

A Example of Page from the Gazetteer

In this section we provide an example of a page (Figure 1) within the Gazetteers collection and its equivalent available in XML (Figure 2).

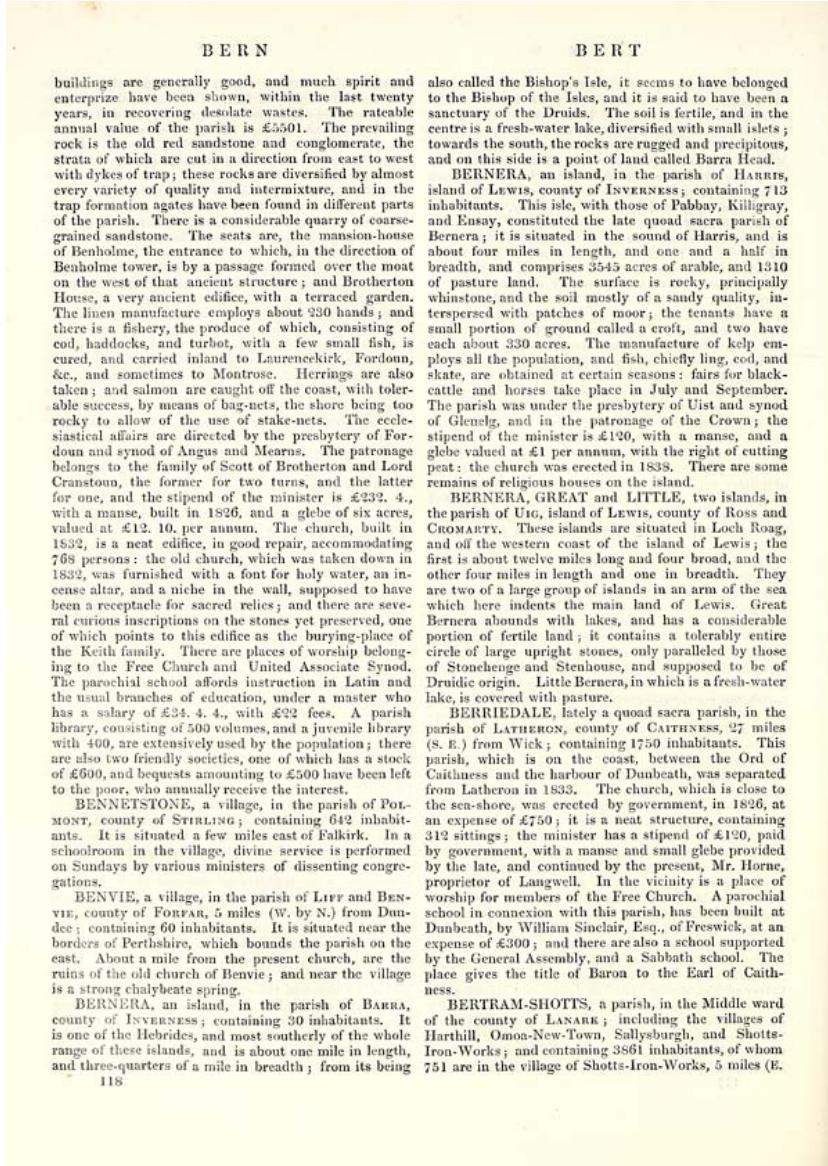


Fig. 1: Page-level encoding for page 118 of the first volume, 1846 edition. Scanned page image available online at <https://digital.nls.uk/gazetteers-of-scotland-1803-1901/archive/97483618>.

```

377   <TextStyle ID="font356" FONTFAMILY="courier" FONTSIZE="6.115" FONTTYPE="serif" FONTWIDTH="proportional" FONTCOLOR="#WWWW
378   FONTSTYLE="superscript"/>
379
380 </Styles>
381 <Layout>
382   <Page ID="Page132" PHYSICAL_IMG_NR="132" WIDTH="2500" HEIGHT="3510">
383     <PrintSpace>
384       <TextLine WIDTH="194" HEIGHT="64" ID="p132_t1" HPOS="591" VPOS="113">
385         <String ID="p132_v1" CONTENT="BERN" HPOS="591" VPOS="113" WIDTH="194" HEIGHT="64" STYLEREFS="font0"/>
386       <TextLine WIDTH="194" HEIGHT="64" ID="p132_t2" HPOS="1714" VPOS="113">
387         <String ID="p132_v2" CONTENT="BERT" HPOS="1714" VPOS="113" WIDTH="194" HEIGHT="64" STYLEREFS="font0"/>
388       </TextLine>
389       <TextLine WIDTH="1080" HEIGHT="25" ID="p132_t3" HPOS="147" VPOS="227">
390         <String ID="p132_w3" CONTENT="buildings" HPOS="147" VPOS="237" WIDTH="173" HEIGHT="25" STYLEREFS="font1"/>
391         <SP WIDTH="33" VPOS="237" HPOS="320"/>
392         <String ID="p132_w4" CONTENT="are" HPOS="353" VPOS="236" WIDTH="57" HEIGHT="25" STYLEREFS="font2"/>
393         <SP WIDTH="33" VPOS="236" HPOS="410"/>
394         <String ID="p132_w5" CONTENT="generally" HPOS="443" VPOS="234" WIDTH="170" HEIGHT="25" STYLEREFS="font3"/>
395         <SP WIDTH="34" VPOS="234" HPOS="613"/>
396         <String ID="p132_w6" CONTENT="good," HPOS="648" VPOS="233" WIDTH="100" HEIGHT="26" STYLEREFS="font4"/>
397         <SP WIDTH="33" VPOS="233" HPOS="748"/>
398         <String ID="p132_w7" CONTENT="and" HPOS="781" VPOS="231" WIDTH="69" HEIGHT="30" STYLEREFS="font5"/>
399         <SP WIDTH="36" VPOS="231" HPOS="850"/>
400         <String ID="p132_w8" CONTENT="much" HPOS="886" VPOS="227" WIDTH="104" HEIGHT="34" STYLEREFS="font6"/>
401         <SP WIDTH="33" VPOS="227" HPOS="990"/>
402         <String ID="p132_w9" CONTENT="spirit" HPOS="1023" VPOS="237" WIDTH="101" HEIGHT="22" STYLEREFS="font7"/>
403         <SP WIDTH="32" VPOS="237" HPOS="1125"/>
404         <String ID="p132_w10" CONTENT="and" HPOS="1157" VPOS="230" WIDTH="70" HEIGHT="31" STYLEREFS="font8"/>
405       </TextLine>
406       <TextLine WIDTH="1079" HEIGHT="25" ID="p132_t4" HPOS="147" VPOS="278">
407         <String ID="p132_w11" CONTENT="enterprise" HPOS="147" VPOS="284" WIDTH="189" HEIGHT="25" STYLEREFS="font9"/>
408         <SP WIDTH="32" VPOS="284" HPOS="336"/>
409         <String ID="p132_w12" CONTENT="have" HPOS="368" VPOS="279" WIDTH="86" HEIGHT="28" STYLEREFS="font10"/>
410         <SP WIDTH="27" VPOS="279" HPOS="453"/>
411         <String ID="p132_w13" CONTENT="been" HPOS="480" VPOS="278" WIDTH="86" HEIGHT="28" STYLEREFS="font11"/>
411         <SP WIDTH="34" VPOS="278" HPOS="566"/>

```

Fig. 2: ALTO XML: words in CONTENT.

B SPARQL Queries for the Competency Questions

In this section we report the SPARQL queries employed for answering the competency questions in Section 5.1 of the main paper.

B.1 CQ1: How is a place described over time?

```

SELECT ?volume_title
?s_page_num ?year ?text
WHERE {
  <https://w3id.org/hto/Concept/gaz2337271726_1>
    hto:hasConceptRecord ?record.
  ?record hto:startsAtPage ?s_page;
    hto:hasOriginalDescription ?desc.
  ?desc hto:text ?text.
  ?volume hto:title ?volume_title;
    schema:hasPart ?s_page.
  ?s_page hto:number ?s_page_num.
  ?edition schema:hasPart ?volume;
    hto:yearPublished ?year.
} ORDER BY ?year

```

Fig. 3: Query for exploring the descriptions of *Brucehaven* across editions.

B.2 CQ2: How is an article extracted?

```

SELECT ?text ?extraction_tool ?source_label ?dataset
WHERE {
  <https://w3id.org/hto/LocationRecord/9931344583804340_97421702_2337271726_0>
    hto:hasOriginalDescription ?desc.
  ?desc hto:text ?text;
    hto:wasExtractedFrom ?source;
    prov:wasAttributedTo ?extraction_tool.
  ?extraction_tool a prov:SoftwareAgent.
  ?source rdfs:label ?source_label.
  ?dataset schema:hasPart ?source;
}

```

Fig. 4: Query to explore the extraction method for *Brucehaven* in 1825.

B.3 CQ3: Where is the place that an article primarily describes?

```

SELECT ?modern_place_wkt ?historical_place_wkt
?modern_place_wktLabel ?historical_place_wktLabel
WHERE {
VALUES ?record
{<https://w3id.org/hto/LocationRecord/9931003343804340_97343436_769119998_0>}
?record hto:refersToModernPlace ?modern_place;
rdfs:label ?record_name.
?modern_place a crmgeo:SP2_Phomenal_Place;
geo:hasCentroid ?centroid.
?centroid a crmgeo:SP6_Declarative_Place;
geo:asWKT ?modern_place_wkt.
?modern_sp crm:P161_has_spatial_projection ?modern_place;
crm:P160_has_temporal_projection ?mp_temporal.
?mp_temporal rdfs:label ?modern_place_temporal_label.
BIND(CONCAT(?record_name, ", ", ?modern_place_temporal_label)
AS ?modern_place_wktLabel)
OPTIONAL {
?record hto:describesPlace ?historical_place.
?historical_place a crmgeo:SP2_Phomenal_Place;
geo:hasCentroid ?historical_centroid.
?historical_centroid a crmgeo:SP6_Declarative_Place;
geo:asWKT ?historical_place_wkt.
?historical_sp crm:P161_has_spatial_projection ?historical_place;
crm:P160_has_temporal_projection ?hp_temporal.
?hp_temporal rdfs:label ?historical_place_temporal_label.
BIND(CONCAT(?record_name, ", ", ?historical_place_temporal_label)
AS ?historical_place_wktLabel)
}
}

```

Fig. 5: Query for exploring the historical location primarily described by the article *Dundee* in the 1803 edition and its modern reference.

B.4 CQ4: What places are mentioned in an article?

```

SELECT ?modern_place_wkt ?modern_place_wktLabel
WHERE {
VALUES ?description
{<https://w3id.org/hto/OriginalDescription/9931344583804340_97421702_2337271726_0NLS>}
?description hto:text ?text.
?annotation oa:hasBody ?location;
oa:hasTarget ?text_segment.
?location rdfs:label ?place_name;
geo:hasCentroid ?centroid.
?centroid geo:asWKT ?modern_place_wkt.
?text_segment oa:hasSource ?description;
oa:hasSelector ?selector.
?selector oa:start ?start;
oa:end ?end.
BIND(CONCAT(?place_name, ", appears in text: | ",
?text, "| at ", STR(?start), "-", STR(?end)) AS ?modern_place_wktLabel)
} ORDER BY ?start_index

```

Fig. 6: Query to explore places mentioned in *Brucehaven* in the 1825 edition.

B.5 CQ5: How do the socio-economic roles of a place change over time?

```

SELECT
?year
?record
?snippet
(GROUP_CONCAT(DISTINCT ?kw; SEPARATOR=", ") AS ?keywordsFound)
?wkt
WHERE {
  # Place concept
  ?concept a skos:Concept ;
    rdfs:label "Edinburgh" ;
    hto:hasConceptRecord ?record .

  ?record a hto:LocationRecord ;
    hto:hasOriginalDescription ?desc ;
    hto:refersToModernPlace ?modernPlace .
  # Edition / year
  ?record hto:startsAtPage ?page .
  ?volume schema:hasPart ?page .
  ?edition schema:hasPart ?volume ;
    hto:yearPublished ?year .
  # Text
  ?desc hto:text ?text .
  BIND(LCASE(?text) AS ?ltxt)
  # Socio-economic keywords
  VALUES ?kw { "harbour" "railway" "factory" "mill" }
  # Keep only articles where this keyword appears
  FILTER(CONTAINS(?ltxt, ?kw))
  # Geometry (for mapping if needed)
  ?modernPlace geo:hasCentroid ?geom .
  ?geom geo:asWKT ?wkt .
  # Snippet
  BIND(SUBSTR(?text, 1, 500) AS ?snippet)
}
GROUP BY ?year ?record ?snippet ?wkt
ORDER BY ?year

```

```

SELECT ?year ?keyword (COUNT(DISTINCT ?record) AS ?articleCount)
WHERE {
  # Fix the place concept (Edinburgh here)
  ?concept a skos:Concept ;
    rdfs:label "Edinburgh" ;
    hto:hasConceptRecord ?record .

  ?record a hto:LocationRecord ;
    hto:hasOriginalDescription ?desc ;
    hto:refersToModernPlace ?modernPlace .

  # Edition / year
  ?record hto:startsAtPage ?page .
  ?volume schema:hasPart ?page .
  ?edition schema:hasPart ?volume ;
    hto:yearPublished ?year .

  # Text of the article
  ?desc hto:text ?text .
  BIND(LCASE(?text) AS ?ltxt)

  # Socio-economic keywords
  VALUES ?keyword { "harbour" "railway" "factory" "mill" }

  # Keep rows where the article mentions this keyword
  FILTER(CONTAINS(?ltxt, ?keyword))
}
GROUP BY ?year ?keyword
ORDER BY ?year ?keyword

```

- (a) Query 5.A: Retrieve *Edinburgh* articles on socio-economic infrastructures with per-edition keyword matches.

- (b) Query 5.B: Aggregate, per edition, the occurrence of socio-economic keywords.

Fig. 7: Queries for CQ5: analysing socio-economic change across editions.