7 March 2025

English

United Nations Group of Experts On Geographical Names
2025 session
New York, 28 April – 2 May 2025
Item 5 (b) of the provisional agenda *
Technical expertise: Geographical names data management

Stadnamnportalen, the Norwegian geographical names portal

Submitted by Norway**

Summary

Stadnamnportalen is an innovative technical platform developed for the management of geographical names under the Norwegian Language Collections project of the University of Bergen.[1] The platform is an example of how modern technology can enable evidence-based standardization while at the same time preserving cultural heritage. Launched in June 2024, it is currently used to manage 1.7 million entries and 3.8 million source forms, which makes it one of the most comprehensive geographical names management systems in the world.

The technical architecture is the result of a deliberate move away from traditional database-centric approaches. Instead, a more flexible infrastructure was adopted that better accommodates diverse user needs and heterogeneous datasets. The system utilizes Git Large File Storage (Git LFS) for data retention and Elasticsearch for search capabilities, which enables direct file-to-index processing and significantly reduces system complexity while maintaining robust functionality. A key innovation is the platform's data management strategy, whose salient features are flexibility and standardization. Each dataset keeps its independent search index while supporting aggregated indixes on the basis of standardized place name identifiers. The document structure is based on JavaScript Object Notation (JSON), which lets it accommodate both unconstrained original data fields and standardized fields with defined constraints, thus ensuring preservation of source material while enabling consistent cross-dataset functionality.

Another feature of the platform is its comprehensive cross-referencing using the Universally Unique Identifier (UUID) standard while maintaining connections to external reference systems such as Wikidata for administrative units and the Norwegian feature type codes (Systematic Organization of Spatial Information,, SOSI) for geographical feature types. Integration with PostGIS software enables sophisticated spatial data handling, while an image server compliant with the International Image Interoperability

* GEGN.2/2025/1

** Prepared by Peder Gammeltoft, Norway, Norwegian Language Collections, University of Bergen

1 Available at https://stadnamnportalen.uib.no.

Framework (IIIF) facilitates the management of historical documentation.

The frontend leverages Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), the JavaScript programming language and the Leaflet.js JavaScript library to ensure interactive mapping capabilities and offer multiple options for data visualization such as list, hierarchical and table views. The interface is fully compliant with the Web Content Accessibility Guidelines (WCAG), which ensure universal accessibility as required under Norwegian law. Automated processes, implemented through GitLab Continuous Integration and Continuous Delivery (CI/CD) pipelines and so-called cron jobs is used to handle data preparation and updates. Scripts written in the Python programming language are used to manage the conversion of various input formats to Elasticsearch-ready JSON files, while RESTful application programming interface (API) support efficient data retrieval and display.

The platform's flexibility demonstrates how the use of modern technical infrastructure can support the inclusive, culturally informed standardization geographical names and the result in evidence-based tools for sustainable development. Its combination of standardization and flexibility, coupled with robust automation and user-friendly interfaces, is a valuable model for geographical names authorities around the world that are considering digital transformation of their resources.

Introduction

Every now and then, one gets the opportunity to create something entirely new. At the Norwegian Language Collections, we have spent the last years working on creating an online resource for geographical names in Norway. Since the establishment of the Language Collections in 2016 at the University of Bergen, creating a unified portal for geographical names across Norway – in both time and space – has been a goal. A database that would show both geographical name entries and their historical spellings – while simultaneously allowing work with individual data sources. Such a resource did not exist in Norway, and it was generally thought to be impossible to create. However, it was and is now available online as Stadnamnportalen (The Geographical Names Portal).

The Stadnamnportalen platform (https://stadnamnportalen.uib.no/) is a comprehen-sive digital infrastructure for managing Norwegian geographical names. This system represents a significant advancement in the preservation and accessibility of geographical names heritage, combining modern web technologies with efficient data management approaches applied to historical and current datasets. The resource represents a significant advancement in geographical names management, currently hosting over 1.9 million entries and 5.9 million source forms, see Figure 1 in the Illustrations appendix at the end of this paper. The platform's technical infrastructure was specifically designed to address the challenges of preserving and standardizing geographical names.

Norway was early in digitizing geographical names sources – perhaps a bit too early. This effort resulted in a series of widely different digital datasets and portals, some nationwide, others regional or local. Some datasets include audio, others feature digital images of paper slips and name lists – and some of these have coordinates, but most do not. While groundbreaking, the digitization efforts resulted in fragmented datasets lacking crucial standardization elements. This fragmentation particularly affected the representation of minority language place names and historical variants, limiting their accessibility and use in official contexts. Recognizing these challenges, the Norwegian Language Collections initiated development in 2018 on an online resource, aiming to create a comprehensive platform that would cater to research, as well as legal-administrative needs and cultural heritage preservation. At the same time, the portal also needed to be of use for the public.

The standardization of geographical names in Norway represents a crucial intersection of cultural heritage preservation and modern digital infrastructure. The Norwegian Language Collections' experience in developing Stadnamnportalen offers valuable insights into how technical solutions can

support inclusive and culturally informed approaches to geographical names standardization at the same time as providing evidence-based source material for standardizing bodies and research in general.

Technical platform

The technical infrastructure supporting Stadnamnportalen deliberately moves away from traditional database-centric approaches, instead adopting a more flexible architecture that better serves the needs of diverse user groups and more easily accepts heterogeneous datasets. At its core, the system utilizes Git Large File Storage (LFS) for data retention and Elasticsearch for search capabilities. This approach allows for direct file-to-index processing, significantly reducing system complexity while maintaining robust functionality.

The platform's data management strategy emphasizes flexibility and standardization. Each dataset maintains its own search index, preserving original field structures while simultaneously supporting aggregated indices based on standardized place name identifiers. Data preparation is handled through Python scripts, with both raw data and Elasticsearch-ready JSON files stored in LFS. This preparation process is fully automated through GitLab CI/CD pipelines, with additional support for scheduled updates via cron jobs for datasets requiring continuous maintenance. This architectural choice proves particularly valuable in managing pluralities of datasets. Document management within the system follows a structured yet flexible approach. The JSON document structure accommodates both unconstrained original data fields and standardized fields with defined constraints. This dual approach ensures preservation of source material while enabling consistent cross-dataset functionality. The system accepts various input formats, with a preference for flattened JSON structures, and implements nested document handling only when essential for search functionality. Each dataset maintains its own search index while supporting aggregated indices based on standardized geographical name UUID-identifiers.

The frontend of Stadnamnportalen leverages modern web technologies to provide an intuitive user experience. Built on HTML, CSS, and JavaScript foundations, the platform incorporates Leaflet.js for interactive mapping capabilities and employs responsive design principles to ensure accessibility across various devices. The integration of PostGIS enables sophisticated spatial data handling, allowing users to explore geographical names through an interactive map interface.

User Experience

The majority of the datasets of Stadnamnportalen are geocoded and primarily organized under common geographical name entries. This makes it possible to search both comprehensively across all datasets, as well as in individual datasets. There are a number of filtering possibilities. A particular powerful filter is the filtering by area – mainly counties and municipalities, but also down to first and second cadastral level. Results are possible in several different views list view, hierarchical view, and table view of data (see Figures 2-4). To help find the preferred way of working with the portal, there is both a search guide and the ability to see what datasets are included on the front page – and select them individually if desired.

In the overall search, if one searches for a geographical name like Mørløs, it is possible to search either the full term or using a wildcard like Mørlø*' (open wildcard search) or 'M?rløs' (single letter wildcard search). In the open wildcard search you will get hits on several geographical name entries, such as *Mørløs*, *Mørløsdalen* and *Møløsvannet*, whereas in the single letter wildcard search you will get just *Mørløs*. Clicking on the name form 'Mørløs' brings up a map showing the location of the name locality and the historical written forms of the name. The dropdown menu to the right of the entry allows you to see all occurrences of the name in different datasets and see if there is e.g. an index card or a sound file associated the geographical name entry. Some of the individual source entries digitized paper slips attached, such as 'Mörlös' in the Bustadnamnregisteret (Settlement Name Register), see Figure 5. As in this case, digitized paper slips often provide precise pronunciation and various types of information about the name locality. Datasets like the Norwegian Place Name Lexicon and Norwegian Farm Names contain information about the origin of the name, earlier spellings, and pronunciation. It

is also possible to use historical forms for retrieving modern name forms. If we query for the historical name form ‘Myrløß’ – then Mørløs will appear as the headform of the query result.

Several of the datasets are digitized cadastres, where coordinates have been added and unique administrative codes enabling to link property from current and historical cadastres under same geographical names entry. Unique localization codes occur as GNIDu (first cadastral level) or MIDu (second cadastral level). Clicking on the name form for the farm Mørløs (Mørlaus) in the 1838 cadastre provides information about the farm name and the holdings under the farm. It is also possible to check against the original cadastre by clicking the ‘Scanned cadastre’ button and see the scanned cadastre with its Gothic script, or access the same information in the ‘Digital cadastre' digitized by the University of Tromsø. It’s not always easy to navigate a cadastre through a regular search, so a hierarchical register view of data is also supplied and accessible via the ‘Datasets’ tab and selecting, for example, the 1838 Cadastre. In the top right corner of the search field, you can select the ‘Hierarchy’ button and get a list with the cadastre’s highest division – county.

Not all datasets are ‘dead’, frozen data. For instance, the geographical names resource like the Norsk Stadnamnleksikon (Norwegian Place Name Lexicon) and the coming Kvenske stedsnavn (Kven Place-Names), for example, are living datasets that will be continuously updated in the coming years.

Supporting Geographical Names Standardization by Design

Integration and cross-referencing capabilities play a vital role in supporting standardization practices. The system implements UUID-based identification (Figure 6) while maintaining connections to external reference systems such as Wikidata for administrative units and Norwegian Mapping Agency SOSI codes for geographical feature types. This comprehensive linking enables users to understand geographical names within their full cultural and historical context, supporting evidence-based decision-making in standardization processes. When specific entities require dedicated search interfaces, separate indices are created, as exemplified by the base word index from the Norwegian Place Name Lexicon.

Media asset management is handled through an IIIF-compliant image server utilizing IIPImage technology. This component includes manual file upload procedures and automatic conversion to pyramid TIFF format, ensuring optimal delivery of visual resources. The system provides standardized image access protocols, facilitating integration with various client applications. The portal’s search capabilities represent a particular strength of the infrastructure. Users can perform detailed searches with multiple filtering options, explore geographical names through an interactive map interface, and access connected archival records and datasets. This functionality is supported by RESTful API services that enable efficient data retrieval and display.

The aim is that all data of the Stadnamnportalen are open data. All datasets so far are published under the open licence CC-BY 4.0. This means that data can be freely downloaded in various formats, such as CSV, JSON and GeoJSON, see Figure 6. An additional benefit of UUIDs is that they enable landing pages for every geographical name entry (Figure 7) as well as every source form entry.

The Future for Stadnamnportalen

Looking toward the future, the development plans of the Stadnamnportalen include emphasizing expanding its role in supporting inclusive standardization practices. Planned enhancements include implementation of data download capabilities, API access for programmatic data retrieval, and expanded multilingual support. The Language Collections is cooperating with the NRK, the national Norwegian broadcasting cooperation, on providing API services to one another. It is expected that these developments will further strengthen the platform’s ability to serve diverse user groups.

The technical challenges encountered, and solutions developed during the implementation of Stadnamnportalen, offer valuable lessons for geographical names management systems. The process of standardizing diverse historical datasets while preserving cultural information required careful attention to both technical, historical and cultural considerations. Solutions included the implementation of unique geographical name expression codes and location codes, the addition of coordinates to historical records. It was necessary also to implement unique administrative codes to keep control of

administrative changes through time and linking to source materials, be they scanned works, digitized paper slips or digital resources.

Stadnamnportalen demonstrates how modern technical infrastructure can support inclusive, culturally-informed geographical names standardization while providing evidence-based tools for sustainable development. The system's architecture successfully balances the needs of standardization with cultural heritage preservation, offering valuable insights for similar initiatives worldwide. For geographical names authorities considering digital transformation of their resources, Stadnamnportalen provides a model of how an evidence-based technical infrastructure can support national standardization practices. Its combination of robust automation, user-friendly interfaces, and comprehensive data management, coupled with careful attention to cultural preservation, offers a blueprint for developing systems that serve both administrative needs and cultural heritage preservation.

The ongoing development and expansion of Stadnamnportalen ensures its continued relevance for both administrative purposes and cultural heritage preservation. The project has undergone several developments since 2018, at first experimenting with a linked data system, to end up with the current technical infrastructure since late 2023. Since the launch of in June 2024, the front end has been overhauled to achieve a better user experience for desktop and mobile use alike, as well as complying to WCAG-guidelines for Universal Design, as is required by Norwegian law. The latest frontend implementation is from January 2025 and differs from the design demonstrated for the 68. UNGEGN Information Bulletin from November 2024.

By implementing a flexible and easily developed and scaled technical infrastructure, Stadnamnportalen successfully combines robust data management with user-friendly access to Norway's geographical names heritage. The portal's architecture demonstrates how modern web technologies and efficient data handling can be applied to preserve, promote and present evidence-based cultural heritage information, making it accessible to researchers, administrators and the public alike.

The success of this infrastructure provides valuable insights for other geographical names management organizations considering digital transformation of their resources. Its combination of standardization and flexibility, coupled with robust automation and user-friendly interfaces, offers a model for similar initiatives in other contexts.

Useful links:

Stadnamnportalen: https://stadnamnportalen.uib.no/

Gammeltoft, P. 2024. 'The Geographical Names Portal Stadnamnportalen (Norway)', United Nations Group of Experts on Geographical Names (UNGEGN) - Information Bulletin, No. 68: The role of geographical names in preserving cultural heritage. United Nations Group of Experts on Geographical Names Information Bulletin, vol 68, 69-70.

Gammeltoft, P. 2024. 'Stadnamnportalen - millionar av namn i tid og rom'. På Høyden - nettavis for Universitetet i Bergen 2024.

Gammeltoft, P. 2023. 'Place-Name Databases - A Spatio-Temporal Mess'. Digital Spatial Infrastructures and Worldviews in Pre-Modern Societies'. Arc Humanities Press (Arc), 137-157.

Points for discussion

The Group of Experts is invited to:

(1) Consider how flexible technical infrastructures can support geographical names standardization while preserving cultural heritage and accommodating diverse datasets.

(2) Discuss approaches for integrating historical records, phonetic/phonemic renderings, sound files, and modern geographical names data using unique identifiers and cross-referencing systems for evidence-based standardization.

(3) Explore how digital platforms can serve multiple user groups – including researchers, administrators, and the public – while maintaining robust data management practices for geographical names.

Illustrations:

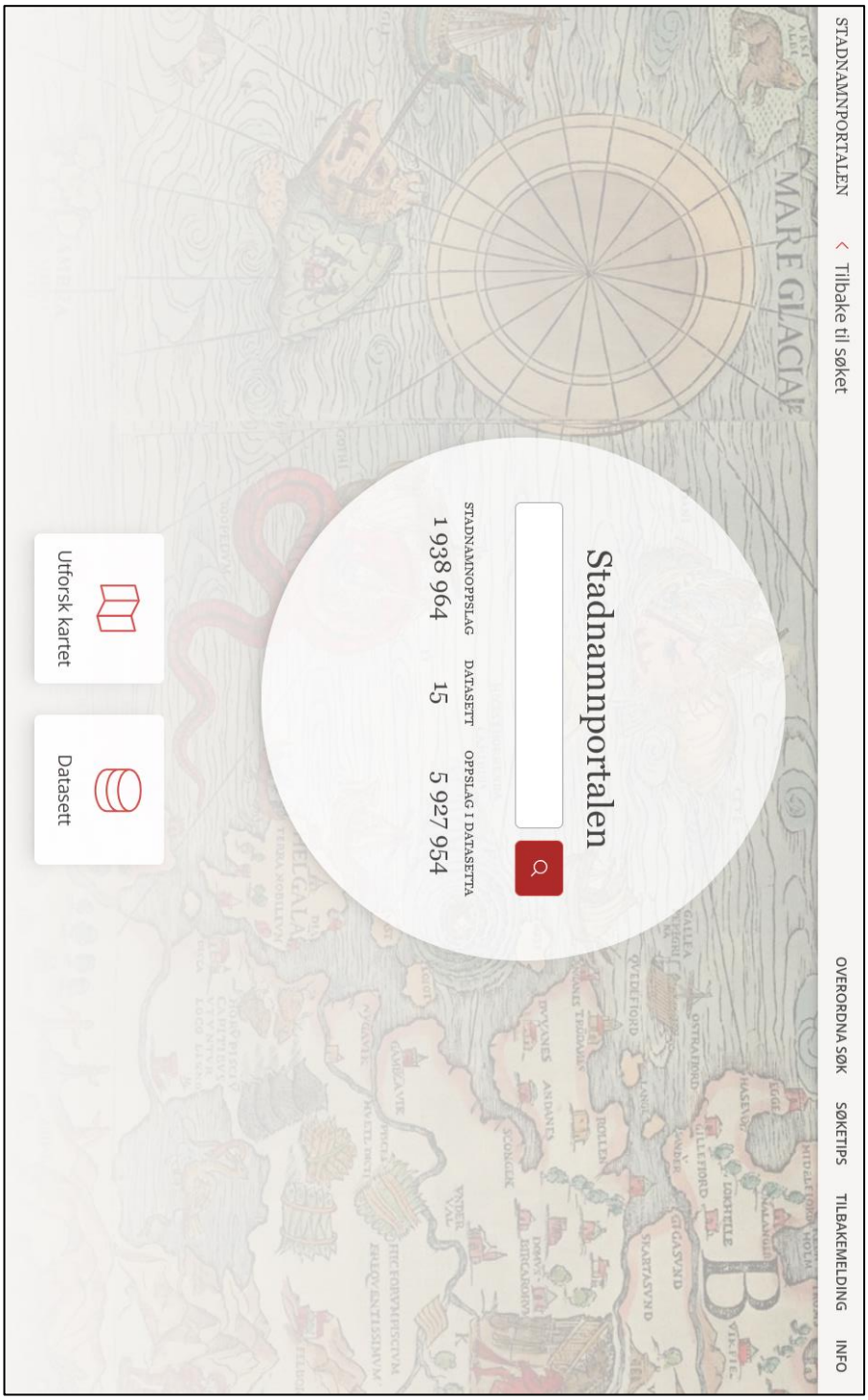

Figure 1: Front page of the new iteration of the Stadnamnportalen (March 2025). The webpage is Universal Design-compliant and has great flexibility in accessing data, such as a general query, a map query entry and queries in individual datasets.

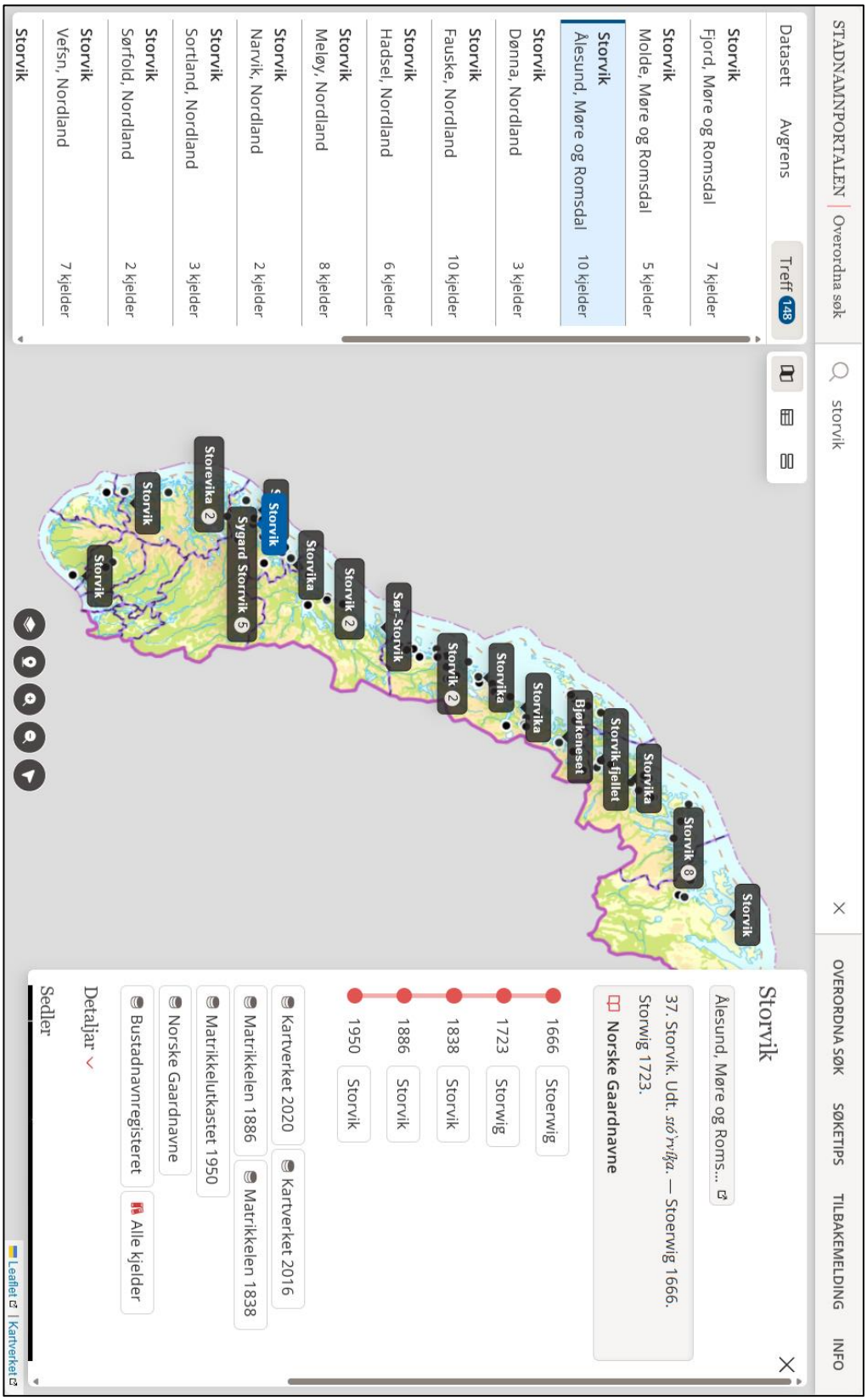

Figure 2: A map-based query in all data for the geographical name ‘Storvik’. By clicking one of the results in the left window, metadata and research data becomes available in the right window, here *Storvik* in Ålesund, Møre og Romsdal.

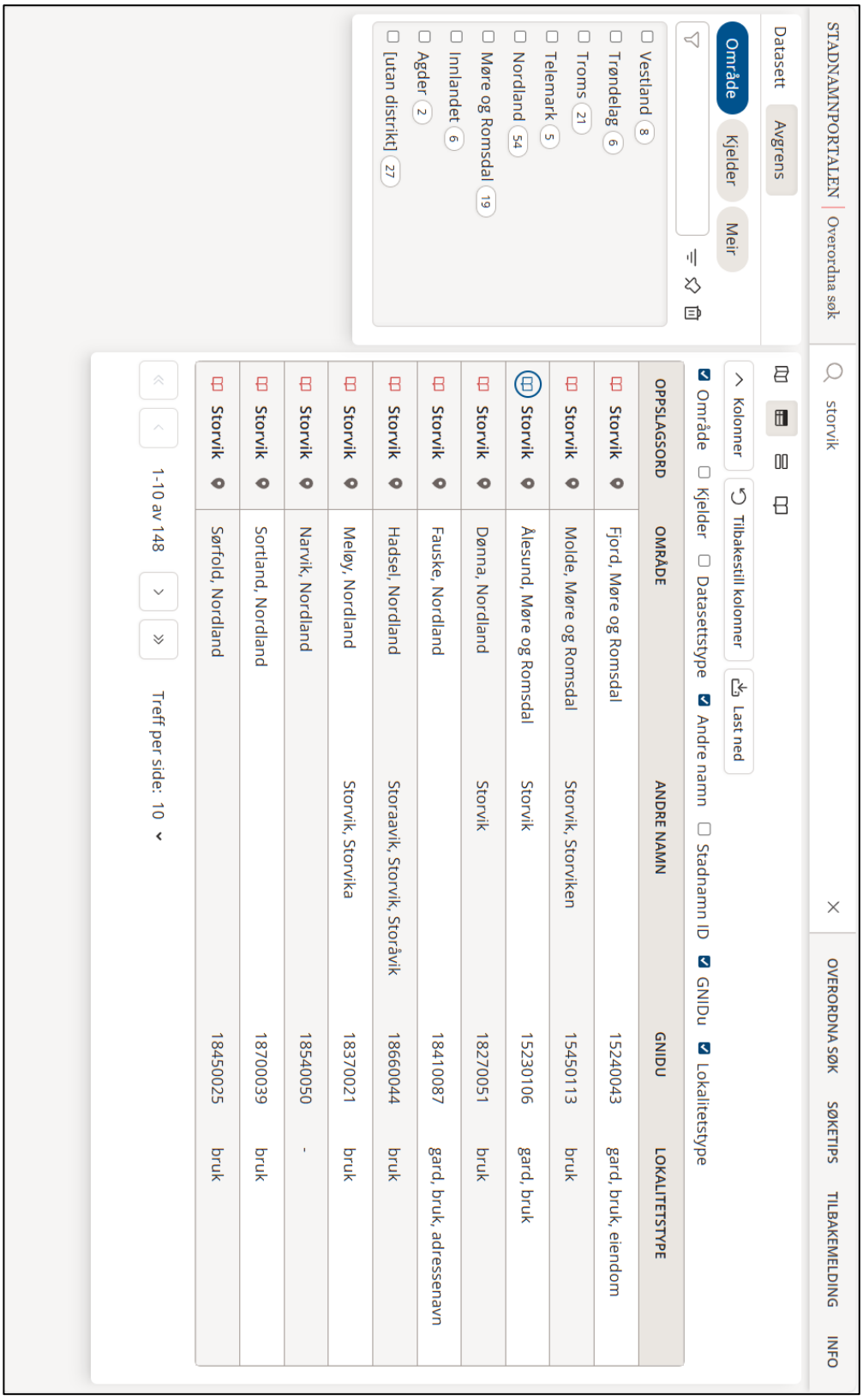

Figure 3: A tabular view of the same query result as in Figure 2. Note also that in this illustration the powerful filtering possibilities have been activated by clicking the button ‘Avgrens’ (= ‘Filter’). Here we can see the modern regions with examples of the query ‘Storvik’.

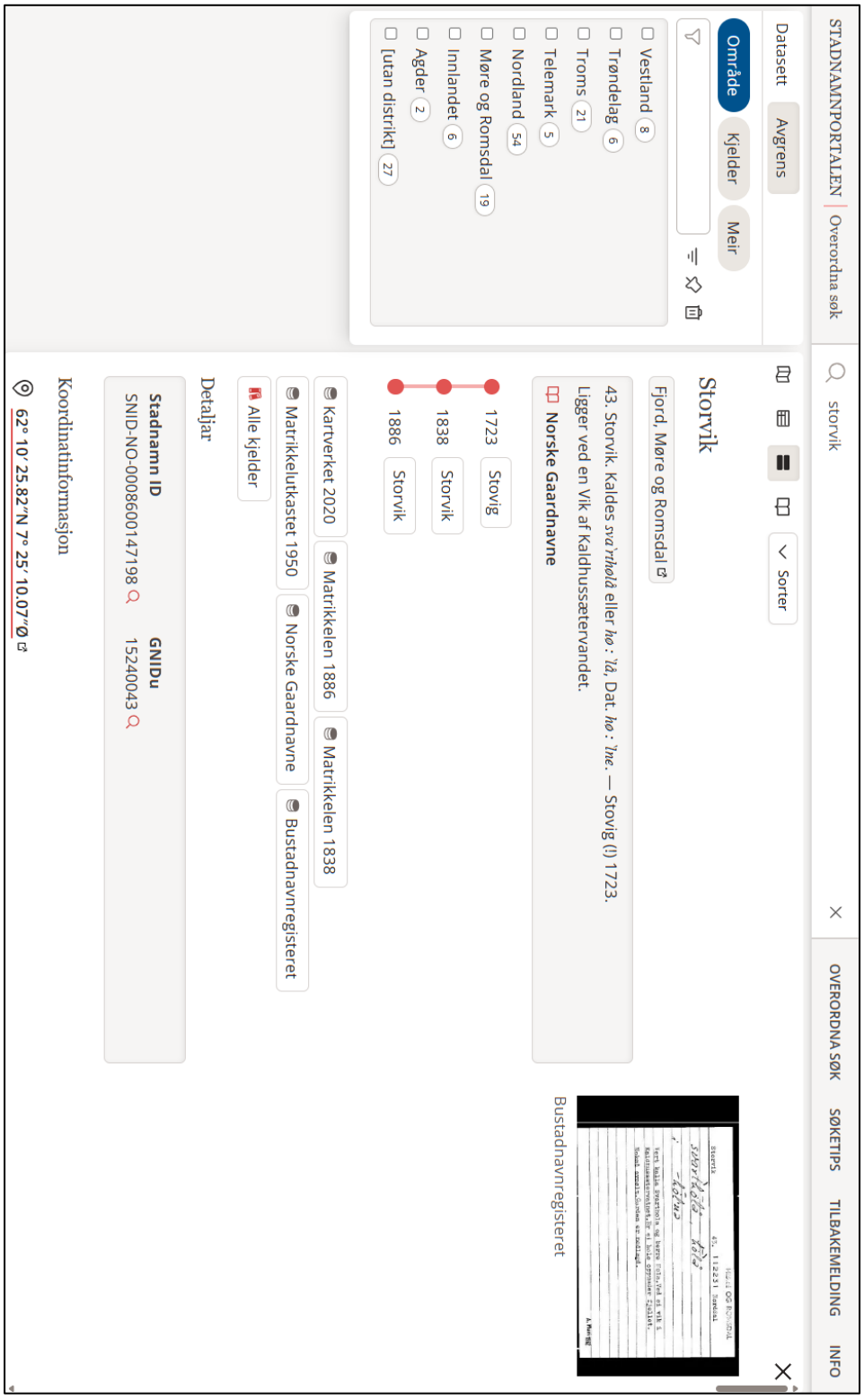

Figure 4: A list view of the same query result as in Figures 2-3. Again the filtering possibilities have been activated by clicking the button ‘Avgrens’ (= ‘Filter’). In this view we see interpretations, historical forms, accessible datasets and index cards, as well as metadata and coordinate information.

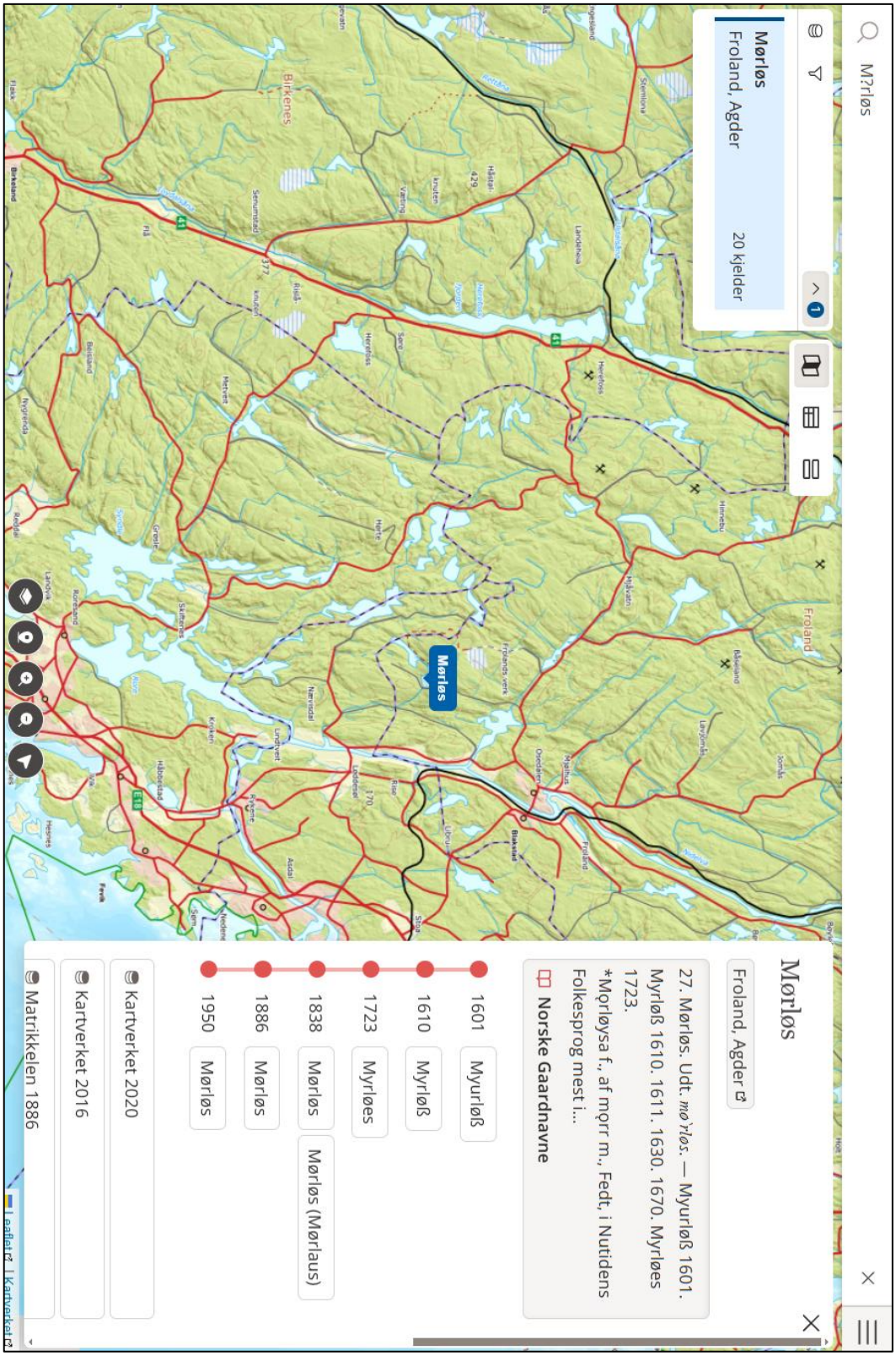

Figure 5: Map view of a so-called single letter wildcard query for ‘M?rløs’, showing the powerful search functionality of Stadnamnportalen.

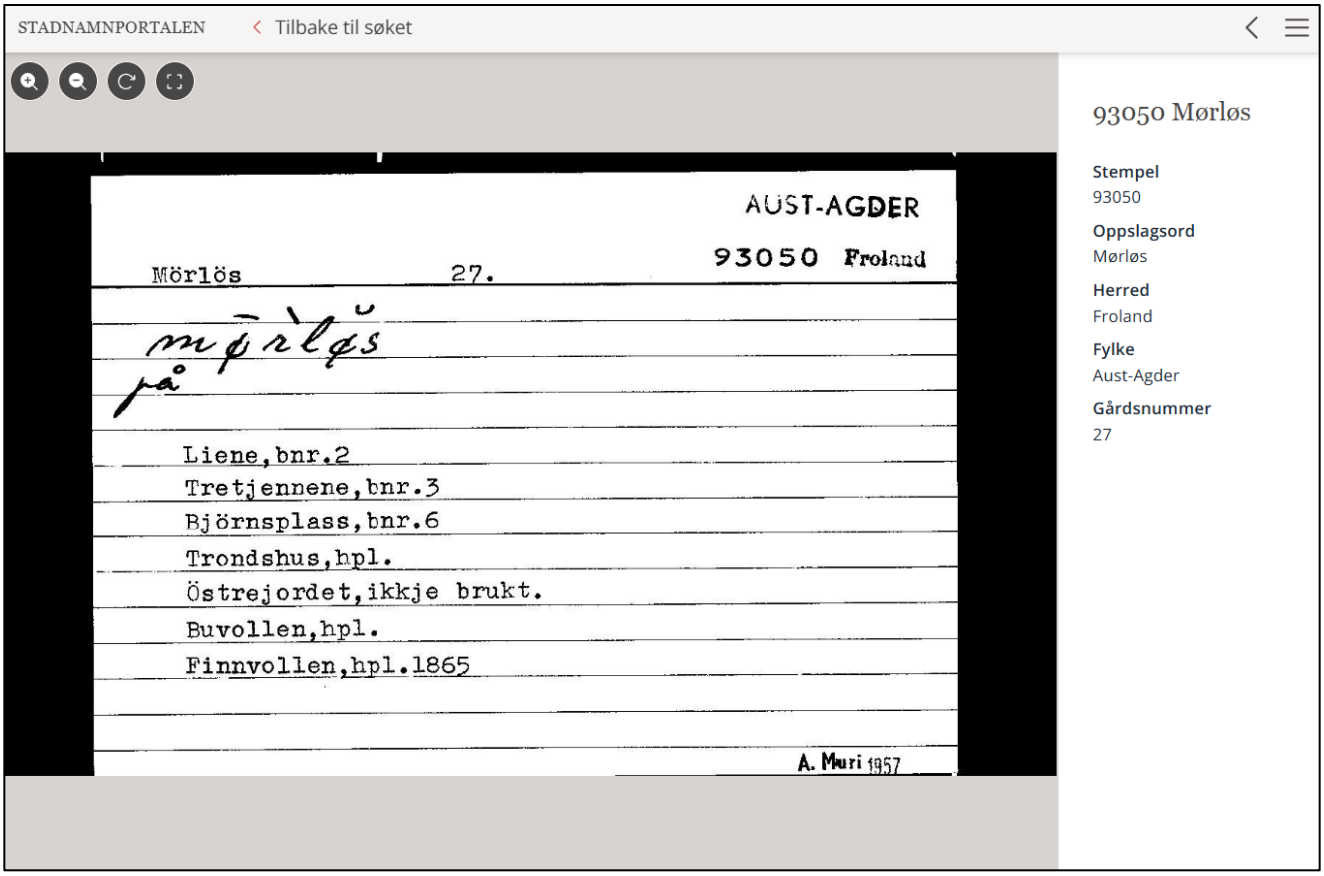

Figure 5a: Example of a IIIF-image showing the available index card for the geographical name of Mørløs. Notice the phonetic rendering in the national phonetic standard *Norvegia*. In addition, the index card also supplies information on proper pronoun use as well as sub-units under the settlement Mørløs.

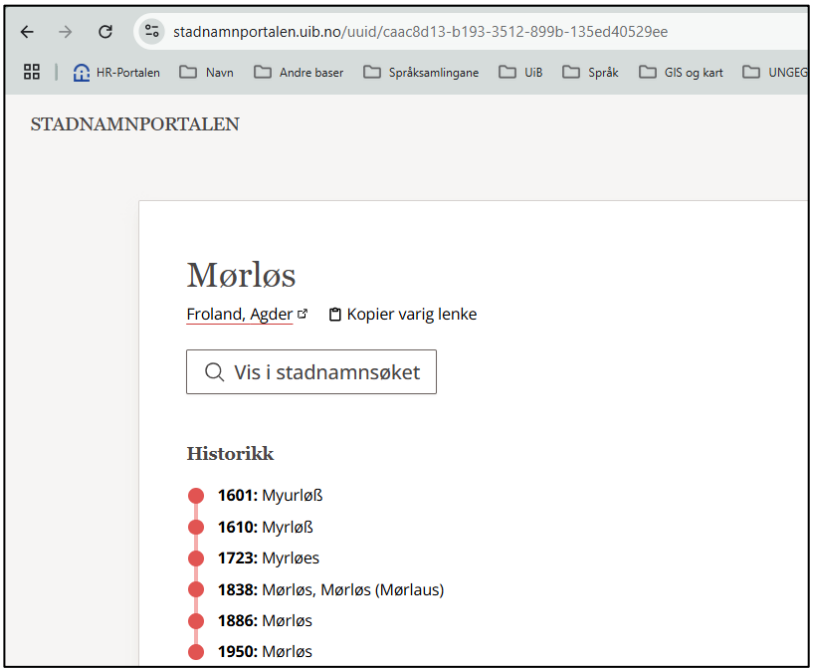

Figure 6: Example of UUID view of a geographical name entry.

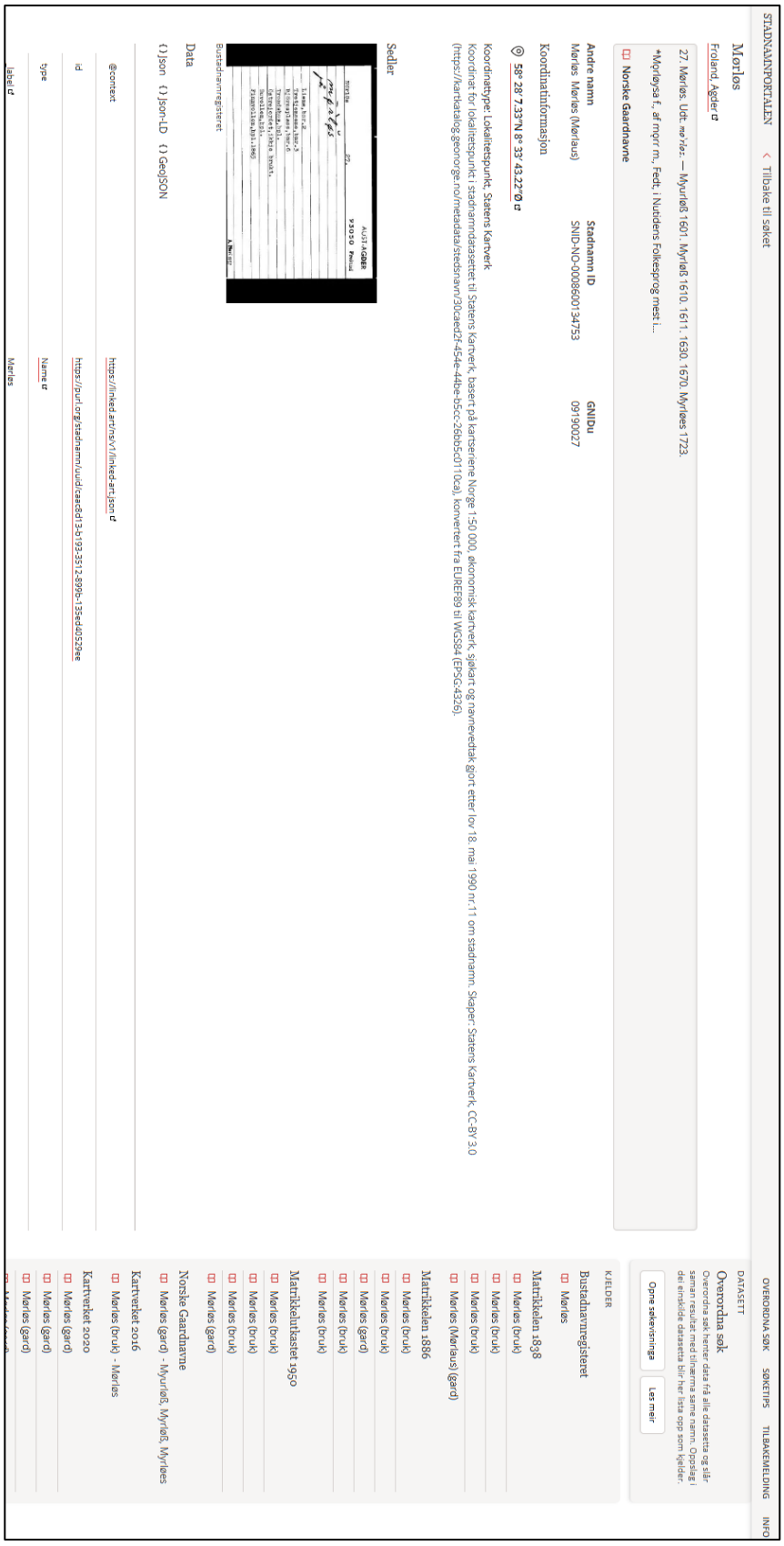

Figure 7: Landing page for the geographical name entry ‘Mørløs’. A landing page is a UUID standalone web page specifically designed to capture data information through a form and provide machine-readable interaction possibilities.