London Lives 1690-1800: XML Data

Author: Sharon Howard

Date: March 2017

[**Overview**](#_ciat5ph3x7ke) **1**

[**Licensing and Citation**](#_q6sam15btke3) **2**

[**The Files**](#_7idmbxbg0sd6) **2**

[The Document Type Definition](#_yi14ohx572ft) 2

[**London Lives Identifiers**](#_cdhl8qt4cde3) **4**

[**Repository, Archive and Document Codes**](#_2ak5fd89hass) **4**

[Repositories](#_k5v6mvn9fjho) 5

[Archives](#_xes24vkr8b9k) 5

[Document types](#_dt4ka8inw7mw) 5

## Overview

Further information about the project, historical background and a searchable version of the data is available at <http://www.londonlives.org>

The technical methods used to create the data are described at: <https://www.londonlives.org/static/Project.jsp#toc7>

The primary goal of the London Lives (originally Plebeian Lives) project was to create a full-text digital resource based on 18th-century institutional, administrative manuscript sources, which would prioritise the many thousands of non-elite individuals named in those sources over institutional and bureaucratic contexts in order to facilitate searching for people and tracing their lives and experiences.

Therefore, the creation of the XML data emphasised adding certain kinds of semantic markup (rather than, for example, marking document structures): names of people (**<name>**) and places (**<geo>**), dates (**<date>**) and occupations (**<occupation>**). This was done by a combination of automated natural language processing and manually by a group of data developers.

The markup did not use a standard schema such as TEI (although it incorporates some TEI-like elements); the dataset includes a copy of the project DTD and this document does not attempt to describe the markup in any detail. Its purpose is rather to explain key elements of the XML structure and project codes, which are essential to make use of the data.

The London Lives project cannot provide unpaid technical support for users of this data. However, we do wish to encourage research and re-use of the data as much as possible. Specific queries (as well as ideas for funding proposals!) may be directed to londonlives@sheffield.ac.uk but be aware that we cannot guarantee that we will be able to help.

## Licensing and Citation

All London Lives data is released under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0). <https://creativecommons.org/licenses/by-nc/4.0/>

Note that our interpretation of 'commercial' normally excludes academic research outputs, even when published with commercial publishers. Please contact the email above if in any doubt.

Please refer to the citation guidelines on the website for general principles:

<https://www.londonlives.org/static/Legal.jsp#toc3>

The most appropriate citation form for research using data derived from this dataset is likely to depend on the exact nature of the new data, but it should include the project name **London Lives, 1690-1800** and the main website URL **http://www.londonlives.org**, the **version number** of the data, and individual IDs where appropriate. If the new data is also made publicly available we suggest the inclusion of a statement along the following lines:

This work builds on [London Lives, 1690-1800](http://www.londonlives.org) that is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Permissions beyond the scope of this license may be available at <https://www.londonlives.org/static/Legal.jsp>.

## The Files

This document describes the following dataset, distributed in a zipped archived (LL\_XML\_1-1.zip).

**London Lives 1690-1800 XML data files** (1624 files)

Accompanying documentation:

* LL\_XML\_notes.docx (this document)
* listLLfiles.txt
* plebeian-ver1.dtd
* licence.txt

The files are organised in 14 top level folders (representing archives) and 59 second level folders (document types). See below for details of the codes used to name files and folders.

The current version of the data is 1.1 (April 2012).

The XML files were created and processed between about 2007-2009. In most cases, a file corresponds to a **piece** or similar level object in a repository catalogue. Exactly what kind of object (or objects) this represents can vary but it is usually either a bound volume or a bundle of papers. (For convenience, the project may often refer to a file as a "Document", even though a file might in fact represent multiple documents.)

You will find references to 'Plebeian Lives' throughout the data. This was the original name of the London Lives project.

### The Document Type Definition

The XML is validated against a custom DTD (*not* a standard schema such as TEI). A copy of the DTD is included in the zipped archive of the XML files and should be studied to understand the XML markup and structure.

The DTD declaration at the beginning of each file points to an online file which may no longer be accessible. If you open the file in an XML editor and it complains that it can't find the DTD, you will need to delete the declaration or edit it to point to the location of the accompanying DTD file.

## 

## London Lives Identifiers

It is crucial to understand the IDs used in the XML files to uniquely identify various entities; they are based on a set of two-letter codes whose meaning is not immediately transparent. The starting point for all IDs is the filename, which consists of the following elements:

* Repository code, eg **GL**
* Archive code, eg **BA**
* Document type code, eg **EP**
* Document five-digit number, eg **10315**

(See below for further explanation of this structuring and the list of codes.)

Name IDs (which appear as @id attributes in the markup) are formed using:

* Paragraph identifier; usually a number, occasionally a number with letter suffix (must be unique within the file), prefixed with the letter 'n'
* Unique numerical identifier within the paragraph

Eg:

<name given="ANN" surname="ROSS" sex="f" **id="n1-1"**>Ann Ross</name>

So it should be noted that the name ID is unique only within the file, **not** in the data overall. To ensure uniqueness when names from multiple files are extracted into a single database, the London Lives practice is to prefix this ID with the Filename. This example is in **GLBAEP10315**, so the full London Lives ID for this name instance is

**GLBAEP10315\_n1-1**

NB: Most paragraph and name IDs were automatically generated during automated processing and thus will be sequential numbers But this cannot be assumed in all cases. When the automated markup processes failed to correctly identify a paragraph or name, new markup and IDs would need to be inserted manually.

Image (**<img>**) tags are used in the XML to point to the corresponding image files, and are also needed to construct London Lives page-based citations and URLs (please note that we don't recommend using URLs directly for citations in publications, as they could change in the future, whereas IDs are permanent). An image id consists of the filename + a 4 digit number, eg

<img>**GLBAEP103150003**</img>

The corresponding URL for this image ID is

<https://www.londonlives.org/browse.jsp?div=GLBAEP103150003>

## Repository, Archive and Document Codes

London Lives uses the term 'repository' to refer to the modern institutions that hold records, and 'archive' to refer to the 18th-century institutions that created the original records. We use the term 'document type' (not to be confused with an XML Document Type) to refer to distinct types of records, normally based on their organisation in repositories' catalogues.

### Repositories

|  |  |
| --- | --- |
| BB | Bethlem Royal Hospital Archives |
| GL | Guildhall Library (all these records apart from CC are now in LMA) |
| LM | London Metropolitan Archives |
| NA | The National Archives |
| WA | Westminster Abbey Muniment Room |
| WC | Westminster Archives Centre |

### Archives

There is more information about these institutions in the London Lives background information: <https://www.londonlives.org/static/Background.jsp>

|  |  |
| --- | --- |
| BA | St Botolph Aldgate |
| BR | Bridewell and Bethlem |
| CC | Carpenters' Company |
| CD | St Clement Danes |
| CL | City of London Coroners' Inquests |
| CO | Middlesex Coroners' Inquests |
| CW | Westminster Coroners' Inquests |
| DB | St Dionis Backchurch |
| HO | Home Office |
| OB | Old Bailey Gaol Delivery |
| SL | City of London Sessions |
| SM | Middlesex Sessions |
| TH | St Thomas' Hospital |
| WJ | Westminster Sessions |

### Document types

These are documented in detail at <https://www.londonlives.org/static/Documents.jsp>

|  |  |
| --- | --- |
| AC | Churchwardens and Overseers of the Poor Account Books |
| AO | Miscellaneous Parish Account Books |
| AP | Regular Parish Payments to Paupers |
| BC | Books of Clothing Provided by the Parish |
| BE | Enfield Books: Registers of Parish Children put out to Nurse |
| BN | Parish Clothing Books, with Names of Children at Nurse |
| CR | Criminal Registers of Prisoners in Middlesex and the City |
| EP | Pauper Settlement, Vagrancy and Bastardy Examinations |
| GO | General Orders of the Court |
| IA | Apprenticeship Indentures and Disciplinary Cases |
| IC | Coroners' Inquests into Suspicious Deaths |
| IW | Workhouse Inquest (Visitation) Minute Books, St Dionis Backchurch |
| LB | Out-Letter Books: Copies of Correspondence Sent from St Thomas's Hospital |
| LP | Lists of Paupers Receiving Parish Relief |
| LT | Lists of Governor Takers-In of Patients to St Thomas's |
| LW | List of Inmates of St Clement Danes Workhouse in 1785 |
| MB | Minute Books of the General Prison Committee of Bridewell |
| MC | Minute Books of Hospital/Guild Courts and Committees |
| MG | Minute Books of the Court of Governors of Bridewell/St Thomas's |
| MO | Minute Books of Parish Vestry Sub-Committees |
| MV | Minutes of Parish Vestries |
| OO | Overseer's Order Books: Goods Ordered by Workhouse Master |
| PA | Parish Apprenticeship Indentures and Registers |
| PM | Miscellaneous Parish and Bridewell Papers |
| PP | Churchwardens' Vouchers/Receipts from St Dionis Backchurch |
| PR | Letters to Parish Officials Seeking Poor Relief |
| PS | Sessions Papers: Justices' Working Documents |
| RA | Registers of Pauper and Bridewell Apprentices |
| RC | Registers of Poor Children under Fourteen years in Parish Care |
| RD | Register of Pauper Settlement and Bastardy Examinations |
| RH | St Thomas's Admission and Discharge Registers |
| RI | Register of Poor Infants under Four Years in Parish Care |
| RP | Registers of Fortnightly/Monthly Parish Pensioners in 1733 |
| RR | List of Securities for the Maintenance of Bastard Children |
| RV | Register of Paupers Removed from the Parish |
| RW | Register of Paupers Admitted to the Workhouse |
| WB | Bastardy Bonds: Securities for the Maintenance of Bastard Children |