process of situated, problem-orientated public discourse. *Landscape Research* 41(2), 212–26.
- English Heritage, 2008. *Conservation Principles, Policies and Guidance*. Swindon: English Heritage.
- Fagerholm, N., K. Samuelsson, S. Eilola, et al., 2022. Analysis of pandemic outdoor recreation and green infrastructure in Nordic cities to enhance urban resilience. *npj Urban Sustainability* 2(1), 1–14.
- Fiorucci, M., M. Khoroshiltseva, M. Pontil, A. Traviglia, A. Del Bue & S. James, 2020. Machine learning for



- cultural heritage: a survey. *Pattern Recognition Letters* 133, 102–8.
- Franzosi, R., W. Dong & Y. Dong, 2022. Qualitative and quantitative research in the humanities and social sciences: how natural language processing (NLP) can help. *Quality & Quantity* 56, 1–31.
- Gatersleben, B., K.J. Wyles, A. Myers & B. Opitz, 2020. Why are places so special? Uncovering how our brain reacts to meaningful places. *Landscape and Urban Planning* 197, 103758.
- Gibson, L. & J. Pendlebury, 2016. Introduction: Valuing Historic Environments, in *Valuing Historic Environments*, eds L. Gibson & J. Pendlebury. London/New York: Routledge, 1–18.
- Ginzarly, M. & F. Srour, 2022. Cultural heritage through the lens of COVID-19. *Poetics (Amst.)* 92, 101622.
- Goerz, G. & M. Scholz, 2010. Adaptation of NLP techniques to cultural heritage research and documentation. *Journal of Computing and Information Technology* 18(4), 317.
- Grenville, J., 2007. Conservation as psychology: ontological security and the built environment. *International Journal of Heritage Studies* 13(6), 447–61.
- Harrison, R., 2010. Heritage as social action, in *Understanding Heritage in Practice*, ed. S. West. Manchester: Manchester University Press, 240–76.
- Harrison, R., 2015. Beyond ‘natural’ and ‘cultural’ heritage: toward an ontological politics of heritage in the Age of Anthropocene. *Heritage & Society* 8(1), 24–42.
- Harrison, R., 2020. *Heritage Futures: Comparative approaches to natural and cultural heritage practices*. London: UCL Press.
- Historic England, 2022. *Survey of COVID-19 Effects on the Heritage Sector*. <http://historicengland.org.uk/coronavirus/heritage-sector/survey/>
- ICOMOS, 1975. *The Declaration of Amsterdam – 1975 – International Council on Monuments and Sites*. Amsterdam: ICOMOS.
- ICOMOS, 1979. *Burra-Charter\_1979*. Canberra: Australia ICOMOS.
- ICOMOS, 2013. *The Burra Charter: The Australia ICOMOS Charter for Places of Cultural Significance 2013*. Canberra: Australia ICOMOS.
- Ireland, T. & J. Schofield (eds), 2015. *The Ethics of Cultural Heritage*. (Ethical Archaeologies: The Politics of Social Justice 4.) New York/Heidelberg/Dordrecht/London: Springer.
- Johnston, C., 1992. *What Is Social Value? A discussion paper*. Canberra: Australian Government Publishing Service.
- Jones, S., 2017. Wrestling with the social value of heritage: problems, dilemmas and opportunities. *Journal of Community Archaeology & Heritage* 4(1), 21–37.
- Jones, S. & S. Leech, 2015. *Valuing the Historic Environment: A critical review of existing approaches to social value*. Manchester: University of Manchester.
- Jones, T., W. Doane & M. Attbom, 2021. *Package ‘textmineR’: Functions for Text Mining and Topic Modeling. Version 3.0.5*.
- Jones, T.W., 2021. *Topic modeling*. 2021. [https://cran.r-project.org/web/packages/textmineR/vignettes/c\\_topic\\_modeling.html](https://cran.r-project.org/web/packages/textmineR/vignettes/c_topic_modeling.html)
- Latour, B., 1993. *We Have Never Been Modern*. Cambridge (MA): Harvard University Press.
- Lau, J.H., D. Newman & T. Baldwin, 2014. Machine reading tea leaves: automatically evaluating topic coherence and topic model quality, in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Gothenburg: Association for Computational Linguistics, 530–39.
- Leeson, W., A. Resnick, D. Alexander & J. Rovers, 2019. Natural Language Processing (NLP) in qualitative public health research: a proof of concept study. *International Journal of Qualitative Methods* 18, 1609406919887021.
- Matrone, F., E. Grilli, M. Martini, M. Paolanti, R. Pierdicca & F. Remondino, 2020. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS International Journal of Geo-Information* 9(9), 535.
- Modesto, G. & E. Waterton, 2020. The elite and the everyday in the Australian heritage field, in *Fields, Capitals, Habitus: Australian culture, inequalities and social divisions*, eds T. Bennett, D. Carter, M. Gayo, M. Kelly & G. Noble. London/New York: Routledge.
- Nardi, S.D., 2014. Senses of place, senses of the past: making experiential maps as part of community heritage fieldwork. *Journal of Community Archaeology & Heritage* 1(1), 5–22.
- Natural England, 2015. *Econets, Landscape & People: Integrating people’s values and cultural ecosystem services into the design of ecological networks and other landscape change proposals*. (NECR180.) Cambridge: Natural England.
- Odacioglu, E.C. & L. Zhang, 2022. Text mining for rendering theory: integrating topic modeling to grounded theory. SSRN. <http://dx.doi.org/10.2139/ssrn.4141372>
- PDNPA, n.d. *Peak District Facts*. <https://www.peakdistrict.gov.uk/learning-about/news/70-years-of-the-peak-district-national-park/peak-district-facts>
- Pearson, M., 1995. *Looking After Heritage Places: The basics of heritage planning for managers, landowners and administrators*. Carlton (VIC): Melbourne University Press.
- Primdahl, J. & L.S. Kristensen, 2016. Landscape strategy making and landscape characterisation—experiences from Danish experimental planning processes. *Landscape Research* 41(2), 227–38.
- Sassolini, E. & A. Cinini, 2010. Cultural heritage: knowledge extraction from web documents, in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, eds N. Calzolari, K. Choukri, B. Maegaard, et al. Valletta: European Language Resources Association (ELRA).

- Schofield, J., 2014. Heritage expertise and the everyday: citizens and authority in the twenty-first century, in *Who Needs Experts? Counter-mapping cultural heritage*, ed. J. Schofield. Farnham: Ashgate, 1–12.
- Shrubsole, G., 2022. *The Lost Rainforests of Britain*. London: William Collins.
- Smith, L., 2006. *Uses of Heritage*. London: Routledge.
- Sporleder, C., 2010. Natural Language Processing for cultural heritage domains. *Language and Linguistics Compass* 4(9), 750–68.
- Stephenson, J., 2008. The cultural values model: an integrated approach to values in landscapes. *Landscape and Urban Planning* 84, 127–39.
- Tenzer, M., 2022. Tweets in the Peak: Twitter analysis – the impact of Covid-19 on cultural landscapes. *Internet Archaeology* 59. <https://doi.org/10.11141/ia.59.6>
- Tenzer, M. & J. Schofield, n.d. The eye of the beholder: using Topic Modelling to identify landscape perception and place attachment – case studies from a national park. (Manuscript in review.)
- txq94, 2022. *Text-Analytics\_LDA*. [https://github.com/txq94/Text-Analytics\\_LDA](https://github.com/txq94/Text-Analytics_LDA)
- Traviglia, A., 2022. Digital technologies (including AI) for cultural heritage in the framework of the Council of Europe's heritage conventions, in *Creating Our Future: Creativity and cultural heritage as strategic resources for a diverse and democratic Europe*. Strasbourg: Council of Europe. <https://www.coe.int/en/web/cdcpp-committee/-/-creating-our-future-creativity-and-cultural-heritage-as-strategic-resources-for-a-diverse-and-democratic-europe>
- Tuan, Y.-F., 1980. Rootedness versus sense of place. *Landscapes* 24(1), 3–8.
- Tuan, Y.-F., 1990. *Topophilia: A study of environmental perceptions, attitudes, and values*. New York (NY): Columbia University Press.
- UNESCO, 2021. *Impact of the COVID-19 Pandemic on World Heritage and Responses by the Secretariat*. (WHC/21/44.COM/INF.5A.2.) Paris: Unesco.
- Verschoof-van der Vaart, W.B., K. Lambers, W. Kowalczyk & Q.P.J. Bourgeois, 2020. Combining deep learning and location-based ranking for large-scale archaeological prospection of LiDAR data from the Netherlands. *ISPRS International Journal of Geo-Information* 9(5), 293.
- Waterton, E., L. Smith & G. Campbell, 2006. The utility of discourse analysis to Heritage Studies: the Burra Charter and social inclusion. *International Journal of Heritage Studies* 12(4), 339–55.
- Welsh, E., 2002. Dealing with data: using NVivo in the Qualitative Data Analysis process. *Forum: Qualitative Social Research* 3(2). <https://doi.org/10.17169/fqs-3.2.865>
- West, S., 2010. Introduction, in *Understanding Heritage in Practice*, ed. S. West. Manchester: Manchester University Press, 1–6.

### Author biographies

**Martina Tenzer** is a PhD researcher at the University of York focusing on methods for analysis and visualization of public perception in cultural landscapes. Prior to this, she gained degrees at the University of Oxford and the University of Heidelberg and worked in commercial archaeology. Her research interests include Contemporary Archaeology, Historic Landscape Characterisation, landscapes of biodiversity and climate change impact, Artificial Intelligence in heritage management, mapping and visualization of complex, abstract concepts, QGIS.

**John Schofield** teaches cultural heritage management and contemporary archaeology in the Archaeology Department, University of York. Prior to this he worked for Historic England's Characterisation Team, where he developed a particular interest in understanding people's social and communal values for everyday places. John holds adjunct status at the universities of Turku (Finland) and Flinders and Griffith (Australia). He is Corresponding Fellow of the Australian Academy of the Humanities and Fellow of the Society of Antiquaries of London.