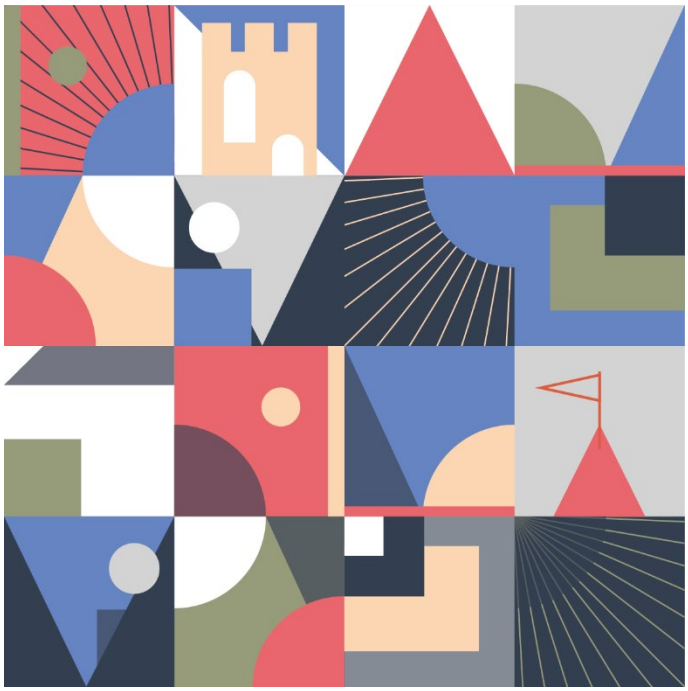
A feasibility study and preliminary framework for an alternative heritage sector statistics methodology

Prepared for the Department for Culture, Media and Sport (DCMS)

August 2024



alma economics

About the authors

alma economics

Alma Economics combines unparalleled analytical expertise with the ability to communicate complex ideas clearly.

www.almaeconomics.com

About the commissioning organisation

Department for Culture, Media & Sport

This independent analysis was commissioned by the Department for Culture, Media & Sport (DCMS). The analysis and findings are those of the authors and do not represent the views of DCMS.

www.gov.uk

Table of Contents

Executive summary

The heritage sector refers to not just historic monuments, sites and buildings, but also a wide range of assets and landscapes that are connected to history in some way. Interest in the sector has recently increased due to the UK Government's ratification of the 2003 UNESCO Convention for the Safeguarding of Intangible Cultural Heritage. As heritage can be found across an array of economic sectors and sub-sectors, the lack of clear boundaries for the heritage sector has made it challenging to quantify the sector's impact on the UK economy. The current proxy for the heritage sector captured in the Department for Culture, Media and Sport (DCMS) Economic Estimates, the single SIC code 91.03, *Operation of historical sites and buildings and similar visitor attractions*, is widely viewed as too narrow of a definition that underestimates the size and impact of the sector, with one alternative measure known as SIC-SOC mapping first developed by the economic consultancy Cebr in 2017 and currently used by Historic England.

To address this issue, DCMS commissioned Alma Economics to carry out a feasibility study of different approaches that could be used to produce a single reliable estimate of the economic contribution of heritage organisations to the UK economy. There were four important specifications on the scope of this project:

1. The project sought to examine the methodologies to measure the economic contribution of heritage through Gross Value Added (GVA), total employment, number and size of businesses in the sector and imports/exports of goods and services. Each approach considered was assessed against six criteria: coverage, disaggregation, robustness, feasibility, replicability and comparability.
2. While heritage generates significant impacts on physical and mental wellbeing as well as broader societal welfare, non-market price impacts would be better captured by different frameworks or approaches not reviewed in this study (these impacts are part of DCMS' ongoing work on developing a formal approach to value culture and heritage assets called Culture and Heritage Capital). These tools complement the approaches considered in this study, as taken together they can present a holistic view of how heritage impacts UK society as a whole.
3. The project sought to establish a formal economic definition of the heritage sector by designating a set of industries and occupations aligning with the UK national accounts framework as "heritage industries" and "heritage occupations".

4. We recognise that intangible heritage assets, referring to non-physical aspects of heritage including traditions or customs, are an important aspect of the value of heritage and may be linked to or supported by additional industries and occupations beyond those listed in the report. Due to a lack of comprehensive data sources which span across the UK for intangible assets of heritage, this was beyond the scope of the report but could be a focus for future research. This project, therefore, focuses on tangible or material heritage assets, including physical artefacts or buildings.

Four approaches were considered by the research team in the feasibility study:

- **Dynamic mapping** classifies industries based on their "heritage intensity" (the proportion of employment within the industry in heritage occupations) (Bakhshi et al., 2013). This approach is currently used within the DCMS Economic Estimates for the creative industries sector.
- **SIC-SOC mapping** addresses the fragmented economic structure of heritage by mapping heritage occupations (SOC codes) to their corresponding industry (SIC codes), capturing both industries that are fully part of the heritage sector and industries which may engage in heritage activities, but heritage does not represent their primary focus. This approach was developed by the economic consultancy Cebr and is currently used by Historic England as part of its Heritage Counts publications (Cebr, 2023). While including both occupations and industries in this approach means it better captures the breadth of the heritage sector, it cannot be directly used to produce estimates of the number and size of businesses in the sector or trade statistics.
- **Satellite accounts** refer to statistical systems aiming to describe the economic contribution of specific areas of the economy that are not directly presented in the country's core national accounts framework. They are widely used internationally for sectors that do not clearly map to SIC codes, such as tourism, the digital economy, and civil society. Currently, almost no countries have developed a standalone heritage satellite account, so our review covered cultural satellite accounts which included a heritage component, focusing on Canada, Finland and the framework developed by ESPON spanning 11 European countries (Statistics Canada, 2024; Statistics Finland, 2023; ESPON, 2020). While a heritage satellite account would potentially produce the most accurate estimate of the economic impact of the heritage sector, it would also be the most difficult approach to implement due to data requirements: Canada and Finland use administrative datasets that are not publicly available, while ESPON uses proxy indicators

(called “keys” within their report) to calculate the fraction of economic activity related to the material cultural heritage of individual industries.

- **Big data** approaches use large online datasets (such as descriptions of businesses and the products/services they provide found on their websites) to estimate the size of the heritage sector. After data on individual businesses has been collected, a machine learning model can be trained to classify businesses as “heritage” or “not heritage” based on information published on the business’s own website as well as on third-party websites such as news sites, forums or social media. By looking at individual businesses as part of a bottom-up approach, big data methods would not require proxy variables to estimate the heritage component of industries such as tourism or construction. However, collecting online data may be resource-intensive, quality assurance of this data is extremely difficult (given the volume of data collected) and the approach cannot be used to produce GVA or employment estimates without first mapping it to Companies House data (which includes SIC codes of registered businesses).

Based on these findings, SIC-SOC mapping emerged as the best of the four approaches in the short-term to provide economic estimates of the heritage sector due to its coverage and feasibility. Importantly, the SIC-SOC mapping approach also provides a flexible template to develop more precise estimates over time by:

- Expanding the list of heritage industries and occupations:
    - Industries included in the ESPON heritage framework such as information and communication technologies (ICT) and insurance
    - Heritage research at higher education institutions
    - Heritage occupations found in heritage career guides and job vacancy postings on heritage organisation websites, mapped to SIC codes
    - Other occupations supported by grants provided by heritage organisations or intended to support heritage projects
- Incorporating additional data on heritage activities and outputs to improve apportionment methods for heritage industries and occupations:
    - Reviewing a broader range of data sources for construction, planning and design-related occupations that could inform potential changes to buildings considered heritage assets

- Collecting firm-specific data on heritage activities through systematic analysis of job vacancy descriptions, heritage grant monitoring frameworks and regular surveys of a sample of heritage umbrella organisations and businesses
- Collecting data on “consumption” of heritage outputs to help apportion tourism-related industries such as accommodation and food services, by drawing on national tourism surveys or surveys conducted by individual heritage sites

These recommendations have been set out to support DCMS in an eventual transition to a fully-developed heritage satellite account, the approach that most robustly captures the full set of linkages the heritage sector has with the broader UK economy.

Introduction

Project aims and objectives

Heritage is a wide-ranging sector that has been defined by the National Lottery Heritage Fund as "anything from the past that [individuals or communities] value and want to pass on to future generations."[1] Heritage is commonly used to refer to historic monuments, sites and buildings (such as Stonehenge or the Tower of London) but can also encompass almost any asset or landscape that is connected to history in some way.[2] The Department for Culture, Media and Sport commissioned Alma Economics to conduct research on two areas. First, Alma Economics was asked to develop a formal definition of the heritage sector by designating a set of industries aligning with the UK national accounts framework as "heritage industries": in other words, industries that are linked to the protection of assets inherited from past generations to be preserved for future generations. Second, Alma Economics was asked to develop a methodology to produce a single reliable estimate of the economic contribution of heritage organisations to the UK economy.

Within the UK, the heritage sector stands as a cornerstone of cultural preservation and historical appreciation, but there currently exists no formal definition. The heritage sector is broadly characterised by a fragmented market structure, with its economic contribution stemming from an array of sectors and sub-sectors and transcends confinement from a single industry. This lack of formal definition and boundary into what economic activities constitute heritage make it challenging to quantify its impact on the UK economy.

A proxy for the heritage sector is currently captured by DCMS estimates for heritage through its Economic Estimates.[3] The DCMS Economic Estimates use the single Standard Industrial Classification (SIC) code 91.03, *Operation of historical sites and buildings and similar visitor attractions*. This narrow definition and measurement results in what is widely viewed as an underestimate of the economic value of heritage, with other estimates of the sector's economic impact ranging significantly across sources. To address this gap in estimating the

[1] Available here.

[2] As a benchmark definition, DCMS' Culture and Heritage Capital programme groups culture and heritage assets into five categories: build historic environment, landscapes and archaeology, collections and moveable heritage, performance and performance venues, and digital assets. For further details, see here.

[3] Available here.

economic value of heritage, this project aims to assess available methodologies and offer insights into the best way to develop reliable estimates to measure the economic contribution of heritage in the UK economy.

In addition to built heritage, it is important to recognise that heritage also includes both intangible heritage (knowledge and traditions such as social practices or cultural events) and natural heritage (landscapes, marine environments and native flora and fauna). This definition is used by the Natural Lottery Heritage Fund in determining which projects are eligible for funding, as well as umbrella organisations for the heritage sector such as the Heritage Alliance.[4] In addition, in December 2023, the UK government set out its commitment to ratify the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage, with a public consultation launched to define and identify intangible cultural heritage in the UK.[5] However, because this project is intended to develop an estimate of the *economic contribution* of heritage organisations, the scope of this project only focuses on tangible heritage (physical assets or artefacts)[6]. This is because tangible heritage can be mapped in a more straightforward manner to industries (defined by SIC codes) in the UK national accounts framework, the standard economic approach for estimating the contribution of individual sectors to the UK economy. Due to a lack of defined metrics for intangible heritage assets, this is not within the scope of this project but is an important area for future research to understand the economic contribution of intangible heritage assets.

Methodology

To understand existing methodologies that could potentially be used to estimate the economic contribution of heritage organisations to the UK economy, our research team reviewed four methodologies: (i) the dynamic mapping methodology used by DCMS to estimate the economic contribution of creative industries, (ii) the SIC-SOC mapping developed by Centre for Economic and Business Research (Cebr) and used by Historic England to estimate the economic contribution and impacts of England's heritage sector in the UK, (iii) international satellite account approaches that encompass heritage (such as

[4] A description of the Heritage Alliance's mission can be found here.

[5] The consultation document can be found here.

[6] Definitions of tangible heritage focus on its material nature, including definitions set out by the EU-funded RICHES (Renewal, innovation & Change: Heritage and European Society) project and the Cambridge Heritage Research Centre, which can be found here.

cultural satellite accounts), and (iv) approaches that work with "big data" (large publicly-accessible online datasets). These four methodologies were compared against the baseline DCMS Economic Estimates methodology using a single four-digit SIC code.

For each of these methodologies, we reviewed published reports describing each methodology, including assumptions used and data sources required as inputs. This included a review of datasets owned by the Office for National Statistics and other Government departments, related entities and regulators (such as the Charity Commission for England and Wales), as well as arms-length bodies, sector support organisations and commercial entities operating in the heritage sector. In addition, for the dynamic mapping and SIC-SOC mapping methodologies, our research team sought to replicate these estimates as far as possible using publicly available datasets to understand their feasibility.

For satellite accounts specifically, we reviewed methodologies to measure the cultural sector used by 11 countries (United States, Australia, Canada, New Zealand, Japan, the Netherlands, Colombia, France, Spain, Finland, Scotland). In addition, three frameworks published by EU bodies were also reviewed. Most of these methodologies were based on a culture satellite account approach. It is important to note that this review was primarily limited by availability of English-language technical papers published by the relevant national statistics authority, and only one satellite account methodology reviewed (ESPON's cultural heritage framework) focused specifically on the heritage sector. All other methodologies reviewed focused on the broader set of creative or cultural industries, though Canada, France, Spain and the Netherlands defined a separate heritage sub-domain. The summary presented in this report focuses on approaches developed by Canada, Finland and ESPON, as these were viewed to be the most representative or informative across all approaches reviewed.

Finally, we reviewed a number of heritage career guides published by universities and sector organisations, including the University of Edinburgh Centre for Research Collections; Heriot-Watt University and the Heritage Alliance; and the Museums Association.[7] We also reviewed occupations posted on job boards of large-scale heritage organisations as well as cultural organisations and museums, as well as job descriptions posted in general online databases such as Indeed that included the keywords "heritage" or "historic".

[7] The Centre for Research Collections's Careers in Heritage guide can be found here. Heriot-Watt University's Heritage Careers Guide 2023-24 can be found here. The Museums Association's Museums, Galleries & Heritage Careers guide can be found here.

In parallel with our review, we also conducted interviews with leading academic experts in heritage research, project managers in Government departments developing satellite accounts for other sectors and stakeholders in heritage sector organisations. A full list of interviewees is included in the “Feedback from stakeholder engagement” section of the report.

Methodology and data review

This section provides an overview of the four potential heritage methodologies reviewed by our team. As mentioned in the above methodology section, the following methodologies were reviewed: (i) the dynamic mapping methodology used by DCMS to estimate the economic contribution of creative industries, (ii) the SIC-SOC mapping developed by Cebr and used by Historic England to estimate the economic contribution and impacts of England's heritage sector in the UK, (iii) international satellite account approaches which encompass heritage, specifically focusing on Canada, Finland and the ESPON Heritage framework, and (iv) approaches that work with "big data" (large publicly-accessible online datasets).

To get an initial sense of the outputs currently produced, we also reviewed the baseline DCMS Economic Estimates methodology. Economic estimates measure the economic contribution to the UK for each sector of which DCMS covers, many of which may overlap or be adjacent to the heritage sector, including (i) civil society, (ii) cultural sector, and (iii) tourism. For SIC codes that fit within each of these sectors and others within DCMS responsibility, various indicators are used to measure economic output. These indicators include (i) Gross Added Value (GVA) in current prices, (ii) chained volume measures (GVA with inflation considered), (iii) number of jobs, (iv) earnings, (v) exports and imports of goods and services, and (vi) the number of businesses. Indicators are currently sourced from various ONS datasets, including supply-use tables, the Annual Business Survey, Inter-departmental Business Register (IDBR), balanced GVA data, the Annual Population Survey, and others.

For each methodology, we offer (i) an overview of how the methodology works and how relevant heritage estimates are calculated, (ii) information on what data sources are required to replicate the methodology, and (iii) how a methodology might be replicated for the heritage sector.

Dynamic mapping

Overview

Dynamic mapping was initially developed by Nesta to classify creative industries in the UK and assess the comprehensiveness of creative industries defined by the UK government (Bakhshi et al., 2013). This method is used within the DCMS Economic Estimates to establish the sectoral definition of which industries should be included as "creative" when calculating the economic contribution of Creative Industries.[8]

Dynamic mapping is considered "dynamic" as it is a data-driven approach which uses the current labour force to inform whether an industry should be classified under the sector of interest. For creative industries, a creative intensity is calculated using the proportion of creative *occupations* in any given creative *industry*. This is shown in the formula below, where subscript *i* refers to a specific industry SIC code.

$$creative\ intensity_i = \frac{total\ creative\ occupations_i}{total\ jobs_i}$$

Since the classification is based on the number of occupations, it captures broader changes in the labour force and therefore is responsive to real structural changes in the economy rather than overreacting to small fluctuations in the data (Bakhshi et al., 2013).

Industries are candidates for inclusion as a creative industry if they have (i) more than 6,000 total jobs within the industry, and (ii) have a creative intensity threshold of higher than 30%. Once an industry is included, it is then fully considered a creative industry even if there are portions of the industry that may not be deemed creative. The threshold for creative intensity was set so that there was an equal likelihood of type I error (wrongly classifying a creative industry) and type II error (wrongly classifying a non-creative industry) so that the threshold lies "an equal number of standard deviations from the mean of the distribution" (Bakhshi et al., 2013).

Data sources and requirements

The DCMS Economic Estimates for Creative Industries uses dynamic mapping and the creative intensity calculation to define which industries should be included as contributing to the wider creative economy. This then feeds into the Creative Industries estimates published

[8] The DCMS Creative Industries Economic Estimates Methodology can be found here.

in DCMS Economic Estimate official statistics, including (i) GVA, (ii) Employment, (iii) Earnings, (iv) Trade – exports and imports of services, and (v) Business Demographics, including the number and characteristics of businesses.

To calculate creative intensity and determine which industries should be included as creative, DCMS uses 4-digit SIC and SOC codes from the Annual Population Survey (APS), which is derived from the Labour Force Survey (LFS). Current industries are determined from a combined sample of 2011 and 2012 APS data. The APS is the UK's largest continuous household survey, covering a diverse array of topics, including personal characteristics, labour market status, work characteristics, education, and health.

The list of SOC codes that are included as a creative occupation comes from a 2013 DCMS consultation.[9] Both main and secondary occupations are included in the total number of creative occupations. For each industry, the total number of creative SOC codes is then divided by the total number of occupations (creative and non-creative), resulting in a creative intensity for each industry. For creative industries the threshold is 30%, so therefore any industry with a creative intensity above 30% and more than 6,000 jobs is labelled as creative, while any industry below this threshold is then labelled as not creative. A full description of how to replicate this method and its application for the heritage sector can be found in Annex 1.

SIC-SOC mapping

Overview

The SIC-SOC mapping methodology was developed as a national model in 2017, following a pilot by Ortus Economic Research. The methodology has since developed and is now produced annually as part of Historic England's Heritage Counts publication series entitled "Heritage and the Economy". The approach is used to assess the economic contributions and impacts of England's heritage sector in the UK. Subsequent studies by Cebr in 2018, 2019, 2020, 2022 and 2023 replicated this methodology, providing estimates of key indicators up to two years prior.

These studies are aimed to generate a range of statistical data to demonstrate different aspects of the value of England's heritage sector. These include its direct contribution to the economy in terms of GVA and employment, as well as the indirect and induced multiplier

[9] The consultation document can be found here.

impacts. Indirect impacts arise through the activities stimulated in the heritage sector supply chain, whereas induced impacts occur through the activities stimulated in the wider economy when the employees linked to both direct and indirect heritage activities spend their earnings on domestic goods and services. For DCMS' statistical purposes, it is possible to exclude indirect and induced impacts, and just focus on direct ones. Furthermore, through a mix of desk and primary research, Cebr's studies investigate a series of ancillary contributions made by the heritage sector. These include spillover impacts through tourism, regeneration, community wellbeing, and the role of the heritage sector in developing skills, nurturing innovation, and fostering growth in other sectors.

The SIC-SOC mapping approach is specifically designed to estimate the heritage sector's direct contribution in terms of employment. The analysis of heritage employment draws upon official data provided by ONS, which is broken down according to the SIC framework. However, a fair share of the activities in the heritage sector either cross the boundaries of individual SIC codes or are relatively niche and therefore buried within broader categories of economic activity. For this reason, SOC codes are used to map relevant occupations in the heritage sector to corresponding SIC industries, enabling the generation of employment estimates drawing upon ONS data sources broken down by SIC codes only. The list of relevant industries and occupations is based on sector literature review and qualitative research, including stakeholder roundtables and surveys conducted during the pilot stage. This list has subsequently been refined, as SIC and SOC codes have changed over time, largely using desk-based research (Ortus Economic Research, 2017).

Data sources and requirements

The estimation of England's heritage sector's direct employment primarily relies on two key data sources: the APS (discussed above for dynamic mapping) and the Business Register and Employment Survey (BRES). The BRES provides employment information by detailed geography or industry, which collects data from all businesses registered for VAT and/or PAYE in Great Britain, supplemented by employment data independently collected by the Department for Finance and Personnel Northern Ireland (DFPNI).

In addition, as the apportionment of construction-related occupations (e.g., architects, civil engineers, town planning officers) to the heritage sector is based on the proportion of pre-1919 properties to the overall building stock, the methodology utilises the Valuation Office Agency (VOA) dataset 'Council Tax: stock of properties' to derive this figure for each region. The total number of workers in those occupations, estimated through the SIC-SOC mapping

approach, is then multiplied by the share of pre-1919 dwellings in each region, to obtain the total number of heritage workers in those occupations.

Finally, for heritage activities captured by neither SIC nor SOC codes (because no corresponding codes exist), secondary data sources are used to establish estimated heritage employment figures or the employment share that could be applied to the relevant SIC or SOC codes. For example, regional employment figures for archaeology are extracted from Landward Research's series of archaeological labour market intelligence reports.[10]

GVA calculation, which is based on employment estimates, requires additional data from the Annual Survey of Hours & Earnings (ASHE). This dataset provides information about the levels, distribution and make-up of earnings for employees in all industries and occupations.

A full description of how to replicate this method to generate statistics on employment and GVA of the UK heritage sector can be found in Annex 2.

Satellite accounts

Overview

Satellite accounts are a statistical system aiming to describe the economic contribution of specific areas of the economy (such as tourism, arts & culture, or the digital economy) that are not directly presented in the country's core national accounts framework. These accounts provide a more accurate estimate of the relative contribution of these specific areas to a country's economy (such as proportion of total GDP or domestic demand), with these estimates directly comparable across countries as they have been developed through a consistent methodology. In practice, heritage has historically been included as part of cultural satellite accounts (CSAs) due to its classification as a cultural good, though this approach excludes both non-cultural activities and natural heritage from the heritage sector, potentially underestimating its contribution to the national economy (Vanhoutte, 2020).

As part of our review, we carried out an in-depth examination of three different CSAs, including (i) Finland, (ii) Canada, and (iii) the ESPON Heritage Framework, which includes 11 European countries. Among the three different methodologies, we reviewed what type of classification systems each of the accounts used.

[10] The Landward Research reports can be found here.

An overview of each satellite account can be found below in Annex 4, outlining the unit of classification and the type of industries included.

Finland and Canada satellite account definitions encompass culture broadly, including any sector that has a creative or heritage component that produces economic output. The Finnish CSA uses methodologies similar to internationally recognised Tourism and Travel satellite accounts. The Canadian CSA relies on existing statistical components within Canada, including the CSMA and the 2011 Canadian Framework for Culture Statistics.

The ESPON Heritage Framework approach uses a narrower definition of cultural heritage which more closely aligns with how heritage is generally defined within a UK context. ESPON defines material cultural heritage within their framework as “objects of immovable (e.g., archaeological sites, cultural landscapes) and moveable (e.g., paintings, books) nature recognised as having heritage value in each country/region…” (ESPON, 2020).

Data sources and requirements

Satellite accounts generally use country-specific national economic statistics. As part of our in-depth review, we reviewed information on the data used by each of the satellite accounts including details on sources, level of granularity and time coverage; as well as the approach used for assigning an appropriate culture share or apportionment to sectors that have both cultural and non-cultural components. An overview of details on data sources can be found below in Table 1.

Table 1. Data sources across satellite accounts

Satellite account	Data sources used	Key data information	Methods for partial sectors
Finland CSA	• Statistics Finland Business register • Household consumption expenditure data • National accounts data	• No separate data collection — relies on existing data collected within Finland. • Data only available at the national level • Available from 1995 onwards • Updated annually with a 22-month lag	• Appears that industries are treated as fully cultural or not • No specific methods on treating partial contributions of industries
Canada CSA	• Canadian System of Macroeconomic Accounts (CSMA) • Supply and Use Tables • Annual Survey of Service Industries	• Available at the national and provincial/territorial levels • Available from 2004 onwards • CAS is updated every 2 years, and cannot be compared between periods	• All businesses where cultural products make up a significant portion of output are included. • For example, food/drinks at a live entertainment venue would be considered part of culture.
ESPON Heritage Framework	• Accumulated data across 11 regional databases • National measures from Eurostat (European LFS, Structural Business Statistics, Business Demography Statistics) • Other ESPON sources: ESPON Heritage project (2019), ESPON project 2006	• Available at NUTS2 and NUTS3 levels • Exploratory framework for research purposes; not updated regularly	• Developed measurement “keys” as proxy indicators for economic activities to calculate the fraction of a sector related to material cultural heritage. • Examples of keys include proportion of pre-1919 dwellings of total dwellings to measure cultural heritage contribution of architecture activities.

Satellite accounts from Canada and Finland rely on national economic statistics which have limited crossover with data availability for the UK. We therefore focus on the ESPON heritage framework, which encompasses relevant data sources that could be relevant when developing a UK heritage satellite account.

The ESPON Heritage Framework attempts to study the economic value of material cultural heritage specifically focusing on available official statistics and not case studies (ESPON 2020). This review included 11 countries/regions of focus, which do not include the UK. However, this framework draws on other ESPON research and data sources which do cover the UK and include relevant indicators or proxies that have potential for measuring material heritage (ESPON, 2019; ESPON Atlas, 2006). These include (i) density of monuments in regions, at the NUTS3 level, and (ii) the proportion of pre-1919 dwellings at the NUTS3 level.[11] These metrics are part of ESPON's novel proposal for estimating the economic impact of heritage by coming up with a "feasible" definition of the heritage sector and corresponding "keys" to apportion individual industries solely based on availability of existing official statistics (using either data collected at the EU level or by individual national statistics authorities). These keys and the data used are the most relevant to the UK and possible measures for the heritage sector's economic impact.

Big data

Overview

No single definition exists for "big data", a term that has become widely used since the 1990s. While data has long been used to inform decision-making, historically individuals, businesses and governments have been limited by the amount of data they could collect and analyse (both tasks were required to be completed manually). The exponential growth of computing power and the rise of the Internet, however, transformed the volume and speed with which data was generated, opening up new avenues for gaining insights, making better predictions and quantifying metrics that previously relied on more limited data sources (McAfee and Brynjolfsson, 2012). However, most of this data is unstructured (not organised in a database), requiring new tools to be developed to process and analyse this data. In other words, big data differs from regular data in the volume and speed at which the data is generated, the diversity in sources for collecting data and the techniques required to analyse

[11] NUTS3 is part of the EU's common classification of territorial units for statistics and refers to smaller regions (approximately corresponding to unitary authorities, council areas and districts in the UK).

and apply the data. In the public policy sphere, big data has been used for a broad range of applications, ranging from improving the effectiveness of restaurant hygiene inspections (Kang et al., 2013) to developing more accurate and timely measures of economic performance (Cohen et al., 2013).

More specifically to heritage, this category of big data methodologies refers to the use of large online datasets to estimate the size of the heritage sector, using organisation-level data from Companies House, the Inter-Departmental Business Register (maintained by ONS), Dun & Bradstreet (which includes data on unregistered businesses) and Bureau van Dijk (through the FAME dataset). Because companies are required to select a SIC code when registering with Companies House, one approach to counting businesses in a specific sector is looking at relevant five-digit SIC codes, then aggregating the total number of businesses across all relevant codes. This approach has been applied to both the UK digital economy (Nathan and Rosso, 2013; Nathan and Rosso, 2015) and video gaming industry (Mateos-Garcia, Bakhshi and Lenel, 2015). However, the main challenge with this approach is that SIC codes define an organisation's primary business activity (across both inputs and outputs), but many heritage activities are produced by businesses in other sectors (such as a large construction firm taking on several projects involving listed buildings each year). In addition, SIC codes, even at the five-digit level, may be insufficiently detailed, with many businesses classified into "other" or "not elsewhere classified" codes within industry groups.

To address these issues, researchers have started to develop bespoke datasets to identify companies within a sector that do not rely on SIC codes published on Companies House. This involves compiling data across publicly available datasets such as social media, newsfeeds (such as Bloomberg and Reuters), blogs, online forums and company websites. Keywords or key phrases in the online text linked to a company can then be used to link the company to a specific SIC code. Alternatively, researchers have also sought to develop more granular industry classifications outside of the SIC classification system, using a "sector-product" mapping (Nathan and Rosso, 2015), clustering methods (Papagiannidis et al., 2018; Stich, Tranos and Nathan, 2022) or network analysis based on keywords (Marra and Baldassari, 2022). Despite the diversity of different approaches that have been developed in recent years, these do not address the overall incompatibility of big data methodologies with the ONS national accounts framework (and thus DCMS Sector Economic Estimates more generally).

Data sources and requirements

Within the heritage sector, much of the existing research on "big data" approaches has been led by Historic England, which has leveraged machine learning approaches to identify heritage organisations using data from the Charity Commission (for charities) and The Data City (for for-profit companies).[12] These datasets are extracted from public websites through web-scraping, then automatically categorised/clustered and manually checked to develop a list of charities and commercial organisations in the heritage sector. More specifically, the Charity Commission publishes data on the charity's aims and objectives, income and expenditure, a description of what the charity does and a sample section of the charity's governing document. A team of trained researchers can manually review this text data to label the charity as "heritage" or "not heritage", with these positive and negative examples used to train machine learning models to classify all charities in the public register as "heritage" or "not heritage". While The Data City's data collection methods are proprietary, its database is primarily based on Companies House, which has been linked to other data sources such as CreditSafe, Dealroom.co, Lightcast, all text found on companies' own websites and other public webpages. This list of charities and commercial organisations can then be mapped to machine-readable financial data from Companies House and the Charity Commission to produce sector-level estimates such as the number of heritage organisations and total sector turnover.

[12] Report not yet published, project scope and high-level overview described here.

Feedback from stakeholder engagement

Learnings from sector workshop and interviews

Throughout the duration of the project, we engaged with stakeholders to gather insights and perspectives crucial to our research. Our engagement efforts included hosting a workshop with ten participants and conducting nine one-to-one or small group interviews with relevant sector stakeholders. Representatives from the following organisations took part in these discussions:

- Architectural Heritage Fund
- Cardiff University
- Creative Industries Policy & Evidence Centre
- DCMS
- Department for Science, Innovation and Technology (DSIT)
- ESPON
- Heritage Alliance
- Heritage Trust Network
- Historic England
- Historic Houses
- National Lottery Heritage Fund
- ONS
- UNESCO UK
- University College London
- University of Edinburgh

These interactions served as invaluable opportunities to delve into various topics pertinent to our study. The discussions touched on several key thematic areas, reflecting the complexity of defining the heritage sector and measuring its contribution to the economy. These include:

Scope - Definition of heritage

A central theme was the challenge of defining the heritage sector's scope in terms of industries, organisations and activities to be included. Some participants highlighted the need for a broader definition that goes beyond traditional boundaries to include not only physical sites open to the public but also venues with different intended uses, intangible cultural elements, and activities in other industries fuelled by heritage's presence. The trade-

off between expanding the breadth of what is considered heritage and maintaining precision and feasibility in measurement was a recurring concern. A broader scope could capture the full economic impact of heritage but could also risk overestimating the size of the sector and poses substantial challenges in terms of scope of data collection and analysis.

Pre-1919 cutoff for heritage buildings

The discussion around the cutoff year for classifying heritage buildings centred on the need to reconsider the pre-1919 standard, with a suggestion to adopt a pre-World War II cutoff. This reflects an understanding of heritage as a dynamic concept, and a recognition of the historical value of interwar buildings. Participants discussed the implications of such a shift for data collection, policy-making, and heritage conservation efforts, suggesting a need for regular updates to heritage thresholds to reflect changing societal values and historical significance of buildings. It is also worth noting that this measure was chosen due to availability of publicly available datasets, which, in general, is an issue when measuring heritage.

Data sources

The limitations of existing datasets, particularly those comprising registered companies, were noted by some participants as a barrier to accurately assessing the heritage sector's economic contribution. Many heritage activities, especially small sites, charities, and non-VAT registered organisations are not captured in datasets such as the BRES. However, other participants pointed out that the organisations typically excluded by these kinds of datasets are very small entities that do not have a considerable economic impact.

Broader impact of heritage

The significance of intangible aspects of heritage was acknowledged as an important component of the sector's impact. This includes the impact of heritage beyond GVA, direct employment or revenue, such as its influence on well-being, community identity, and the potential for heritage sites to enhance the value of their surrounding areas. Measuring these intangible contributions, however, presents substantial challenges, and would necessitate complex, ad-hoc approaches, such as monitoring frameworks or value-chain approaches. These issues are currently being explored as part of DCMS' Culture and Heritage Capital Programme, including social cost benefit analyses and cultural and heritage capital accounting.[13]

[13] More details on the Culture and Heritage Capital Programme can be found here.

Methodologies to estimate the economic contribution of the sector

Various models for measuring the economic impact of heritage were examined, including SIC-SOC mapping, dynamic mapping and satellite account, each model presenting unique advantages and limitations in capturing the complexity of the heritage sector. Among the models discussed, SIC-SOC mapping emerged as the generally preferred approach due to its ability to delineate the economic activities associated with the heritage sector more clearly, and to account for the contribution of both heritage industries and occupations. Some participants also mentioned alternative approaches, such as incorporating methodologies that allow for the valuation of non-use value of heritage assets (e.g., contingent valuation or choice modelling), also explored as part of DCMS' Culture and Heritage Capital Programme, or identifying a list of heritage organisations through website scraping.

There have been efforts to build better statistics around quantifying the societal and economic impact of cultural heritage, including indicators for immoveable cultural heritage and heritage-related GVA/jobs generated by sectors such as (i) construction, (ii) real estate and property, (iii) tourism, and (iv) cultural and creative industries. However, as mentioned above, there are concerns around overestimating the economic contribution of the sector and how to define what should be considered within the scope of heritage.

Learnings from other sector definition efforts

In addition to discussion with relevant heritage stakeholders, we also conducted in-depth interviews with Government teams trying to develop definitions for other sectors whose economic activity does not align with standard statistical classifications, namely sport and the digital economy. These conversations demonstrated that across sectors, there were similar challenges in coming to a consensus around defining boundaries of a sector and how this translated to using existing economic measures.

Defining the sector

Much like heritage, there have been difficulties defining the boundaries for other types of satellite accounts. For the digital economy, there is no well-established definition, and it overlaps with many standard sectors, including health or finance. This lack of definition for both heritage and the digital economy makes it difficult to map to SIC or SOC codes, and there is a lack of consensus over what should or should not be included. However, for the

sport economy, there is a well-established definition (the Vilnius Sport definition).[14] This is already used internationally across other sport satellite accounts and is directly based on a list of relevant SIC codes.

Identifying partial industries

There are many industries which are expected to fit within orthodox definitions of heritage, for example sectors which encompass architecture or historical artefacts. However, there will be partial industries which contribute to the overarching economic sector, even though they do not fit within the obvious heritage definition. This is relevant even for satellite accounts which have clear boundaries and scope, such as the sport economy. The sport satellite account has attempted to measure these partial industries by first identifying relevant businesses which contribute to the sport economy in some way (e.g., a farm which grows horse-feed for horse racing), and then delivering a survey to identify what percentage sport activities comprise out of their total business (using either output or turnover). How to measure these partial industries is part of ongoing discussions in the heritage and digital economy sector as part of defining the boundaries of these sectors.

[14] The Vilnius Definition of Sport can be found here.

Assessment of methodologies

This section of the report details the strengths, weaknesses and limitations of each of the four methodologies (dynamic mapping, SIC-SOC mapping, satellite accounts and big data) reviewed by the research team. Our assessment of each methodology has been mapped to a RAG (red, amber, green) chart to provide a visual summary of how the methodology performs against each criteria that we considered.

Each of the four methodologies was assessed by our research team against six criteria:

- **Coverage**: the extent to which the breadth of the sector is captured by the approach (a key policy objective of an updated heritage sector methodology)
- **Disaggregation**: enabling disaggregation across different parts of the sector (through producing country- and regional-level estimates)
- **Robustness**: the defensibility of the approach given its methodological steps;
- **Feasibility**: the public availability of required data, or possibility of developing the approach in a proportionate way considering resources required, other costs and complexity of the approach;
- **Replicability**: the potential for publishing the approach at least annually, given that the new method will form the basis of DCMS statistical outputs, including the production of key economic variable estimates for the sector (GVA, total employment, number and size of businesses in the sector and imports/exports of goods and services). Note that replicability both refers to data availability (are underlying data inputs also updated annually?) as well as resources required (is it difficult or time-consuming to collect data annually from all data sources required?)
- **Comparability**: the extent to which the method aligns with UK national accounts methodology published by ONS and the DCMS Sectors Economic Estimates Methodology. Alignment with the Economic Estimates Methodology would allow estimates of heritage sector GVA, for example, to be added to the GVA of other DCMS sectors to produce a single top-level estimate.

Note that all methodologies considered that align with UK national accounts (dynamic mapping, SIC-SOC mapping, satellite accounts) include the activities and contributions of charities and universities through the category “Non-profit institutions serving households”. However, volunteering activities are not currently included in the national accounts

framework, and there have been growing calls to develop a "third sector" or social economy satellite account that can formally capture the impact of all charities (including those that generate income by selling products and services) as well as volunteering activities, both of which are critical components of the heritage sector.[15]

Dynamic mapping

Strengths

The strengths of the dynamic mapping methodology lie in its simplicity and data-driven nature. The formula for calculating the heritage intensity of an industry is straightforward, and defining an industry's heritage intensity is directly a function of the number of heritage-specific occupations that an industry has. Because this methodology abstracts away from proportioning individual industries into heritage and non-heritage components, it is more easily explainable to stakeholders and avoids double-counting between industries.

Weaknesses and limitations

Calculating heritage intensity would be heavily focused on inputs to the industry (individual labour) and does not account for an industry's output. Therefore, while this method worked well for the creative sector, it may be less suitable for measuring heritage economic impact which is less focused on the "production process" of heritage and more focused on the types of products or services that are produced instead. Sports economy and digital economy satellite accounts appear to be more similar to the creative sector, where focus tends to be on production processes rather than outputs. Weaknesses and limitations of this method include the following:

- Current data sources do not include other types of employment essential to the heritage industry, including informal labour like volunteering.
- Labelling an industry by whether they meet a certain intensity threshold creates an "all-or-nothing" system. This will inherently exclude industries that are only partially related to heritage through other means, for example having mostly heritage outputs but fewer heritage occupations, and also potentially overestimate the sector by including non-heritage industries.

[15] For example, see the report "Taking account: the case for establishing a UK social economy satellite account" published by the Law Family Commission on Civil Society, which can be found here.

- Current publicly available APS data does not have secondary jobs at the right granularity of SIC codes, therefore forcing use of the LFS instead for heritage replication which relies on smaller sample sizes.

Assessment

- **Coverage**: Data-driven approach based on "heritage intensity" to determine heritage and non-heritage industries excludes key industries such as construction, food services and accommodation services whose heritage components are large in absolute terms but are small relative to the non-heritage components of these industries.
- **Disaggregation**: Can produce country- and regional- level estimates based on standard econometric approaches to generating regional economic multipliers.
- **Robustness**: The use of "heritage intensity" focuses on employment, which may exclude other dimensions of heritage (for example, communities with a high concentration of historic sites may support a thriving tourism sector even if this sector does not include any occupations recognisable as heritage occupations). In addition, the LFS faces issues with very small sample sizes[16] for heritage occupations, but this can potentially be overcome by transitioning to either (i) the transformed LFS, to be released later in 2024, or (ii) secure-access versions of the APS which include the correct granularity of SIC and SOC codes.
- **Feasibility**: Some data limitations (SIC codes of secondary jobs are not published at a sufficiently granular level), but otherwise estimates can be produced using publicly available datasets. Straightforward approach with clear explainability.
- **Replicability**: Fully replicable, as underlying data inputs are updated at least annually (if not more often). Can produce all desired economic variable estimates for the sector. There may potentially be an issue in establishing the correct "threshold" for heritage intensity, as this is likely to vary from year to year (it would be important to look at the behaviour of the threshold across industries over time, especially with larger industries such as construction or accommodation). Requires the least amount of resources to produce annual dates, due to ease of accessing underlying data sources from ONS and the UK Data Service.

[16] Some of these concerns are addressed by the current DCMS Economic Estimates methodology – for example, the creative intensity dynamic mapping approach uses two years of APS data to avoid unusual survey results in a single year, which could also be applied to heritage.

- **Comparability**: A definition based on SIC codes means the methodology is fully compatible with UK national accounts methodology through supply-use tables published by ONS as well as with the DCMS Economic Estimates. In addition, the methodology can be used to produce separate heritage component sub-sector estimates, and estimates of GVA could be added to and compared against the GVA of other DCMS sectors.

SIC-SOC mapping

Strengths

The SIC-SOC mapping methodology exhibits a number of strengths that render it suitable for assessing the heritage sector's impact on the economy.

- The methodology provides a precise definition of the heritage sector, closely aligned with publicly acknowledged sector 'boundaries'. This alignment with established sector definitions, as observed through university career guides and publications from heritage organisations, ensures accuracy in delineating the sector's scope.
- The methodology effectively captures the contributions of both heritage occupations and industries. Furthermore, it employs a systematic approach to address potential double-counting between these components.
- The methodology is flexible to changes in the scope of the heritage sector. Even if adjustments are necessary to the list of heritage industries and occupations to be included, the core mechanism of the methodology remains fully applicable.

Weaknesses and limitations

One notable limitation of this methodology is its inability to generate trade statistics, or other metrics that rely solely on industrial codes, such as number and size of businesses in the sector.[17] The SIC-SOC mapping enables the attribution of a share of a given industry employment or GVA to the heritage sector. However, this mechanism cannot be directly applied to other statistics. In other terms, the share of an industry attributed to the heritage sector for employment or GVA purposes may not reflect its proportion in trade statistics, also

[17] Trade statistics in general are relatively less important for the heritage sector. In the DCMS Sector Economic Estimates: Trade in Services (2021) publication, almost all trade indicators for "Operation of historical sites and similar visitor attractions" were marked as "None recorded in the survey" or "Suppressed to prevent disclosure of respondent information".

considering that heritage goods and services are often not tradable, and when they are, they are most likely captured by trade figures of other sectors, such as arts, craft, and museums.[18]

Another limitation arises from using an outdated coefficient to estimate the number of heritage workers in the built environment sector (SIC07 codes 433 and 439). This coefficient is based on the 2012 estimated count of individuals involved in heritage building craft skills in England, as reported in the 'Skills Needs Analysis 2013' report published by English Heritage, Historic Scotland, and Construction Industry Training Board.[19] As this report has never been updated, the figures on which this coefficient is built are now over a decade old. Finally, calculating GVA based solely on multiplying employment and earnings does not account for the two other components of GVA in the ONS national accounts framework (taxes less subsidies and gross operating surplus), which may be substantial and lead to an underestimate of the economic impact of the heritage sector.

A number of additional limitations, which are not related to the methodology itself, arise from the public availability of data hindering the full replication of the methodology. In summary:

- The publicly available BRES data are disaggregated either by geography or industry but not both simultaneously. Consequently, sub-national estimates cannot be produced. Note that this can be overcome either through the secure-access version of the BRES dataset or the Inter-Departmental Business Register (IDBR), which the BRES is used to update and includes a comprehensive list of all UK businesses.
- The industrial disaggregation in BRES data extends to the 5-digit SIC code level for Great Britain only. For the whole of the UK, employment data is disaggregated only to the 3-digit SIC code level. To attain the necessary level of disaggregation for UK-level data, an assumption is made that the distribution of employment at the 5-digit SIC level within the respective 3-digit SIC codes in the UK mirrors that of Great Britain.
- The publicly available APS data does not reach the requisite level of disaggregation (5-digit SIC codes and 4-digit SOC codes) at both the industrial and occupational levels. However, this limitation can be addressed through DCMS access to non-public APS

[18] One key challenge in addressing this issue is that most datasets with detailed business-level data such as BRES and the Annual Business Survey (ABS) do not include occupation-level breakdowns of employment (only the total number of employees is reported). Other than developing a new survey of businesses that requests occupation-level data, it is generally not possible to generate a robust set of trade or business demography statistics that is consistent with the apportionment produced by SIC-SOC mapping.

[19] Available here.

microdata.

Assessment

- **Coverage**: This methodology encompasses a broader set of industries in the heritage sector compared to dynamic mapping. This is because it avoids an “all or nothing” approach by using SIC-SOC mapping and strategies to “net out” double-counting to map heritage employment across all industries. The published Cebr methodology currently does not include tourism, as the spillover effects of heritage on the tourism sector are estimated and presented separately.
- **Disaggregation**: Can produce country- and regional- level estimates based on APS data at the regional level for employment statistics, and standard econometric approaches for GVA.
- **Robustness**: Uses the income approach to calculate Gross Value Added (standard methodology consistent with ONS). Similar issues as dynamic mapping with the LFS including data of insufficient granularity (though this can be addressed by secure-access versions of the APS). Assumptions are required around correct apportionment of construction (through the number of people involved in heritage building craft skills) and some building and planning occupations (through the proportion of pre-1919 housing stock).
- **Feasibility**: Some data limitations (some apportionment methods require employment breakdown by 5-digit SIC codes, which is not reported in publicly available versions of the LFS), but otherwise estimates can largely be produced using publicly available datasets. More complex due to (i) the need to apply different apportionment methods to specific industries, and (ii) differences in the level of SIC code granularity requiring additional data aggregation or disaggregation.
- **Replicability**: Fully replicable, as underlying data inputs are updated at least annually (if not more often). However, this methodology cannot be used to produce estimates of imports/exports of heritage goods and services due to challenges around apportionment (for example, the share of heritage GVA relative to total sector GVA is unlikely to be equivalent to the share of heritage goods and services produced by the sector that are exported, due to differing levels of export intensity). For goods specifically, standard commodity codes are in general not sufficiently granular to distinguish between heritage goods and non-heritage goods and a robust mapping would require further research.

Requires a broad range of data sources, though these are generally straightforward to access online and can be merged using SIC codes as a common variable.

- **Comparability**: A definition based on SIC codes (or occupations mapped to SIC codes) means the methodology is fully compatible with UK national accounts methodology through supply-use tables published by ONS. In addition, the methodology can be used to produce separate heritage component sub-sector estimates. Estimates of GVA could potentially be added to the GVA of other DCMS sectors, though adaptations to the methodology would be required to ensure comparability: GVA would need to be calculated using approximate GVA from the ABS (in line with the current DCMS Economic Estimates methodology) instead of earnings from ASHE (see Annex 2 for further details).

Satellite accounts

An overview of the key strengths and weaknesses of all satellite accounts reviewed can be found below.

Strengths

- **Canada CSA:** Codes used are highly detailed and reflect a very broad definition of culture which does include heritage more specifically. Also incorporates a wider range of industries, including those that are supporting cultural sectors without being directly related to culture (for example, food and beverage services at a live music venue). Note that the Canada approach is somewhat similar to the dynamic mapping approach as it is based on identifying individual industries that are “in” or “out” of the cultural sector based on the proportion of cultural output in products/services.
- **Finland CSA:** Uses SIC codes making it easier to compare internationally.
- **ESPON Heritage Framework:** Uses a data-first approach, where all definitions of cultural heritage are tied to indicators where data is available. The initial proposal included some discussion of the UK, providing insight into what data would be available for certain measures of cultural heritage. Data is also available for a sufficiently high level of granularity (NUTS2 and NUTS3), so there is an opportunity for comparisons with the EU more broadly. The ESPON framework also uses keys as measures to approximate the percentage of an industry that could contribute to heritage overall.

Weaknesses and limitations

- **Canada CSA:** Data cannot be compared year over year at the national level but can be at

more granular regional levels (provincial/territorial). Definition does not use SIC codes, which makes comparison to the UK across industries and occupations more challenging.

- **Finland CSA:** Estimates only available at the national level. Does not include cultural components of non-cultural industries (e.g., the graphic designer in a non-creative industry would not be included).
- **ESPON Heritage Framework:** Significant variation across regions, so calculating a single EU measure would have a decent amount of discrepancy. In other words, differences in heritage levels across countries may be due to differences in regional definitions and availability of data (requiring use of proxy variables) compared to actual differences. There is also some disagreement among stakeholders over the use of specific keys as proxy indicators for measuring partial contribution to heritage, including what the best measures are and whether they effectively capture an industry's contribution. Framework also uses NACE codes instead of SIC codes, though the UK SIC is identical to NACE down to and including the 4-digit level.

Assessment

- **Coverage**: The use of administrative data collected by national statistics authorities (or other government departments) provides the most accurate coverage of the heritage sector. Instead of having to consider how to proportion entire SIC codes between heritage and non-heritage components, individual organisations can be reviewed to determine whether they should be included in the heritage sector. Alternatively, supply-use tables might include a sufficiently detailed product-level classification (the approach taken by the United States and Australia, which include around 6,000 and 900 items on detailed commodities, respectively, compared to 100 items in the UK). This means that a sector definition can be fully based on heritage goods and services rather than occupations or industries (which are higher-level and more difficult to include or exclude in full).
- **Disaggregation**: Can produce country- and regional- level estimates based on standard econometric approaches to generating regional economic multipliers.
- **Robustness**: Satellite accounts are widely used by national statistics authorities and this methodology is seen as very robust due to the lack of assumptions required other than limitations around data inputs. In addition, many national statistics authorities have linked industry classifications to products in their supply-use tables, meaning that any definition of the heritage sector can take into consideration the proportion of heritage goods and

services produced as output (which more closely aligns with general definitions of heritage used by the sector).

- **Feasibility**: The Canada and Finland approach is not feasible using publicly available data sources. The ESPON approach is feasible, but appropriate "keys" for the UK need to be identified, and these may not fully align with the ESPON approach due to differences in data collection and publication (which focused on a sample of EU countries). Satellite accounts are in general the most complex of the four methodologies due to the need to (i) ensure that administrative datasets are aligned within a consistent framework for measurement, and (ii) create bespoke supply-use tables to correctly apportion industry (or activity) output across heritage products/services.
- **Replicability**: This approach is replicable, as national statistics authorities often update satellite accounts each year. One potential consideration is the amount of supplementary data collection required – if a satellite account can be produced using routinely-collected administrative datasets, then replication can be done more frequently than if a satellite account requires supplementary data sources (that might not be feasible to update annually). Difficult to fully assess resource requirements but would probably be challenging to update annually as this would be a joint effort by ONS and DCMS.
- **Comparability**: A definition based on SIC codes means the methodology is fully compatible with UK national accounts methodology through supply-use tables published by ONS. In addition, the methodology can be used to produce separate heritage component sub-sector estimates.

Big data

Strengths

"Big data" approaches can be potentially more accurate than SIC-based approaches in estimating the total number of businesses in the heritage sector, as they take a bottom-up approach and consider each organisation for "inclusion" in the heritage sector on a case-by-case basis using data specific to that organisation. A wide variety of publicly available data sources can be drawn on as part of this decision-making process, including information published on the organisation's own website as well as on third-party websites such as news sites, forums or social media.

Weaknesses and limitations

The main challenge with big data approaches is that their primary output is the number of

businesses within a sector and they are not aligned with the UK national accounts methodology. The final list of heritage organisations must first be mapped to SIC codes using data sources such as Companies House before estimates of GVA, employment and trade metrics can be calculated. This step cannot be avoided as SIC codes act as the common identifier between individual businesses and ONS supply-use tables, although an automated framework for mapping businesses to SIC codes would substantially reduce the resources required for big data methodologies. In addition, by focusing on the organisation level, these approaches abstract away from the breakdown of occupations within organisations. Organisations will be included or excluded within the heritage sector using an "all or nothing" approach, which might not be accurate for large businesses (such as construction firms) with both heritage and non-heritage projects (as well as heritage and non-heritage occupations).

Assessment

- **Coverage**: A bottom-up approach better allows each potential organisation to be treated on a case-by-case basis (to determine if it should be considered a heritage organisation or not). This allows many organisations not assigned to traditional heritage SIC codes to be identified as part of the heritage sector. In addition, the use of web-scraping techniques for data collection can include small organisations that do not pay VAT or are not currently registered on Companies House. On the other hand, without further detail on the specific methodology or the training data used for automated classification models, it is difficult to tell if currently-used big-data approaches can capture larger organisations where heritage comprises only a workstream or source of revenue.

- **Disaggregation**: A bottom-up approach allows for country- and regional-level estimates to be produced if the dataset includes geographic characteristics (such as office addresses or geographic scope of projects). This may require additional data collection, and there are potential data quality issues with using registered and operating address data alone (for example, this does not reflect how employees within a single organisation are distributed across offices, or if heritage projects are a focus for all offices or just one or two specific ones).[20]

- **Robustness**: Big data approaches draw on standard, widely-used machine learning and

[20] The Data City has written a more detailed description of the practical challenges encountered when mapping businesses, which can be found here.

quantitative data aggregation techniques (note that methodologies developed by for-profit companies are generally not made publicly available and may involve proprietary data analysis and merging techniques). However, additional work may be needed to understand the reasoning behind classification decisions (through procedures such as estimation of Shapley values or local interpretable model-agnostic explanations). One key challenge with big data approaches is quality assurance, as text from multiple inputs (company name, company description, company aims/objectives, list of products or services provided) must be correctly cleaned and triangulated to ensure meaningful data is passed to predictive models.

- **Feasibility**: All datasets are publicly available. There is a potential trade-off between developing a web-scraping tool in-house (to collect data directly) or partnering with a consultancy (which may already have collected the relevant data but by using a proprietary methodology). Cleaning data and verifying the correct data (without data cleaning errors) has been passed to training models will likely require significant resources.

- **Replicability**: No constraints in terms of frequency of updates (the underlying dataset can be collected and updated as frequently as needed and as resources allow). However, big data approaches can only be used to estimate the number and size of businesses, and mapping to SIC codes is required to produce GVA, employment and trade metrics.

- **Comparability**: Currently not comparable with the UK national accounts methodology and does not align with the DCMS economic estimates unless individual organisations can be mapped to SIC codes. Lack of alignment with SIC codes also means that the methodology cannot be used to produce separate heritage component sub-sector estimates.

Comparative analysis

Table 2 below, presented in Red-Amber-Green format, offers a visual comparative analysis of the four methodologies according to the six key assessment criteria, as discussed in previous sections.

Table 2. Comparison of methodologies based on six key assessment criteria

Criterion	Dynamic mapping	SIC-SOC mapping	Satellite Accounts: ESPON Framework	Big data
Coverage	Red: Excludes key industries whose heritage components are relatively small	Green: Covers a broad set of industries and occupations	Green: Accurate coverage of the heritage sector based on material heritage goods and services	Green: Accurate coverage of the heritage sector based on heritage organisations
Disaggregation	Green: Support 4-digit SIC and SOC codes and regional disaggregation	Green: Support 4-digit SIC and SOC codes and regional disaggregation	Amber: Supports conversion from NACE codes to 4-digit SIC codes and regional disaggregation. Conversion to SOC codes is not available	Red: May support regional disaggregation but no SIC or SOC codes are considered
Robustness	Red: "Heritage intensity" focuses on employment, excluding relevant dimensions of heritage and not including any partiality of sectors (industries are heritage or not)	Amber: Assumptions required for apportioning construction industries (e.g., pre-1919 cutoff)	Green: Widely used by national statistics authorities and no assumptions required	Amber: Widely used ML and data aggregation techniques, but work needed to validate classification decisions
Feasibility	Green: Can be produced with publicly available data, with some minor limitations	Green: Can be produced with publicly available data, with some minor limitations	Red: The ESPON approach may be feasible, but "keys" for the UK must be identified and validated	Amber: Datasets are public, but trade-off on web-scraping development vs. outsourcing
Replicability	Green: Fully replicable, as data inputs are updated annually	Amber: Fully replicable, but cannot be used for trade statistics	Amber: Replicable, but supplementary data collection may be required	Amber: No constraints in frequency, but SIC codes needed for GVA
Comparability	Green: Definition based on SIC codes - fully compatible with UK national accounts methodology	Green: Definition based on SIC codes - fully compatible with UK national accounts methodology	Amber: Definition based on NACE codes (convertible to SIC codes) - fully compatible with UK national accounts methodology	Red: Not comparable with UK national accounts, unless mapped to SIC codes

Next steps on a bespoke methodology

Based on an assessment of the strengths and weaknesses of each of the four methodologies and feedback provided by sector stakeholders, SIC-SOC mapping emerged as the preferred alternative methodology out of the four to replace the baseline DCMS Economic Estimates methodology. This is due to its breadth of coverage (particularly its ability to account for the contribution of both heritage industries and occupations), feasibility and replicability. In this section, we outline a set of potential next steps to develop a bespoke methodology (building on SIC-SOC mapping) to more precisely measure the contribution of the heritage sector to the UK economy. These steps are presented in order of timeframe, starting from steps that could be taken in the short-term to steps that would potentially require additional stakeholder engagement, resources and data to implement. Additional data collection and research would enable eventual development of a heritage satellite account in line with international guidelines for developing satellite accounts such as those developed by the OECD (van de Ven 2021).

Expanding the list of heritage industries and occupations

The list of industries and occupations used by Cebr in the SIC-SOC mapping methodology represents a much broader definition of the heritage sector than the single SIC code currently used by DCMS in its Economic Estimates. This list is based on scoping research to develop Historic England's Heritage Economic Impact Indicators 2017 (Ortus Economic Research, 2017), with the two aims of ensuring (i) the breadth of the sector was captured and (ii) that sufficient data was available to fulfil this definition.

The framework developed by ESPON, which focuses on activities that form part of the material cultural heritage value chain (economic activities dependent on material cultural heritage), includes several industries that are not included by Cebr, such as tourism, information and communications technology (ICT) and insurance.The flexibility of the SIC-SOC mapping methodology means that these industries can directly be "added", using the "keys" proposed by ESPON to apportion these industries between heritage and non-heritage components:[21]

[21] For example, total employment in the tourism sector can be multiplied by 30% to calculate the heritage

- **Tourism**: approximately 30%[22] [23]
- **ICT**: approximately 0.5%
- **Insurance**: approximately 3%

These keys would serve as a substitute for directly estimating the size of the heritage component of each industry through dividing employment in heritage occupations within that industry by total employment, as tourism, ICT and insurance all involve individuals in non-heritage occupations working to support heritage organisations (for example, implementing an online payment system so potential visitors to heritage sites can book tickets or digitising museum collections to make them accessible to a wider audience). ICT is a particularly important industry to include in an expanded definition, as digital technologies have played a growing role in managing heritage infrastructure and sharing heritage with a wider audience (through programmes such as Digital Skills for Heritage, the UK Digital Heritage Symposium and the European digital cultural platform Europeana).[24]

In terms of occupations, there is general consistency between the list of occupations (SOC codes) included in the current SIC-SOC mapping methodology, occupations mentioned in heritage career guides (published by universities and sector organisations) and job vacancy postings on heritage organisation websites. To consider additional occupations that should be included in the definition, one potential approach is to follow the heritage equivalent of the criterion set out by the United States' Bureau of Economic Analysis (BEA) and Statistics Canada to define the arts and cultural occupations: would the occupation cease to exist if heritage assets and activities were removed? (National Endowment for the Arts, 2013). Under this criterion, one set of activities that does not currently appear to be captured is heritage research at higher education institutions (for example, lecturers and researchers at the University of Birmingham's International Centre for Heritage). Data on this set of occupations could potentially be collected through scraping online data of university staff profiles, searching

component of tourism (i.e. the proportion of tourism employment and GVA attributable to heritage). This is a similar calculation to how total employment in the built environment sector is multiplied by the share of people involved in heritage building craft skills in the SIC-SOC mapping methodology.

[22] The Tourism Satellite Account is published by ONS with a two-year lag, which means in general it would not be possible to estimate the contribution of heritage to tourism on the same timeline as producing estimates for other industries via SIC-SOC mapping.

[23] The 30% key for tourism is derived by ESPON from the proportion of leisure tourists to total tourists.

[24] Digital technologies are also widely used in archaeology, as these technologies can be used to conduct archaeological research without disturbing potentially fragile sites as well as more effectively record/archive findings.

for “heritage” and heritage-related keywords, and then estimating the proportion of staff at each university involved in heritage research.

The determination of industries and occupations to be included within the definition of heritage for estimating its contribution to the UK economy may benefit from a dedicated public consultation. The purpose of this consultation could be extended to capture input from the public and heritage stakeholders on the use of the SIC-SOC mapping methodology to produce estimates of employment and GVA, as well as on the proposed data sources, and the feasibility of incorporating additional statistics, such as trade figures and number and sizes of businesses in the sector. An indicative list of potential questions for inclusion in such a public consultation can be found in Annex 3.

Collecting additional data on heritage

In addition to broadening the scope of heritage industries and occupations, another step that could be taken is improving assumptions related to how individual industries are split into heritage and non-heritage components (the “keys” developed by the ESPON framework). These include:

- **Revising the proportion of architects, town planning officers, chartered surveyors, and building/civil engineering technicians working in heritage:** The current apportionment method used for the SIC-SOC mapping methodology is based on the proportion of pre-1919 housing stock, as 1919 is the threshold for “historic buildings” used by Historic Environment Scotland, Historic England and Cadw (Cadw and Historic England, 2019). Changing this threshold to 1945 could allow for interwar heritage to be captured (for example, Art Deco Underground stations on the Piccadilly Line in London or modernist social housing estates such as St Andrew’s Gardens in Liverpool). This change would significantly increase the number of buildings classified as heritage. For example, according to the Valuation Office Agency’s 2023 data, moving the cutoff year from 1919 to 1945 would raise the proportion of heritage properties in England and Wales's total building stock from 21% to 36%, proportionally increasing the share of construction professionals to be considered heritage workers.[25] Similarly, in Scotland and Northern Ireland, such adjustments would shift the percentages of heritage buildings from 19% to 30% and from 11% to 19% respectively, according to the 2021 Scottish

[25] Valuation Office Agency data can be found here.

House Condition Survey and 2016 data from the Northern Ireland Housing Executive.[26] On the other hand, changing the threshold to 1945 might conflate the threshold used to define heritage and the threshold used to apportion heritage-related occupations related to building (these two concepts are different and should not necessarily be applied to the same year). As Historic England have previously pointed out, 1919 represents a key transition year in shifting from widespread use of traditional building techniques to more modern methods.[27] Alternative apportionment methods could be the density of scheduled monuments, density of listed buildings (which is set by Historic England and based on more holistic criteria beyond building age) or the Heritage Index developed by the Royal Society of Arts and the National Lottery Heritage Fund (which is based on "use" of heritage assets by combining over 100 indicators, including the number of young people who are active in heritage, Blue Plaques, Heritage Open Days and land designated for protection of wildlife).[28]

- **Collecting firm-specific data on heritage activities:** Instead of using top-down approaches to estimating the heritage component of industries, a potentially more accurate bottom-up methodology would draw on more granular data sources. This bottom-up methodology would then inform apportionment methods for specific industries in the SIC-SOC mapping methodology. For example, a bottom-up methodology could include:
  - Examination of annual financial statements of heritage-related businesses filed with Companies House and financial data from Bureau van Dijk's FAME dataset
  - A bespoke survey of heritage umbrella organisations in the third sector (to understand the types of projects they fund or support)
  - A monitoring framework for grants provided to heritage organisations (to understand how grants are spent on individual suppliers/activities and the broader supply-chain impacts of spending by different types of heritage organisations)
  - A bespoke survey of heritage/heritage-adjacent businesses (directly asking about the percentage of employment or expenditures on heritage activities)

[26] The 2021 Scottish House Condition Survey can be found here. Data from the Northern Ireland Housing Executive can be found here.

[27] See footnote 1 here.

[28] One potential concern with using density of listed buildings as an apportionment method is that it would be biased towards London and the South East.

- Analysis of job vacancy descriptions (using centralised databases such as Lightcast that can track the relative breakdown of heritage vs. non-heritage job vacancies within individual companies over time)

By combining these data sources, a more robust picture of the heritage sector could be developed that captures the heterogeneity of supply chain impacts across heritage organisations of different sizes, sectors and activities.

- **Collecting data on "consumption" of heritage outputs:** This data could help apportion tourism-related industries such as accommodation and food services. Potential data sources include surveys of consumer spending on heritage sites/attractions as well as the GB Tourism Survey/GB Day Visits Survey (which identifies the proportion of tourists in the UK visiting heritage sites).[29] A more robust alternative would be surveys conducted by individual heritage sites (to understand typical visitor journeys and separate out the impact of the heritage site as a "pull factor" in attracting visitors compared to other nearby visitor attractions).

Developing a heritage satellite account

As discussed previously, while SIC-SOC mapping represents the best available methodology (out of the four methodologies reviewed) in the short-term to capture the breadth of the heritage sector's economic contributions, a heritage satellite account approach would be the most robust methodology in the long-term and would provide the most accurate coverage of the heritage sector in the UK. A heritage satellite account is meant to spotlight heritage-related production and spending already present in ONS supply-use tables (in other words, to re-arrange current data published as part of the UK national accounts to isolate heritage production and spending). However, given the challenges and complexities around collecting the required data, developing such an account would likely not be feasible in the short-term. To help illustrate the process, we have outlined a set of three key steps that would be required to develop a heritage satellite account:

[29] For example, data from the GB Day Visits Survey can be filtered by main activity of the trip, with one option including "went to a visitor attraction". In addition, Visit England conducts the Annual Survey of Visits to Visitor Attractions, which contains year-on-year attendance data that can be filtered by specific heritage sites.

Step 1: Establishing a definition of heritage that captures both a conventional and broad conception of heritage

This would require input from heritage experts across academia, government, businesses, charities and potentially the public. For example, a dedicated public consultation could be conducted by the UK Government (similar to DCMS' 2013 consultation on creative occupations). Note that any definition would need to focus on industries and occupations rather than products, due to the lack of sufficient granularity in the product-level classifications of the ONS national accounts framework. This definition should be informed by a detailed review of firm-specific data on heritage activities (described above). The final definition agreed on should capture both industries and occupations that are fully part of the heritage sector or partially engage in heritage activities.

Step 2: Estimating the proportion of individual industries specific to heritage

Note that this step is also a key part of the SIC-SOC mapping methodology. While SIC-SOC mapping uses an employment-based apportionment method, satellite accounts typically use an output- or expenditure-based apportionment method, drawing on a broad range of data sources on expenditures on economic activities including administrative data collected by national governments, private industry reports and proprietary business data. This would be used to estimate the share of each industry's output specific to heritage, which could then be applied to employment and compensation as well. However, the share of heritage output would ideally be estimated separately from the share of heritage employment, with data provided to either confirm the assumption that the proportions are the same or to replace the assumption with better estimates.

Step 3: Compiling GVA and employment aggregates for heritage industries

The final step would be to apply the heritage shares of industries to the ONS supply-use tables, in line with standard satellite account methodologies in other sectors. This process could either follow an income-based approach used for SIC-SOC mapping (multiplying heritage employees in an industry by average industry compensation to estimate gross value added), or a production-based approach used for satellite accounts in sectors such as sport and tourism (calculating final expenditures based on export, output, changes in inventories and intermediate consumption). The economic contribution of the heritage sector to the UK economy would then be calculated from supply-use tables based on standard input-output multiplier analysis.

Conclusion

This report considered four different approaches to produce a single reliable estimate of the economic contribution of heritage organisations to the UK economy as part of a broader feasibility study. Each methodology was assessed on six criteria (coverage, disaggregation, robustness, feasibility, replicability, and comparability), set against the baseline methodology currently used by DCMS in its Economic Estimates, the single SIC code 91.03 *Operation of historical sites and buildings and similar visitor attractions.*

All four approaches included three steps: identifying relevant businesses engaging in heritage activities, estimating the proportion of these businesses related to heritage, then applying these proportions to the ONS national accounts framework. However, our review uncovered key differences, including:

- The dynamic mapping approach, based on calculating the heritage employment intensity of industries, is easy to understand and straightforward to apply, but potentially excludes many industries that engage in heritage activities such as tourism and construction.
- The SIC-SOC approach covers both heritage industries and occupations, but cannot be used to estimate trade statistics or the number of heritage businesses.
- Developing a heritage satellite account would potentially produce the most accurate estimate of the economic impact of the heritage sector, aligning with the ONS national accounts framework while also allowing for direct comparison with other countries, but would be the most difficult approach to implement due to data requirements.
- Big data approaches support a bottom-up approach and allow for the heritage activities of individual businesses to be captured. However, collecting online data may be resource-intensive, quality assurance of this data is not straightforward and the approach cannot be used to produce GVA or employment estimates without the use of Companies House data.

Based on these findings, SIC-SOC mapping emerged as the best of the four approaches in the short-term to provide economic estimates of the heritage sector due to its coverage and feasibility. Importantly, the SIC-SOC mapping approach also provides a flexible template to develop more precise estimates over time (by expanding the list of heritage industries and occupations and incorporating additional data on heritage activities and outputs), allowing for an eventual transition to a fully developed heritage satellite account, the approach that most robustly captures the full set of linkages the heritage sector has with the broader UK economy.

References

Bakhshi, H., Freeman, A., & Higgs, P. (2013). A dynamic mapping of the UK's creative industries. Accessed here.

Cadw and Historic England. (2019). Teaching Resource: Understanding Traditional (pre-1919) and Historic Buildings for Construction and Built Environment Courses. Accessed here.

Cebr. (2018). The heritage sector in England and its impact on the economy: a report for Historic England. Accessed here.

Cebr. (2023). The heritage sector in England and its impact on the economy: An updated report for Historic England. Accessed here.

Cohen, S. N., Lui, S., Malpass, W., Mantoan, G., Nesheim, L., de Paula, Á., Reeves, A., Scott, C., Small, E., & Yang, L. (2023). Nowcasting with signature methods (arXiv:2305.10256). arXiv. Accessed here.

ESPON Atlas. (2006). Mapping the Structure of the European Territory. Accessed here.

ESPON. (2020). Measuring economic impact of cultural heritage at territorial level. Accessed here.

ESPON. (2019). The Material Cultural Heritage as a Strategic Territorial Development Resource: Mapping Impacts through a Set of Common European Socio-economic Indicators. Accessed here.

Kang, J. S., Kuznetsova, P., Choi, Y., & Luca, M. (2013). Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews (SSRN Scholarly Paper 2293165). Accessed here.

Marra, A. & Baldassari, C., 2022. Using text data instead of SIC codes to tag innovative firms and classify industrial activities. PLOS ONE 17, e0270041. Accessed here.

Mateos-Garcia, M., Bakhshi, H., & Lenel, M. (2013). A map of the UK Games Industry. Accessed here.

McAfee, A., & Brynjolfsson, E. (2012, October). Big Data: The Management Revolution. *Harvard Business Review.* Accessed here.

Nathan, M., & Rosso, A. (2013). Measuring the UK's Digital Economy with Big Data. Accessed here.

Nathan, M., & Rosso, A. (2015). Mapping digital businesses with big data: Some early findings from the UK. Research Policy, 44(9), 1714–1733. Accessed here.

National Endowment for the Arts. (2013). NEA Guide to the U.S. Arts and Cultural Production Satellite Account. Accessed here.

Ortus Economic Research. (2017). Heritage Economic Impact Indicators 2017: Technical report. Accessed here.

Papagiannidis, S., See-To, E. W. K., Assimakopoulos, D. G., & Yang, Y. (2018). Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age? Computers & Operations Research, 98, 355–366. Accessed here.

Statistics Canada. (2024). Canadian Framework for Culture Statistics. Accessed here.

Statistics Finland. (2023). Culture Satellite Accounts. Accessed here.

Stich, C., Tranos, E., & Nathan, M. (2023). Modeling clusters from the ground up: A web data approach. Environment and Planning B: Urban Analytics and City Science, 50(1), 244–267. Accessed here.

van de Ven, P. (2021), Developing thematic satellite accounts: The example of a thematic satellite account for transport, OECD Statistics Working Papers, No. 2021/02, OECD Publishing, Paris. Accessed here.

Vanhoutte, C. (2020). The Flanders Heritage Satellite Account: A Strategic Instrument for Policy. Österreichische Zeitschrift für Kunst und Denkmalpflege, 74, 12-19. Accessed here.

Annex 1: Dynamic mapping

Application to the heritage sector

Replicating dynamic mapping for the heritage sector involves choosing the appropriate occupations to classify as heritage. These occupations come from several sources, including (i) the Cebr SIC-SOC mapping methodology, (ii) Alma's own review of SIC and SOC codes, (iii) other sources from the literature, including a 2019 Historic England consultation response[30] on new SOC codes, and (iv) SOC codes from SIC 91 (Libraries, museums, archives, and other cultural activities) which passed a 10% threshold. [31] The final list of included heritage SOC codes can be seen below in Table 3.

Table 3. Heritage SOC for primary occupations

Heritage Occupations Group	SOC
Archivists and Curators	2452
Conservation professionals	2141
Conservation and environmental associate professionals	3550
Gardeners and landscape gardeners	5113
Architects	2431
Town Planning Officers	2432
Chartered Surveyors	2434
Building and civil engineer surveyors	3114
Sports and leisure assistants	6211
Librarians	2451
Library clerks and assistants	4135
Engineering professionals not elsewhere classified	2129
Social and humanities scientists	2114
Engineering technicians	3113
Architectural and town planning technicians	3121
Civil engineers	2121
Steel erectors	5311
Bricklayers and masons	5312

[30] This is taken from Historic England's response to the ONS 2019 SOC user engagement exercise, which can be found here.

[31] This means that of the total number of a specific SOC code, at least 10% of the occupations were concentrated within SIC 91.

Heritage Occupations Group	SOC
Roofers, roof tilers and slaters	5313
Plumbers and heating and ventilating engineers	5314
Carpenters and joiners	5315
Glaziers, window fabricators and fitters	5316
Painters and decorators	5323
Construction and building trades supervisors	5330
Quantity surveyors	2433
Chartered architectural technologists	2435
Construction project managers and related professionals	2436
Quality control and planning engineers	2461
Floorers and wall tilers	5322
Moulders, core makers and die casters	5212
Sheet metal workers	5213
Pipe fitters	5216
Tool makers, tool fitters and markers-out	5222
Construction operatives not elsewhere classified	8149
Scaffolders, stagers and riggers	8141
Construction and building trades not elsewhere classified	5319

With heritage occupations, we can then calculate the heritage intensity for each industry using the same formula for creative industries but modified for the heritage sector. This modified formula is:

$$heritage\_intensity_i = \frac{total\_heritage\_occupations_i}{total\_jobs_i}$$

Key steps to apply dynamic mapping and calculate heritage intensity are summarised below.

1. The first step is to narrow down the scope of employment. Only jobs where individuals are either employed or self-employed should be included, as well as primary and secondary jobs.
2. With the list of SOC codes from Table 3, the total number of heritage occupations is then calculated for each industry. This is done with survey weights applied, so the weights of each heritage SOC code are summed by each industry code (SIC). This step calculates the numerator of the heritage intensity formula.

3. For each industry, the total number of jobs (both heritage and non-heritage) is calculated with weights applied. Weights for each SIC code are summed to calculate the total, which is the denominator of the intensity formula. This step calculates the denominator of the heritage intensity formula.
4. Divide the output of Step 2 by the outputs of Step 3 to calculate the heritage intensity for each industry (SIC).

Dynamic mapping replication results

Applying the above steps based on LFS data produces the following heritage intensities, as shown in Table 4. These are the only industries with heritage intensities above 50%. The highest intensity is SIC code 4391 (Roofing activities) with an intensity of 80.2%. Most of the industries with a heritage intensity above 50% appear to be centred around construction activities, suggesting that most occupations focus on repairing and maintaining heritage buildings.

Table 4. Heritage Intensities (above 50%) by SIC code (primary job only)[32]

Industry Name	SIC07 Code	Heritage Intensity
Roofing activities	4391	0.802
Landscape service activities	8130	0.754
Other building completion and finishing	4339	0.752
Painting and glazing	4334	0.751
Joinery installation	4332	0.734
Operation of historical sites and similar visitor attractions	9103	0.579
Other specialised construction activities	4399	0.566
Sales and retails assistants	7111	0.556
Plumbing, heat and air-conditioning installation	4322	0.543
Floor and wall covering	4333	0.542
Sewerage	3700	0.505

[32] Only primary jobs were included due to inconsistencies in the publicly available LFS data that the research team had access to. Secondary jobs should in general be included as part of the dynamic mapping methodology, however the SOC codes for secondary jobs in the publicly available version of the LFS dataset were not sufficiently granular (and thus could not be directly added to primary jobs).

Below, Figure 1 shows the distribution of heritage occupations across industries, partitioned by whether an industry is considered fully within the heritage sector by Cebr (Cebr, 2018). For example, museums are considered fully within the heritage sector based on common definitions of heritage, while construction is only considered partially within the heritage sector (as not all construction projects involve heritage assets).

Figure 1. Distribution of heritage occupied jobs by industry by heritage and non-heritage SIC codes

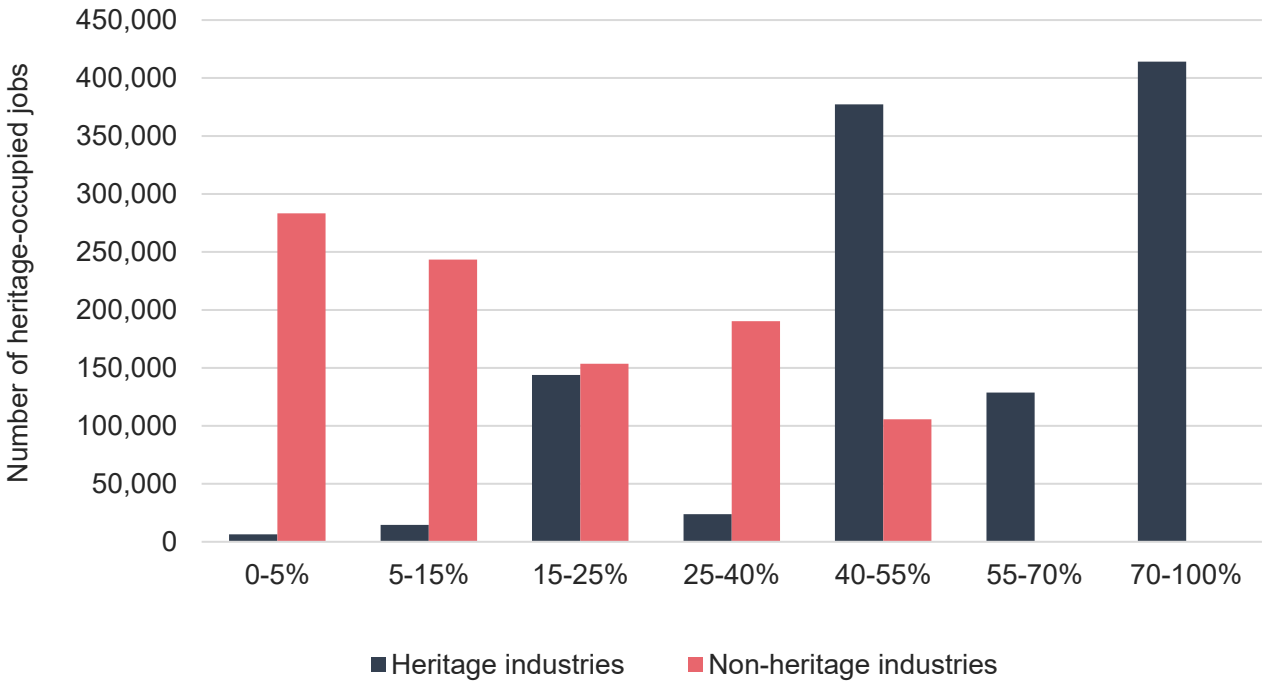


Figure 1 shows that many industries defined as heritage by Cebr have an increasing number of heritage occupations as heritage intensity increases (Cebr, 2018). This suggests that the occupations defined as heritage comprise a crucial component for calculating heritage intensity. These occupations should be selected to ensure that the resulting heritage industries assigned as outputs accurately reflect the heritage sector's activities.

While an occupation may fully be heritage in nature, it may only comprise a small number of occupations in the overarching industry. This is demonstrated in Table 5, which focused on the heritage intensities for each heritage industry defined by Cebr (Cebr, 2018). Most of these industries would also be classified as heritage industries under dynamic mapping due to their intensity threshold surpassing 50% (as seen in Table 5). However, there are some industries including 9103 (museum activities) which may have been expected to have a higher intensity.

Table 5. Heritage intensities – SIC codes from SIC-SOC Mapping

Industry Name	SIC07 Code	Heritage Intensity
Roofing activities	4391	0.802
Painting and glazing	4334	0.751
Joinery installation	4332	0.734
Operation of historical sites and similar visitor attractions	9103	0.579
Other specialised construction activities	4399	0.565
Floor and wall covering	4333	0.542
Museum Activities	9102	0.305
Plastering	4331	0.265
Library and archive activities	9101	0.235
Botanical and zoological gardens and nature reserve activities	9104	0.080

Annex 2: SIC-SOC mapping

Application to the heritage sector

Estimating employment

The estimation of total employment within the heritage sector involves aggregating estimated employment figures across specific industries and occupations. This process includes different approaches tailored to different groups of industries and occupations. A summary of the key steps is outlined below:

1. For constituent heritage industries fully captured by relevant SIC codes, employment figures are obtained directly from the BRES.[33] These industries include:
    a. Museum activities – SIC07 code 9102
    b. Operation of historical sites and buildings and similar visitor attractions – SIC07 code 9103
    c. Archive activities – SIC07 code 91012
2. To estimate the number of heritage workers in building and completion finishing (SIC07 code 433) and other specialised construction activities (SIC07 code 439), a coefficient is applied to the total employment in these industries obtained from the BRES. This coefficient was calculated as the number of individuals engaged in heritage building craft skills in England – found in the report 'Skills Needs Analysis 2013' – divided by total employment in the built environment sector (SIC07 codes 433 and 439) in 2012. The same coefficient is then used for all years the estimates are produced.
3. For heritage activities where occupations rather than industries are well defined, employment estimates are derived by integrating data from the APS and BRES. The SIC-SOC mapping, generated using APS data, enables the estimation, on a representative sample of UK workers, of the weight of specific occupations within the industries where they exist. This is determined as the percentage of workers within those industries that are engaged in heritage occupations. By applying these weights to industrial-level employment data from the BRES, total employment for specific

[33] Employment estimates could also be obtained from the APS, though this may produce slightly different estimates as the APS classifies jobs on a residence basis while the BRES classifies jobs on a workplace basis.

occupations can be estimated. Heritage occupations fully captured by relevant SOC codes include:

a. Archivists and curators – SOC10 code 2452
b. Conservation professionals – SOC10 code 2141
c. Conservation and environmental associate professionals – SOC10 code 3550
d. Gardeners and landscape gardeners - SOC10 code 5113[34]

Moreover, the SIC-SOC mapping prevents double counting of workers in heritage occupations within heritage industries. This means that if the share of workers in these occupations are employed in any of the heritage industries listed above, their employment is not counted again in the estimation process for these occupations.

4. The same methodology is employed to estimate the number of workers in construction-related occupations linked to heritage through the age of the buildings they are associated with. These occupations include:
    a. Architects – SOC10 code 2431
    b. Town planning officers – SOC10 code 2432
    c. Chartered surveyors – SOC10 code 2434
    d. Building and civil engineering technicians – SOC10 code 3114

    To estimate the number of heritage workers in these occupations, an additional step is needed. This involves multiplying the estimated total employment in each occupation by the proportion of pre-1919 building stock in each region.

5. Finally, to account for archaeologists in the calculation of total employment within the heritage sectors – given the absence of SOC or SIC codes identifying this profession – their employment figures are directly sourced from the periodic Landward Research reports 'Archaeology Labour Market Intelligence'.

These steps, taken together, produce employment for all heritage component sub-sectors: each sub-sector has a specific apportionment method to determine the proportion of heritage employment in that sub-sector, based on the principles set out in steps 1 through 6. Total employment in the heritage sector can be calculated by summing up the employment for each heritage component sub-sector.

[34] Only gardeners and landscape gardeners involved in holiday centres and villages (SIC07 code 55201) and general public admin activities (SIC07 code 8411) are included in the estimation.

Estimating GVA

Calculation of GVA for the heritage sector is based on estimated employment figures following the income approach for calculating regional GVA set out by ONS, which adds up income generated by individuals in the production of goods and services.[35] It involves the following steps:

1. Data from ASHE is used to obtain median earnings for relevant heritage sub-sectors and occupations. Given the insufficient sample sizes at the regional level, estimates for the UK as a whole are used; these are weighted by region according to the ratio of total median earnings per region and the overall national average.
2. Median earnings for relevant heritage sub-sectors and occupations are then multiplied by the corresponding employment numbers. Because ASHE includes earnings data for both SIC and SOC codes, this step produces estimates of weighted earnings for individual heritage sub-sectors as well as weighted total earnings for the entire heritage sector in each region.
3. Total weighted earnings of the heritage sector in each region are divided by the total weighted earnings of all industries within that region. This yields the share of total earnings attributed to the heritage sector.
4. This share is then applied to the ONS regional GVA estimate to estimate the total GVA for the heritage sector in each region.
5. Step 3 can be replicated to calculate the share of total earnings for individual heritage sub-sectors (out of all industries within that region). This share can be multiplied by the ONS regional GVA estimate (as in step 4) to estimate the total regional GVA for each heritage sub-sector. Aggregating across all UK regions will produce estimates of the total GVA of each sub-sector.

One key limitation of the earnings-based approach using ASHE (described in the steps above) is that it does not fully cover all components of GVA. Because GVA refers to the value of an industry's outputs minus the value of intermediate inputs, it can be calculated by summing compensation of employees (such as wages and earnings), taxes (net of subsidies), and gross operating surplus and mixed income (such as profit). While previous

[35] The ONS has published a methodology document on calculating regional gross value added (income approach).

ONS publications indicate that compensation of employees comprise the majority of GVA[36], it is possible that the SIC-SOC mapping methodology following the earnings approach understates the total GVA of the heritage sector (as it focuses on one component of employee compensation). Further research is required to understand the magnitude and significance of this discrepancy.

Alternatively, the SIC-SOC mapping methodology could be adapted to the current DCMS Economic Estimates methodology based on approximate GVA from the ABS. Because SIC-SOC mapping splits individual SIC codes into heritage and non-heritage components, adapting this approach would involve multiplying approximate GVA for each 4-digit SIC code by heritage employment shares. The remainder of the DCMS Economic Estimates methodology (summing approximate GVA at the division level and applying the industry share relative to total division approximate GVA to the ONS supply-use tables) would proceed as with other DCMS sectors. Compared to the earnings approach, the approximate GVA approach better captures all three components of GVA. However, DCMS has previously reported a negative approximate GVA for SIC 91 in ABS data in its GVA technical and quality assurance report (requiring the substitution of sales data to overcome).[37] As this may affect other heritage SIC codes as well, further research would be required to establish the validity of the approximate GVA approach (based on the quality of ABS data at sufficiently granular industry breakdowns).

[36] For example, this data can be found in the ONS UK Regional Accounts.

[37] This data is taken from the DCMS Economic Estimates 2019, which can be found here.

Annex 3: Questions for consultation on SIC-SOC mapping and heritage definition

The following list of questions serves as a preliminary framework for a potential public consultation concerning the use of the SIC-SOC mapping methodology and the determination of industries and occupations that define the heritage sector. The consultation would outline the rationale behind the need for a revised approach in estimating the contribution of the heritage sector to the UK economy. Subsequently, it would introduce the use of the SIC-SOC mapping methodology, providing details on its procedural steps and the current list of industries and occupations in scope. A series of questions would then be asked to solicit feedback from sector experts on critical aspects of this proposed approach, as detailed below.

Question 1 – *What are your thoughts on the proposed use of the SIC-SOC mapping methodology to generate DCMS Heritage Sector statistics?*

Question 2 [re: Museum activities (SIC07 code 9102), Operation of historical sites and buildings and similar visitor attractions (SIC07 code 9103), and Archive activities (SIC07 code 91012)] – *What are your views of the list of industries to be considered in full as heritage industries? Are there industries which have not been included which you think should be, in full or in part? Are there industries which have been included which you think should not be? What evidence do you have (if any) to support your view on inclusions or exclusions? Please note: we can only consider industries if they are stated in terms of the Standard Industrial Classification system.*

Question 3 [re: Building and completion finishing (SIC07 code 433), Other specialised construction activities (SIC07 code 439) and the approach to estimate the share of heritage workers within these industries (coefficient in the report 'Skills Needs Analysis 2013')] – *What are your views of the list of construction-related industries to be considered in part as heritage industries and the approach to estimate the share of heritage workers within these industries? Are there industries which have not been included which you think should be? Are there industries which have been included which you think should not be? What evidence do you have (if any) to support your view on inclusions or exclusions? Could you suggest a better or alternative approach to estimate the share of heritage workers within these industries? Please note: we can only consider industries if they are stated in terms of the Standard Industrial Classification system.*

Question 4 [re: Archivists and curators (SOC10 code 2452), Conservation professionals (SOC10 code 2141), Conservation and environmental associate professionals (SOC10 code 3550), and Gardeners and landscape gardeners (SOC10 code 5113) in holiday centres and villages (SIC07 code 55201) and general public admin activities (SIC07 code 8411)] – *What are your views of the list of occupations to be considered in full as heritage occupations? Are there occupations which have not been included which you think should be, in full or in part? Are there occupations which have been included which you think should not be? What evidence do you have (if any) to support your view on inclusions or exclusions? Please note: we can only consider occupations if they are stated in terms of the Standard Occupational Classification system.*

Question 5 [re: Architects (SOC10 code 2431), Town planning officers (SOC10 code 2432), Chartered surveyors (SOC10 code 2434), Building and civil engineering technicians (SOC10 code 3114) and the approach to estimate the share of heritage workers within these industries (pre-1919 cutoff)] – *What are your views of the list of construction-related occupations to be partly considered as heritage occupations and the approach to estimate the share of heritage workers within these occupations? Are there occupations which have not been included which you think should be? Are there occupations which have been included which you think should not be? What evidence do you have (if any) to support your view on inclusions or exclusions? Could you suggest a better or alternative approach to estimate the share of heritage workers within these occupations? Please note: we can only consider occupations if they are stated in terms of the Standard Occupational Classification system.*

Question 6 – *What are your thoughts on the proposed methodology to estimate the GVA for the heritage sector? Do you think there is a better or alternative approach that aligns with the SIC-SOC mapping methodology?*

Question 7 – *What are your thoughts on the data sources employed by the methodology? Do you have any suggestions for different or additional data sources that could potentially substitute or complement those included?*

Question 8 – *Do the SIC and SOC codes adequately and accurately capture the full range of economic activity within the heritage sector? If not, how would you better define a feasible and replicable approach to reach this goal?*

Question 9 – *What are your thoughts on extending the use of this methodology to generate additional statistics, in particular trade statistics and the number and size of businesses in the sector?*

Annex 4: Overview of cultural satellite accounts

Satellite account	Unit of classification	Industries included
Finland CSA	SIC codes No SOC Codes	Architectural and industrial design Motion pictures, videos and computer games Amusement parks, games and other entertainment and recreation Libraries, archives, and museums, etc. Production and distribution of books Education and cultural administration Organisation of cultural events and related activity Advertising Printing and related activities Radio and television Newspapers, periodicals, and news agencies Manufacture and sale of musical instruments Art and antique shops Artistic, theatre and concert activities Photography Manufacture and sale of entertainment electronics Sound recordings
Canada CSA	Input-Output Classification (IOIC) codes North American Industry Classification System (NAICS) No SOC codes	Visual and applied arts Audio-visual and interactive media Written and published works Sound recordings Heritage and libraries (including archives, culture and natural heritage) Multi-domain (e.g. Information services) Education and training Sport Live performance Governance funding and professional support
ESPON Heritage Framework	NACE codes No SOC codes	Archaeology Architecture Museums, galleries, and libraries Tourism Construction Information and communication technology (ICT) Insurance Real estate

+44 20 8133 3192 43 Tanner Street, SE1 3PL, London, UK
+30 21 2104 7902 Ifigenias 9, 14231, Athens, GR

LinkedIn company/alma-economics
almaeconomics

alma economics