

## **Summary**

This dataset provides the first fully comprehensive boundary summary for English [conservation areas](#). It contains public sector information licensed under the [Open Government Licence v. 3.0](#) (OGL).

It has been compiled from the [Historic England](#) INSPIRE dataset (covering around 75% of conservation areas, CAs), supplemented by direct contributions from the missing Local Planning Authorities (LPAs) and, in the case of Chesterfield (for which no adequate quality dataset was available) and South Bucks (where Buckinghamshire Council has so far withheld the data), a specifically prepared good quality substitute. The boundary data has been geocoded over background maps made available by the Ordnance Survey (OS) as [open data](#) or under the [Public Sector Geospatial Agreement](#) (PSGA).

The consolidated data has then been subjected to a quality review aimed at standardising names and unique identifier (UID) codes, correcting LPA allocations, correcting invalid geometries, simplifying multi-feature CAs, eliminating significant overlaps, and disaggregating CAs that materially overlap LPA boundaries. The resultant dataset retains numerous small overlaps, largely arising from geocoding inconsistencies, as well as several CAs with smaller components still spanning LPA boundaries, but should nevertheless represent the best available summary of the spatial extent of conservation areas across England. It is hoped that future releases may further reduce these residual anomalies.

## **Version control**

Version 1.0 (raw) represents the boundary data as provided by Historic England (HE), the Greater London Authority (GLA) and individual LPAs between October 2020 and February 2021, subject to only a few attribute changes such as duplicate UID corrections and first phase name standardisation.

Version 2.0 represents the first quality-enhanced dataset released in February 2021, including the initial changes to boundaries and attributes described in the data quality review section below.

## **Document contents**

- 2 Data attribute structure
- 3 Project motivation, licensing considerations
- 4 Data consolidation process
- 5 Data quality review process
- 8 Dataset maintenance, acknowledgements

## **Contact**

Any questions or comments regarding this dataset, or contributions to enhance and maintain its data quality and comprehensiveness, should be addressed to Ian Hall ([ian@bedfordpark.net](mailto:ian@bedfordpark.net)).

## Data attribute structure

<b>Attribute</b>	<b>Type</b>	<b>Description</b>
NAME	String	Standardised unique CA name
DESIG_DATE	String	Designation and expansion dates, where available
UID	Integer	Unique identifier, mirroring the HE Conservation Areas at Risk Survey UID for the HE dataset and assigned separately for newly added areas
LPA	String	Local Planning Authority ( <a href="#">ONS boundaries</a> as of October 2020)
NPA	String	Indicator for CAs managed by National Park Authorities (NPA) or London Development Corporations (DC)
LA_2019	String	Principal Local Authority area ( <a href="#">OS boundaries</a> as of December 2019)
LA_2017	String	Earlier principal Local Authority area ( <a href="#">OS boundaries</a> as of February 2017)
AREA	Integer	Area in square metres (calculated)
ADDRESSES	Integer	Addressable locations in CA from <a href="#">OS Open UPRN</a> (calculated)
LB_COUNT	Integer	Listed buildings in CA from <a href="#">HE spatial dataset</a> (Jan 2021, calculated)
NEW	Integer	Change flag relative to HE dataset 1=new, 2=change, 3=changed UID
SPLIT	Integer	Numbered components of CA split across several LPA areas (default 0)
ACTION	String	Notes on changes made during data quality review
SOURCE	String	Original source for CA boundary data
SOURCE_CAP	String	Scale at which the original CA dataset was captured, where available
SOURCE_UPD	String	Date when the original CA dataset was last updated, where available
MECHANISM	String	Data provision mechanism (via HE, downloaded, direct etc)
MECH_UPD	String	Date mechanism version was last updated, where available
CONSOL_UPD	String	Date consolidated dataset was last updated using the mechanism

## **Project motivation**

LPAs are responsible for defining and mapping a range of locally designated areas (LDAs); with CAs, covering around 10% of English properties, representing the most widespread form of protection for the local built environment. Comprehensive national datasets have been compiled for some LDAs (such as [brownfield land](#)), but until now no such dataset has been available for CAs.

The [EU INSPIRE Directive \(2007\)](#), implemented domestically in 2009 (with minor Brexit [amendments](#) in 2018), requires public bodies to publish certain datasets in an appropriate spatial format, including the majority of LDA boundaries. In addition, in 2012 the Government initiated a new push to [open up access](#) to public sector data, encouraging the broadest possible use of OGL licensing.

The [Local Government Association](#) (LGA) has set out [further guidance](#) on how these initiatives should be implemented by local authorities, including for [geospatial data](#). In its [guide to INSPIRE compliance in Local Government](#) (2014), it summarised milestones for compliance across all three INSPIRE Annexes, with Annex I datasets (including CAs) scheduled for June 2017, and encouraged publication under OGL wherever possible. These objectives have been further reinforced by the recently published [UK Geospatial Strategy 2020-2025](#) and related (Dec 2020) guidance on improving [access to spatial data](#).

Under its INSPIRE Memorandum of Agreement with the LGA, HE bears responsibility for the creation and maintenance of a national CA dataset, which it makes available on request under OGL. However, although the internal HE dataset covers over 96% of LPAs, many of these have not granted permission to publish, so that the publicly available INSPIRE dataset covers only around 77%. In addition, the HE dataset lacks update timestamps and has been slow to evolve towards full coverage, with only around 2% of LPAs and 3% of CAs added during 2017-2020. This continuing partial coverage acts as a barrier to more widespread use of the dataset by academics, policymakers, and the private sector.

## **Licensing considerations**

Historically, many LPAs have held back from permitting OGL publication of their LDA datasets because of concerns that they might incorporate Crown Copyright material derived from background maps provided by the OS under their public sector licensing agreements. However, an increased recognition of the government's open data objectives has led to a gradual relaxation of constraints in this area, and the April 2020 [replacement of the former Public Sector Mapping Agreement \(PSMA\)](#) by the PSGA has formalised this process. Since then, the OS has significantly enhanced its [guidance](#) to make clear that most LDA datasets, and specifically CA datasets, are suitable for OGL publication under the OS [presumption to publish criteria](#), which simply request completion of a short online notification form.

In the course of preparing this consolidation, this treatment was confirmed with the head of PSGA and government relations at the OS, and was also confirmed direct to numerous LPAs via the OS [PSGA helpline](#). Although some LPAs were initially cautious, all have now confirmed (via HE, the GLA or direct) the OGL licensing treatment. This has enabled the consolidated dataset to be published under an open access [Creative Commons Attribution 4.0 Licence](#), subject to cross-referencing the OGL content.

## **Data consolidation process**

Compilation began with the creation of an initial consolidated 'raw' dataset as described below.

### ***Historic England dataset***

A copy of the current HE INSPIRE dataset was obtained via [on-request download](#) on 2<sup>nd</sup> October 2020, and subsequently confirmed by HE as unchanged on 20<sup>th</sup> January 2021. The HE dataset attaches a unique identifier (UID) to each CA feature, and where possible these UIDs have been retained in the final dataset, to facilitate cross-referencing. All other non-geometric attributes in the HE data were also carried forward into the final dataset, in order to make maximum use of their prior work. For five features, the UIDs were zero or duplicated, so these were replaced with UIDs 1-5. The HE dataset was then compared to an earlier version obtained in January 2017, to understand the evolution of the data in the interim, and this comparison used to generate an estimate for the last update (since HE confirmed that they do not maintain update timestamps). Data for Nottingham LPA was incorporated in the 2017 version of the dataset but not the 2020 version, so this was also added back in using the previous UID codes. Finally, a first-cut revision of names was carried out, by adding the LPA name in brackets to all repeated CA names, in order to begin the move towards a unique naming convention.

### ***Data available for download***

To supplement the HE dataset, those LPA datasets already [indexed via data.gov.uk](#) and made available online were downloaded. In addition, the existing [London-wide consolidation](#) prepared by the GLA was downloaded as a starting point for the remaining London gaps, although the underlying LPA data was ultimately cross-checked for consistency in the majority of cases. Wherever downloadable LPA data was offered under different licensing arrangements to OGL, the LPA was written to in order to confirm the acceptability of OGL treatment, which was ultimately provided in all cases.

### ***Import procedures for additions***

All LPAs absent from the resulting expanded dataset were written to in October 2020, requesting copies of their CA spatial data under OGL. These were gradually provided in a variety of formats, converted into shapefiles with a standardised attribute profile, and added in batches to the evolving dataset. As each LPA's data was received, a basic visual check against background maps was undertaken in the hope of weeding out any obvious placement or coordinate reference system (CRS) errors. Some of the new features provided duplicated those already provided by other LPAs, typically where they crossed LPA borders or fell within National Parks (NPs). In these cases, if the features were spatially identical, the new versions were generally discarded (as the HE dataset typically had richer attributes). Where features varied, both versions were generally retained for future review, mirroring the widespread duplication already present in the HE dataset. The main exception was where a newly provided dataset was clearly more current than the HE version (identified by a quick cross-check of varying features back to the current local appraisal documents), in which case the new features were used in place of the HE ones. One example of this was the replacement of all features in the Craven LA area of the Yorkshire Dales NP by the more current ones provided by the council.

[Chesterfield LPA](#) declined to contribute its boundary dataset on the basis that the data was of too low a quality for broader publication, while Buckinghamshire LPA chose to withhold the data for only the former [South Bucks District Council](#) area of the county, based on continuing licensing concerns. In both cases, good substitute datasets were directly geocoded from the local pdf boundary representations.

A *NEW* attribute was incorporated into the data schema to identify newly added features (1), modified versions of pre-existing HE features (2, where an *ACTION* attribute described the changes) and those features whose UIDs had been changed relative to the HE original (3).

### **Data quality review process**

Having constructed an initial consolidated dataset as described above, a series of data quality review and revision exercises were then undertaken to address various shortcomings in the preliminary data.

#### ***Name standardisation***

Names recorded in block capitals were converted to a conventional 'proper' format with initial capitals only, and trailing 'Conservation Area' components were removed from names. When CA names were repeated in different LPAs, the original name was augmented with the LPA name in brackets to facilitate differentiation between individual CAs, while names that were still repeated were given additional geographical qualifiers. Further minor modifications were made to correct formatting or spelling errors. Overall, around 20% of CA names were modified in this process.

#### ***LPA corrections***

Many of the CAs lying in areas managed by National Park Authorities (NPAs) or London Development Corporation (DCs) had their LPA recorded as the supplier local authority rather than the actual LPA. All CAs lying principally in NPA or DC areas were identified by spatial analysis, had their LPA checked and, where necessary, corrected, and were tagged with an *NPA* attribute set to NPA or DC.

#### ***Correction of invalid geometries***

Around 2% of the original dataset incorporated invalid spatial data geometries (215 in total), largely arising from 'ring self-intersection' errors, with about 65% of these originating in the HE dataset and the remainder in the newly added data. As a necessary precursor to further analysis, these errors were eliminated using GIS software, with the areas of revised features compared to the originals to ensure that no accurate geometries had been compromised.

#### ***Consolidation of multi-feature conservation areas***

While most LPAs consolidated CAs with several disjointed areas into single multi-part features, some provided separate features for each section. In addition, some LPAs split even conjoined areas into several separate features, generally without any clear rationale. Sometimes these separate areas represented identifiable historic expansions (although often the associated attributes were identical), but many appear to have arisen as an artifice of the original geocoding process.

While future versions of this dataset may benefit from a consistent disaggregation of CAs into sub-units corresponding to different designation or expansion dates, the approaches taken in the current raw data were too inconsistent to support such an approach for this iteration. It was consequently decided to consolidate all multi-feature CAs into single multi-part features, so that each line in the attributes file corresponds to a single designated CA (or at least the portion thereof managed by a specific LPA). This correction affected around 0.5% of CAs.

## ***Elimination of overlaps***

The ideal treatment of overlaps between CA features depends on both their origin and the expected uses of the combined dataset. Overall, it is expected that this dataset will be used to identify areas lying within at least one CA and to perform statistical analysis on aggregate CA coverage, so that overlaps should be avoided where possible. There are some deliberate overlaps between CA coverage, generally where smaller local CAs duplicate parts of larger wide-area CAs, and it is possible that a future version of this dataset could treat these areas as separate multi-CA features. However, most overlaps in the raw data are duplications or unintentional errors. Consequently, in the interests of consistency, the aim has been to minimise the area and degree of overlaps in the revised dataset.

The commonest, and most material, category of overlaps arises from *CAs which cross LPA boundaries*, where alternative representations have been provided by the affected LPAs. Sometimes these representations are clipped (often quite approximately) to the LPA boundary, sometimes they partially overlap the border, and sometimes they fully replicate the whole CA. In all cases, the aim of the review process has been to provide two (or more, in the case of multi-LPA overlaps) contiguous CA features split along the LPA boundary, with the boundary line taken from either one of the original features (if apparently accurately clipped) or from the [ONS LPA boundary shapefile](#) (Apr 2019, which is known to have its own shortcomings, slightly modified). Where differences were observed between alternative representations of the outer CA boundary, these were generally checked back to the relevant local maps or appraisal documents, with the most accurate representation of each component used for the replacement features (or the locally generated version used in the event of a conflict).

The second category of overlaps arises from the *apparently deliberate overlaps* described above, and in this case one CA has been treated as 'dominant' with its spatial extent removed from the representation of the 'secondary' CA. For smaller CAs imbedded in larger landscape-scale CAs (such as are common in Norfolk), the smaller CAs have generally been treated as dominant, given their particular local focus. However, there are also numerous smaller CAs overlapping canal or river-based CAs, and in this case the latter have usually been treated as dominant (given their continuous heritage characteristics) and appear as a cut-out through the more locally focused CAs. Some LPAs have already taken their own approach to the elimination of this kind of overlap, often in a way that reflects the underlying appraisal documents, so in this case either existing LPA eliminations or the general LPA local approach have been used, even if they run counter to the default approach to overlaps.

The final category of overlaps arises from *accidental overlaps* between the spatial extent of adjacent CAs, often because the original geocoding failed to 'snap' the boundaries to a consistent line. This has resulted in many boundaries that are meant to be single lines but actually constitute a long string of small overlaps and gaps. These have been addressed according to their materiality, as described below, but remain in significant numbers in the current revision. Their resolution is a matter for future work, and may be addressed in certain areas (such as London) by those with a specific local interest.

Overall, overlaps were approached by generating a consolidated summary via spatial analysis and then systematically addressing them in order of materiality, ranked by both absolute size (area), degree (size relative to the smaller overlapping CA, to pick up substantial overlaps between small CAs) and number of properties. Initially, all 90 overlaps above 0.3ha in area were individually reviewed and modified. Then the smaller spatial overlaps were ranked by degree, and any further overlaps exceeding 5% of the smaller CA area were addressed. Finally, any overlaps still containing properties were eliminated. This process eliminated around 99% of the aggregate overlap area but only around 8% of the total overlaps. However, 98% of the remaining overlaps were minimal (under both 1000m<sup>2</sup> and 1% of the smaller CA area), meaning that only 25 slightly material overlaps remain awaiting review. Overall, around 2% of CAs were modified during this process.

### **Disaggregation of conservation areas spanning LPA boundaries**

While the original LPA boundary datasets and overlap correction process above dealt with many of the CAs spanning LPA boundaries, a number remained represented by a single feature. While this is more acceptable than an overlap, it means that the CAs covered by each specific LPA were not adequately disaggregated, ruling out certain types of spatial analysis. Given that most of this kind of CA were split in the original data, it was decided to split the single feature representations of the more material elements of the residue.

An intersect analysis of CAs and LPA boundaries was performed, and CAs ranked by the proportion of their area or the number of properties lying in a second LPA. These were then assessed systematically, and all CAs with over 15% of their area or over 10 properties in a second LPA were split into separate components. The objective was to capture the most material splits while leaving the less significant ones to be addressed in a subsequent revision.

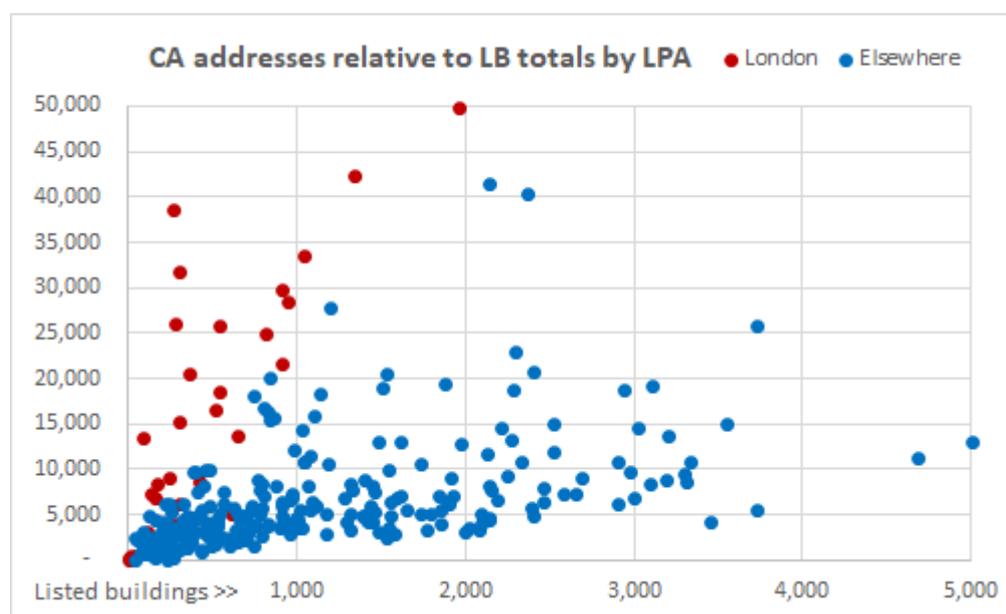
### **Overall results**

In addition to the name standardisation, LPA changes, and geometry error corrections, the initial data quality review process resulted in over 300 changes or additions to the raw dataset. The resulting dataset still contains an array of small overlaps (under 0.3ha and 5% of the smaller CA) and CAs with a smaller part of their area (under 15%) or property count (under 10) lying in other LPAs, but should be fit for most practical purposes. The combined area of residual overlaps is under 0.005% of the total area within CAs, while the proportions of area and properties still represented under the wrong LPA amount to 0.027% and 0.002% respectively.

### **Dataset characteristics**

Version 2.0 of this dataset incorporates 9,779 unique English CAs plus a further 139 areas where CAs overlap a secondary LPA. In aggregate, these CAs cover just over 300,000 ha of land area (2.3% of the national total) and 2.5m addressable locations (10.3%), as well as incorporating around 214,000 listed buildings (56.5%).

There is [marked variation in the coverage of CAs across the country](#), with usage broadly correlated with the number of listed buildings but CAs playing a much more prominent heritage protection role in some areas such as London and other historic cities, as illustrated in the chart below.



## **Dataset maintenance**

While this dataset presents a useful snapshot of the current state of CAs across England, the picture remains one in flux. Although there is a core static element to CA coverage, LPAs are frequently reviewing boundaries, extending and reducing CA coverage, and adding or removing CAs. For any dataset to remain as useful as possible, it is important to address the question of maintenance.

The existing understanding between LPAs and HE means that many LPAs proactively send updates to HE as changes are made, and HE are also understood to request annual updates from contributors on a reasonably systematic basis. However, this process is inherently manual and probably somewhat patchy in terms of consistent participation. The process has also historically lacked timestamping or data quality review, as well as excluding many LPAs who had not granted permission to publish, reducing the usefulness of the data. However, this is undoubtedly the easiest existing mechanism for updating this dataset, and a periodic cross-check using consistent UIDs and basic spatial analysis would allow any user to benefit from updates via HE. If HE were to cross-reference their UIDs for unpublished data to those presented here, and use the expanded agreement to OGL publishing to extend their permission to publish coverage, the process would be further improved and streamlined.

An alternative approach is for all LPAs to maintain current copies of their LDA spatial datasets in a publicly accessible location online, where they can be downloaded manually or via WFS software approaches. Many LPAs have already implemented this methodology, and it is to be hoped that MHCLG and the LGA will continue to encourage further developments in this area. For certain datasets that are already widely available on this basis, the [MHCLG Digital Land](#) team has been developing standardised software to check for updates and create automatic 'live' consolidated summaries. Once all the relevant data is available in this accessible format, ideally with a consistent data schema, the maintenance process will be significantly simplified.

Whichever route is chosen for accessing future updates from LPAs, some cross-referencing will be necessary if the data quality enhancements imbedded in this dataset are to be retained rather than overwritten with new data similar to the original. The best approach would probably involve some coordination between LPAs and the maintainers of any consolidated dataset, to ensure that consistent approaches to the data schema and questions such as overlaps and border splits are taken. This is a question that HE, MHCLG and the LGA may wish to consider further if they see merits in the maintenance of an accurate and current national dataset for CAs or other LDAs.

## **Acknowledgements**

This project would have been much more difficult without the prior consolidation work undertaken by Historic England, and their role in maintaining the existing INSPIRE dataset is much appreciated.

It has been extremely encouraging to see the constructive and helpful approach to open data taken by the vast majority of LPAs during the compilation of this dataset; including, in some cases, a willingness to reconsider their historic licensing concerns. Thank you to everyone who has made the time to help during what has been a difficult period for many local authorities.

Thanks are also due to the PSGA team at the OS for helping to clarify the licensing context, the LGA digital team for their moral support, and the MHCLG Digital Land team for their collaboration and advice on the questions of future maintenance and the broader context.