

Historical Settlement Units as Linked Open Data

Olof Karsvall¹[0000-0003-4350-7533]

¹ The Swedish National Archives • TORA

Abstract. TORA is a historical geographic register provided by the Swedish National Archives. It is built around the concept Historical settlement unit, referring to villages, hamlets and farms in the medieval and early modern rural society. Currently, approximately fifty percent of all such units in Sweden are covered, which are mentioned in written sources and maps before c. 1800. This article discusses and presents the definition and the method of the core concept. Using linked data technologies, the goal is to increase, simplify and improve the use of historical sources in research and applications.

Keywords: historical settlement unit, linked open data, gazetteer.

1 Introduction

Most data on the Web can be related to places. Place as a resource on the Semantic Web is therefore a fundamental component, which sets data about people, events and other things into a geographical context. Location data make delimitation of large data sets possible. It also enables spatial presentations and analyzes, and thus is a complement to lists and graphs on the linked data cloud.

The availability of spatial data today is however limited. Authorities and other actors offers free map services but rarely exposes the underlying data, which is necessary for places to be linkable resources. Map services or gazetteer, providing place names and geolocations, such as GeoNames and Open Street Maps, partially fills this gap, although data quality is not guaranteed, as these services are aggregations of various data sources of different kind, and furthermore require CC-BY licenses or similar for re-use.

The lack of geographical data applies not least to historical places. Few map services handle historical locations, which means they are not particularly useful in relation to historical sources and research. Of course, many historical places — not least towns and major settlements — are still places today. In situations where the aim is to get an overview of historical data, it might not be necessary to have accurate historical references and positions. But in other cases, when a place referred in a historical document is to be defined and described, then proper historical geo data are required.

To allow accurate historical geocoding of places, exposed as linked data nodes on the Semantic Web, the concept Historical Settlement Unit has been created. The initiative has been taken within the research project TORA at the Swedish National Archives.¹

The goal is to establish a comprehensive register that could be used by various researchers as an authority register or gazetteer, which reduces the need for others to create their own historical place registers. A further purpose is to encourage geographical and historical surveys. It's exposed in the linked data format (RDF) and could be used in various search application and, for example, with named entity recognition tools.² At present the register covers 27 000 settlements, which corresponds to between one third and half of all settlements during the Middle Ages until the period around c. 1800 (see Fig. 2).

This article has a methodological aim: first to define the core concept, thereafter to describe the data model and delimitations. Similarities and differences with other historical gazetteer will be discussed throughout. The main question raised is: How could a historical-geographical register and concept be created, that meets the need of being intelligible on the Semantic Web, while fulfilling scientific criteria, as source criticism and accuracy, that makes it relevant as a resource in research?

2 The definition of Historical Settlement Unit

From the Middle Ages until the agriculture and industrial revolution of the 18th and 19th century, the society consisted exclusively of rural villages, hamlets and farms. Historians have emphasized that the village organization, as we know it from historical sources, began to be established during the 9th and 10th centuries; a system combining individual ownership of land with collaboration farming practices.³ With an increasing amount of sources from the High Middle Ages, as e.g. law books and cadasters, these settlements appear as spatial or territorial units, with rules for how land were distributed within the villages and with defined boundaries towards neighboring villages.

The purposed definition of Historical settlement unit is based on the medieval and early modern rural conditions where settlements, such as villages, hamlets and single farms, appeared as basic spatial, functional and economic units. Also towns and larger farms, such as demesnes belonging to lords, churches and monasteries, could constitute such a unit. Although they differ in size and function they appear in a similar way as unit in the historical rural landscape. Also abandoned villages and farms, that often are vaguely documented but likely had been settled, could be classified as Historical settlement units.⁴

¹ Since 2003, a major work has been carried out at the National Archives in Stockholm concerning digitalization of the oldest large-scale maps from the 17th century over Sweden (<https://riksarkivet.se/geometriska>).

² An example of this is discussed and described in Karsvall & Borin 2018.

³ See e.g. Wickham 2005 p 514–518.

⁴ Karsvall 2011 p 22; 2016.

A demarcation around year 1800 is needed since, as a result of a major landscape changes, the historical settlement units are less relevant. Land reforms, industrialization and urbanization did re-shape and re-define many settlements. Not least the enclosure land reform *Laga skifte* in Sweden, which resulted in a new division of the land, had great impact on settlement patterns, as villages and hamlets more or less ceased being functional units and instead became groups of individual farms. Although the place names often were kept, the spatial definition of a historical place is not necessarily relevant after the last two centuries rapid changes. In other words, the slow development in landscapes and settlements – the long duration or *La longue durée* to use the term of Fernand Braudel (1958) – motivates a creation of ‘historical’ settlement as a concept for the Semantic Web.

3 Creation of gazetteer

A historical gazetteer can be created in different ways, either from the ground up on the basis of original sources such as maps and written sources or fully or partly based on published secondary sources. The conditions for establishing a register vary between different countries depending on access to source material, especially historical maps and related geographic and topographical sources.

Researchers can use historical atlases and other reference works about places. Digitization of such books would simplify access but not improve usage.⁵ A greater benefit is achieved when the reference works go through a digitalization process and becomes machine-readable using a data model and place identities, that various sources could refer to. The development of geographical applications (GIS) has considerably simplified the ability to create, analyze and publish geographic data. Different methods have been examine in related to geocoding of historical sources. Research projects have put focus in creating digital historical atlas (“Historical GIS”, or “HGIS”) developing Web-GIS search applications.⁶

In recent years, initiatives have been taken to establish digital geographical registers or gazetteers, and services that use them.⁷ Open data formats and aggregation of geo-data has gained greater attention. Also crowd-sourcing has been used as a method of transcribing and converting large amounts of text sources into data.⁸

⁵ E.g. <https://sok.riksarkivet.se/rosenberg>.

⁶ Examples of such projects: National Edition of the Oldest Geometrical Maps 1630–1655, <https://riksarkivet.se/geometriska>; SVEA-Pommern Karten und Texte der Schwedischen Landesaufnahme von Pommern 1692–1709, <http://www.svea-pommern.de>; Digitalt atlas over Danmarks historisk-administrative geografi, <http://digdag.dk>, see also <http://www.hgis.org.uk> for a list of other similar resources.

⁷ E.g. <http://commons.pelagios.org>; <http://whgazetteer.org>.

⁸ The GB1900 Gazetteer, which publishing place names from the period around 1900, is likely one of the largest datasets of this kind so far online, holding over 1.7 million identifications and related coordinates, created through crowd-sourcing (Aucott et al 2018; Southall et al 2017; <https://geo.nls.uk/maps/gb1900>).

The editorial work of creating a register or gazetteer of Historical settlement units have so far (in the TORA project) considered the conditions in the Nordic countries and especially Sweden. There are two comprehensive source materials that enables accurate or precise definitions of historical places in Sweden: firstly, the Crown's cadastres from the 1530s and beyond; secondly the land survey maps from the 1630s and onwards.⁹ When used together, these two sources are invaluable when creating a historical gazetteer. The cadastre reports almost all villages and farms by name in a uniform way for the whole Sweden by mid 16th century. The survey maps, especially the oldest series from 17th and 18th century, provides a detailed view of the rural landscape. Without access to historical maps, and primary large-scale maps, there is hard to make precise identification of historical places. Although, as an alternative, it could be possible to reconstruct the locations of settlements by combining data from older written sources with more present maps.¹⁰ It is nevertheless the great availability of historical maps in Sweden that is the foundation of Historical settlement unit, that is described in more detail below.

4 Point objects vs. polygons: How the spatial coordinates are set

So far we have specified what kind of object Historical settlements units are and their time period. In the following, a deeper insight is given on how the geographical coordinates are extracted, that allow these historical records to be exposed as spatial data objects.

As mentioned, settlement such as villages and hamlets were legally defined by property boundaries, which separated one unit from other surrounding units. One way to define these would be to reconstruct the external property boundary. In practice, however, this is rarely possible. Two reasons can be stated. Firstly, it is difficult and not always possible to interpret and transfer the boundaries on a historical map into a modern map. It assumes that all boundaries are visible on older maps, which seldom is the case. Secondly, it takes far too much time to make complete rectifications of village boundaries.

In a gazetteer, it is reasonable to define administrative divisions, such as parishes, using polygon boundaries, but not rural settlements. Boundary rectification could be an aim of a research projects, but not the foundation in a digital register holding thousands of settlements. Consequently, it is better to define historical settlement as point objects, i.e. as coordinate pairs (latitude and longitude).

The choice of using point objects requires a method; a definite way for how the coordinate pair are set. A settlements such as a village forms an area of one or more square kilometers. There are therefore several alternatives locations where the coordinate points could be set. The official map service of place names in Sweden today holds coordinates of place name positions. The aim is cartographic, and the place names have

⁹ For further information on these sources, see, for example Kain & Baigent 1992 p 49; Karsvall 2007; 2016.

¹⁰ See e.g. Panecki et al 2018.

been set next to the settlement area to improve visibility. Such principle is less suitable for defining actual settlement locations.¹¹

A better way to define historical settlements would be to define them as spatial (geographical) units, focusing on the cultivate land, and primary the most valuable land, as in most cases the arable land. Only one coordinate pair is to be set for each settlement unit. Since there could be several farmsteads within a settlement it is not suitable to use the location of farmsteads. Individual farms could be disbanded or displaced, although the settlement as a unit remains. The farmsteads itself could also carry their own coordinates in other registers, and should not to be mixed with the settlement unit. One way would be to put the coordinate as centrally as possible. The problem with such a principle is, however, that both farms and land often appear at a certain part of the village, which rarely represents the center of the settlement area.

A better way to define settlements is therefore to use the historical map available, and set the coordinates at a location that according to the map constitutes the core cultivated part of the farmland. From a geographical and economic perspective, this location correlates to the most valuable land, usually the arable land lying next to the building plots. This means that surrounding land, such as forests and pastures, will be irrelevant for the spatial definition, which instead aims to point out the central part of the settlement. In situations when farmsteads are dispersed or when there are several scattered arable fields, the coordinates are set at the assumed core arable area.

For identification of the settlements, land survey maps created before c. 1800 are used in the first place. In the absence of such older historical maps, identifications can also be done (but with less accuracy) with the help of more modern maps from 19th century and later. Such a procedure – using much younger maps – which could be perceived anachronistic, does not have to be so. By adding the criterion the place should be mentioned in a historical written source (before c 1800), there are reason to believe the settlement found the later map (holding the same place name) is the same object. Rural settlements have often, at least in the Nordic countries, been stable over extensive period of time. Most changes occurred *within* the villages – the number of farms could increase or decreased during periods of expansion and crises, but the external boundaries that form the settlement as a unit were often kept intact. Even during periods of decline, settlements have often been re-settled.¹²

To conclude this discussion: The aim of setting coordinates to Historical settlement units is to identify what likely was the accurate location during the middle ages and/or the early modern period. The use of later modern maps for identification results in greater uncertainty. Two statements should be emphasized. Firstly, one coordinate pair (latitude and longitude) represent a settlement unit, more precisely the farm land – roughly one or a few square kilometers. Secondly, the coordinates are set at a central location, relative to the core cultivated land, nearby the farmsteads and nearby arable land.

¹¹ The official place names in Sweden is searchable at <https://kso.etjanster.lantmateriet.se>.

¹² The method of analyzing older historical circumstances on basis of younger sources is well proven and known as *retrogressive* by historical geographers (see e.g. Karsvall 2013 p 411).

5 Data model and vocabulary

The registration of historical settlement units in Sweden started in 2003 in a previous project, but the data model and the format have recently been adjusted to linked data standards, the Resource Description Framework (RDF). This section will describe the adaptation into linked data standards and the design choices made. There are two main reasons why RDF has been chosen as the data model: firstly, it is a web standard for how data should be published online, secondly RDF as a model (subject – predicate – object) is easy to understand for most users, which is important when creating a gazetteer.

Historical place as a semantic web concept needs to address at least the following four basic classes: *place name*, *spatial location*, *source* and *time*. The principles behind the last three are discussed in the sections above. Regarding time, we have found that the concept *Historical settlement units* are considered to be rather stable over time. For this reason time is *not* an attribute of the data model, rather the units are considered to have a potential validity extending from c. 1000–1800. Information about when a specific settlement appear in various sources could be linked to a settlement unit, and does not need to be attributed in the core class definition. An attempt to define validity in time for individual settlements would much likely fail. Information on starting and ending rarely occur in the source material, and has to be based on subjective assumptions, which is not suitable in basic register. Temporal attributes are likely better suited for more general geographical data (as e.g. parishes).¹³

The usage of *source* in the data model specifically refers to the historical maps that points out the coordinates. A separate class is used to store these data (TORA:CoordinateSource), which holds references to the core class (TORA:HistoricalSettlementUnit). In other words, the map(s) that have been used to identify the location constitutes the coordinate source. The oldest map is primarily referred, but in many cases more than one map are relevant as a source.¹⁴

Place names that are associated with a particular settlement could also be managed with the help of related classes. Term search related to place names is of course essential for a gazetteer. But how to attribute a place name to a historical settlement unit is not always obvious. A historical name of a settlement, known from sources, does not necessarily correspond with the contemporary name for the same place. Different sources can also provide different names. And it is often not reasonable to manage all occurrences of names when defining a settlement unit in a basic register. A concept like Historical settlement unit need a preferred name, to be shown along with the coordinates. The name to be used is not always given. It should be a name valid and relevant

¹³ A historical gazetteer with temporal attribute is proposed by i.e. <http://whgazetteer.org>.

¹⁴ The coordinates that are extracted from the historical maps have been stored using the Swedish official coordinate system or projection, SWEREF 99 TM (EPSG:3006). Today's GIS tool allows easy conversion between different coordinate projections. In order to simplify reuse, both the national and international standard WGS 84 (EPSG:4326) are exposed as linked data. However, when the coordinates are rendered on a Web page it is often preferable to use the Mercator projection (EPSG:3857) which translates WGS 84 into a sphere.

for the period (in this case c. 1000–1800) that is documented in the sources. To exemplify, if a rural village at some time has developed into a town with a different name, or if the name changed for other reasons, the historical name should be given primarily, while the later represents alternative names.

The chosen way to manage different place names for a settlement is to choose a preferred name (`skos:preLabel`), while all other names are handled in a separate related class with alternative names (`TORA:AlternativeName`), that could be derived from different sources.

Another aspect of place name concerns spelling. It is accepted that names should be spelled according to the usual spelling on present official maps. However, if names only appear in older sources, it may be difficult to normalize the spelling. Then it may be better to quote the name according to a source instead. The chosen way to handle this have been to classify each name as either *modern*, *normalized* or *cited* in a related class (`TORA:NamingPrinciple`), where “modern” relates to names appearing on official maps, “normalized” refer to lost place names with a standalone spelling, and “cited” refer to a quote in written source. Many historical names have a known name forms (prefix/suffixes) and could therefore be normalized, which facilitates reading and use. Another reason to normalize names is that older spellings vary – the same name can be spelled in several ways in the same document.¹⁵

¹⁵ The naming principle used in the TORA project has been developed and discussion in cooperation with researchers at the Institute for Language and Folklore in Uppsala, and may be further refined.

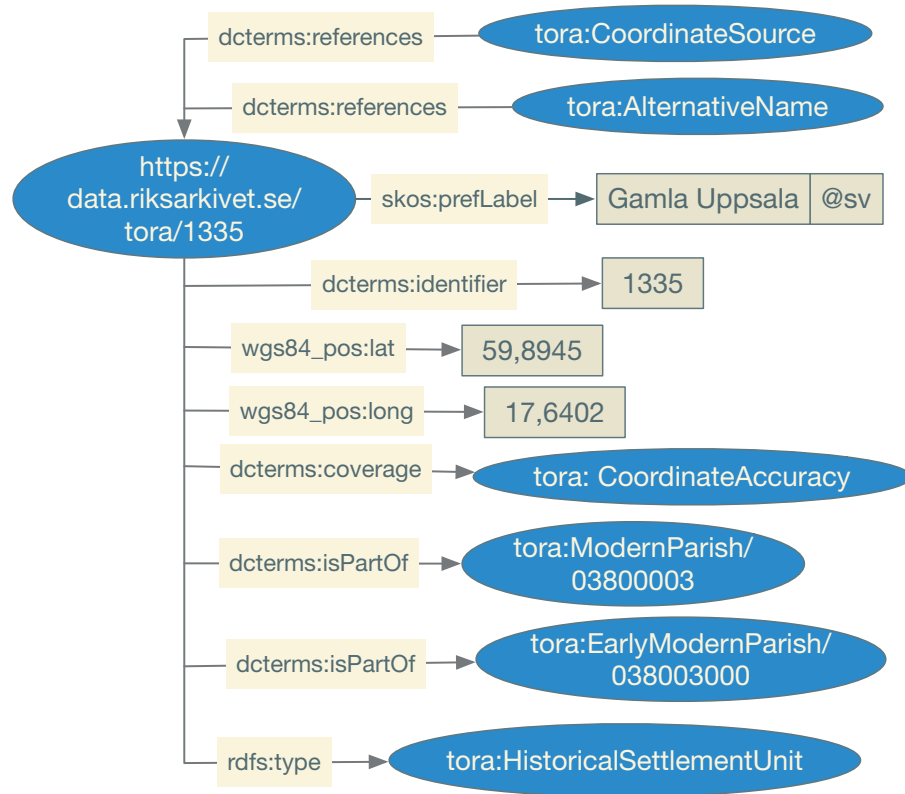


Fig. 1. The Historical settlement units are exposed as linked data in the RDF format. Each unit has a stable URI which constitute the subject or the resource. The relations are expressed as predicates using standard vocabulary or TORA terms. The objects are either nodes or attributes with a value string. The complete TORA model includes other referenced data, but this figure focuses on the core concept of Historical settlement unit.

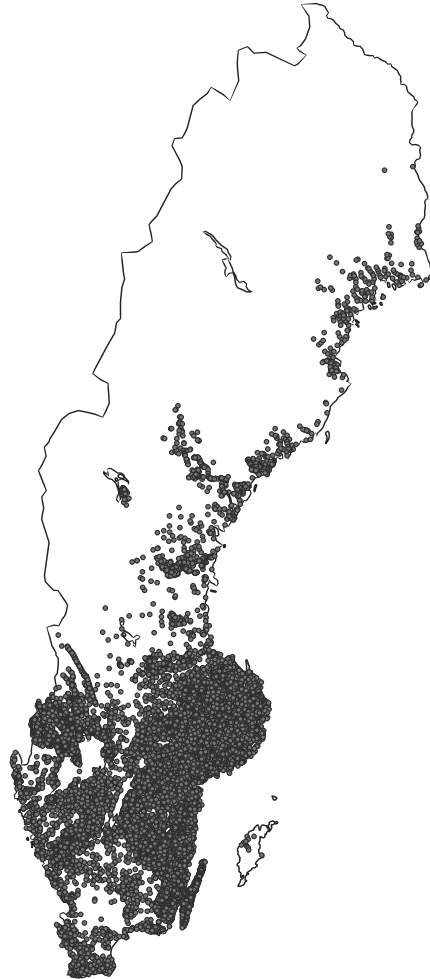


Fig. 2. An overview of the current (February 2019) Historical settlement units registered in TORA, c. 27.000 units covering large parts of Sweden, derived from large scale maps and historical source before c. 1800.

6 Conclusion

A *Historical settlements unit*, as defined in this article, refers to villages, hamlets and single farms which formed the building blocks in the rural society from the Middle Ages until the early modern period. It intends to define historical places, which lack clear definition on the Semantic Web today. It tries to meet the researchers need for better historical context and data definitions, and uses linked data technologies for modeling and publication.

In brief the following criteria constitute a Historical settlement unit: 1/An accurate coordinate pair (latitude and longitude), that are 2/derived from historical map(s), preferable large-scale survey maps; 3/a chosen place name that is valid for the settlement in its historical context, along with 4/a specified spelling principle and 5/alternative names when so occurs.

This paper concludes that the simple RDF data format is sufficient when creating a gazetteer, although historical settlement is a complex spatial object, that could be of different kind (village, town, manor etc.) and hold temporal aspects and several place names. The simplicity of the RDF is seen as a strength, which clarifies what constitutes the core classes and its relations and metadata. In the TORA project, only spatial coordinates as point objects (latitude and longitude) have been handled. If additional geo-data is to be included, such as polygons with time intervals for administrative units such as parishes and larger areas, the model likely needs to be expanded.

Different fields of research, such as history, archeology and onomastics, would benefit from online historical geographical registers, for instance when analyzing relations between people, events and places. It could also be used as a tool in community planning and be a resource in genealogy and local historical research.

7 Acknowledgement

This work presented here is a part of the TORA project, that has been initiated by a group of researchers and co-workers at several institutions in Sweden, with funding (2015–2019) from Kungl. Vitterhetsakademien, Riksbankens Jubileumsfond and the Swedish National Archives. A first public launch will take place 2019 (<https://riksarkivet.se/tora-english>).

References

1. Aucott, P., Fleet C. and Southall, H. 'The GB1900 project – from the horse's mouth', Sheetlines 111 (April 2018), <https://s3.eu-west-2.amazonaws.com/sheetlines-articles/Issue111page46.pdf>
2. Braudel, F. Histoire et sciences sociales: La longue durée. *Annales: Économies, Sociétés, Civilisations*, vol. 13, n. 4, pp 725–753 (1958).
3. Kain, Roger J. P. & Baigent, E. *The Cadastral Map in the Service of the State: A History of Property Mapping*. Chicago: University of Chicago Press (1992)
4. Karsvall, O. Uninhabited cadastral units on large-scale maps 1630–1655: indicators of late medieval deserted farms? In: *Landscape History* Vol. 29, 2007:1, pp 75–87 (2007), <https://www.tandfonline.com/doi/abs/10.1080/01433768.2007.10594590>
5. Karsvall, O. Utjordar and the question of deserted farms – a case study of the parish of Svanshals. In: *Bebyggelsehistorisk tidskrift*, vol. 61, 2011, pp 22–38 (2011), http://media.bebyggelsehistoria.org/pdf/BHT61_2011_22-38.pdf
6. Karsvall, O. Retrogressiv metod: En översikt med exempel från historisk geografi och agrarhistoria. In: *Historisk tidskrift* (Sweden) 133:3, pp 411–435 (2013), http://www.historisk-tidskrift.se/fulltext/2013-3/HT_2013_3_411-435.htm

7. Karsvall, O. Utjordar och ödegårdar: En studie i retrogressiv metod. Diss. Uppsala (2016), <https://pub.epsilon.slu.se/13711/>
8. Karsvall, O. & Borin, L. SDHK meets NER: Linking place names with medieval charters and historical maps. In: DHN 2018, Digital Humanities in the Nordic Countries 3rd Conference. Mäkelä & Tuominen (eds), pp 38–50 (2018), <http://ceur-ws.org/Vol-2084/paper3.pdf>
9. Panecki, T., Związek, T. & Myrda, G. The Spatial Database Development for Research on Settlements and Afforestation in Nowy Tomyśl Plain in Early Modern Times. In: *Studia Geohistorica*, Nr 06. 2018, pp 186–205 (2018), <http://atlasfontium.pl/index.php?article=webhauand>
10. Southall, H., Aucott, P., Fleet, C., Pert, T. and Stoner, M. ‘GB1900: Engaging the Public in Very Large Scale Gazetteer Construction from the Ordnance Survey ‘County Series’ 1:10,560 Mapping of Great Britain.’, *Journal of Map & Geography Libraries*, 13 (1), (2017), pp 7–28
11. Wickham, C. *Framing the Early Middle Ages: Europe and the Mediterranean 400–800*. Oxford: Oxford University Press (2005)