Norwegian Communications Authority

BEACON

Broadband Expansion through Address Clustering and Optimization Network - Method description

Report nr. 2503527-3

18 June 2025

Table of contents

1 Introduction

The Norwegian Communications Authority (Nkom) has developed a geographic information system model called BEACON (Broadband Expansion through Address Clustering and Optimization Network) to calculate cost estimates related to the development of broadband infrastructure in Norway. The model is a tool for assessing the economic consequences of political objectives for broadband coverage, especially with regard to the goal of gigabit speed for all by 2030.

BEACON is based on extensive register data, including address data for households, holiday homes and businesses, as well as infrastructure data such as road networks and existing poles that can be used as conduits. The model also uses detailed coverage data that shows which addresses already have access to different broadband technologies.

Since fibre is in practice the technology that is currently being developed to achieve gigabit speed, BEACON focuses on estimating the necessary fibre development to addresses that lack this. The model uses advanced geographical algorithms for clustering addresses and optimizing fibre stretches. This gives an estimate of the total number of meters of fibre that must be laid to reach the addresses in question.

For the cost calculation, the model takes into account several factors, including population density in the development areas, distance to the nearest available fibre connection and the need for new nodes. Although fibre in Norway is being developed with both GPON and P2P technology, Nkom considers that the main cost lies in the establishment of the actual cable routes (digging, trenching, overhead stretches and submarine cables). Therefore, the difference between these technologies is as of now considered negligible in the overall cost estimate.

BEACON has been developed as a modular and flexible solution, which allows for the incorporation of new data sources as they become available. This methodology document will describe the various components of the model, with particular emphasis on the algorithms and methods used to produce the cost estimates.

2 Model architecture

The architecture of the BEACON model is divided into modules to facilitate both conceptualization and flexibility. Below is a general description of the main modules that make up the model and how they work together. The BEACON model consists of four main modules:

1. **Input data**: This module represents all data sets used as a basis for the model's calculations. This includes:

a. Address register (all addresses in Norway + cadastral data on buildings)
b. Household data (addresses with persons registered in the National Population
c. Register)
d. Enterprise data (all businesses at address level)
e. Coverage data (coverage and subscriptions, at the address level)
f. Node location (Coordinates of the largest operators' fibre nodes)
g. Infrastructure data (road network, poles and other conveyance options)
h. Cost variables for development (taking geographical conditions into account)

2. **Geographic analysis**: In this module, geographic characteristics are analyzed based on the input data. This includes:
    a. Characterization of coverage areas
    b. Clustering of relevant addresses
    c. Mapping of possible cable routes
    d. Identifying favorable location for new nodes
    e. Calculation of distance to the nearest available fibre connection
3. **Indicator calculation**: This module performs the actual calculations of:
    a. Total fibre length that needs to be laid
    b. Distribution between different types of conduit routes based on the characteristics of the area
    c. Number of new nodes that need to be established
4. **Result generator**: The last module compiles the calculations into specific cost estimates:
    a. Total costs for fibre development
    b. Costs by geographical area
    c. Costs per address type (Household, business, possibly holiday homes...)
    d. Result export (to Excel, csv, PowerBI, ...)

The model works sequentially where each module builds on the results from previous modules. The input data forms the basis for the geographical analysis, which in turn provides input for the calculation of key indicators. These calculations ultimately lead to specific cost estimates and other relevant results.
The modular structure is designed to adapt to changes in:

- Data basis (when new or improved datasets become available)
- Geographic algorithms (for continuous improvement of precision)
- Cost parameters (to reflect changes in market prices)

- Policy objectives (to analyze different scenarios)

A key strength of the BEACON model is its ability to simulate different development scenarios based on varying assumptions. This makes it a valuable tool for estimating the costs associated with various policy objectives for broadband development in Norway.

3 Input data

The geographical analysis developed in the BEACON model is data-intensive and is based on extensive data from several sources. This data basis has been built up through multiple data collection processes, including publicly available information and specific datasets to which Nkom has access.

Several databases have been used for the development of the geographical analysis. The main files used in the model are:

Data	Description	Source	Accessibility
Address database	Dataset containing all addresses in Norway including cadastral data	The Norwegian Mapping Authority	Publicly available
Household database	Overview of addresses with registered persons	National Population Register/Statistics Norway	Limited availability
Business database	All businesses at address level	The Brønnøysund Register Centre	Publicly available
Coverage data	Files that contain broadband coverage for each address	Providers/Nkom	Not publicly available
Node locations	Geographical coordinates of the largest providers' fibre nodes	Providers/Nkom	Not publicly available
Road database	GIS file with all information related to the national road network	The Norwegian Mapping Authority	Publicly available

The data sources used represent the most important data for the geographic analysis, and their content therefore has a direct impact on the model's characterization of the network and the results generated.

The first step in the geographic analysis is the preparation and adaptation of the raw data to the specific needs of the model. To this end, a data cleansing process has been established with the following main objectives:

- Adjust inaccurate information
- Remove unnecessary information
- Aggregate data depending on the needs of the analysis

This data cleansing procedure is applied to all the data sources before they are applied to the model. The following sections elaborate on the content and activities performed for each of the databases included in the analysis.

3.1 Directory

The address register is henceforth a compilation of the address database, the household database and the business database. In connection with coverage mapping, we have carried out a linking and cleaning of this data, so that we now have an address register that contains the following information:

- AddressId
- Municipalities
- Name of municipality
- County name
- Er_i_tettsted (located in a settlement where at least 200 persons live in houses not more than 50m from each other)
- Number of households
- Number of people
- Number of enterprises
- Number of holiday houses at the address
- Street Names
- House No.
- Letter
- Postal code
- City
- Farm number
- Use number
- Attachment number
- Subheading
- East (Geocoordinate in UTM zone 33N)
- North (Geocoordinate in UTM zone 33N)

3.2 Coverage data

Data from Nkom's coverage survey. This is an annual survey that collects data from all providers with their own infrastructure or have broadband customers based on other people's infrastructure. The data we have available from this survey is the following at the address level:

- Offers
- Technology
- Maximum download speed
- Maximum upload speed
- Owner of infrastructure
- Homes Connected (is the building connected?)

We also have an overview of actual subscriptions at address level.

3.3 Node locations

In connection with collecting data for the annual coverage survey, we also collect the position of the providers' fibre nodes. It also contains the link to which addresses are connected to the different nodes.

3.4 Infrastructure data

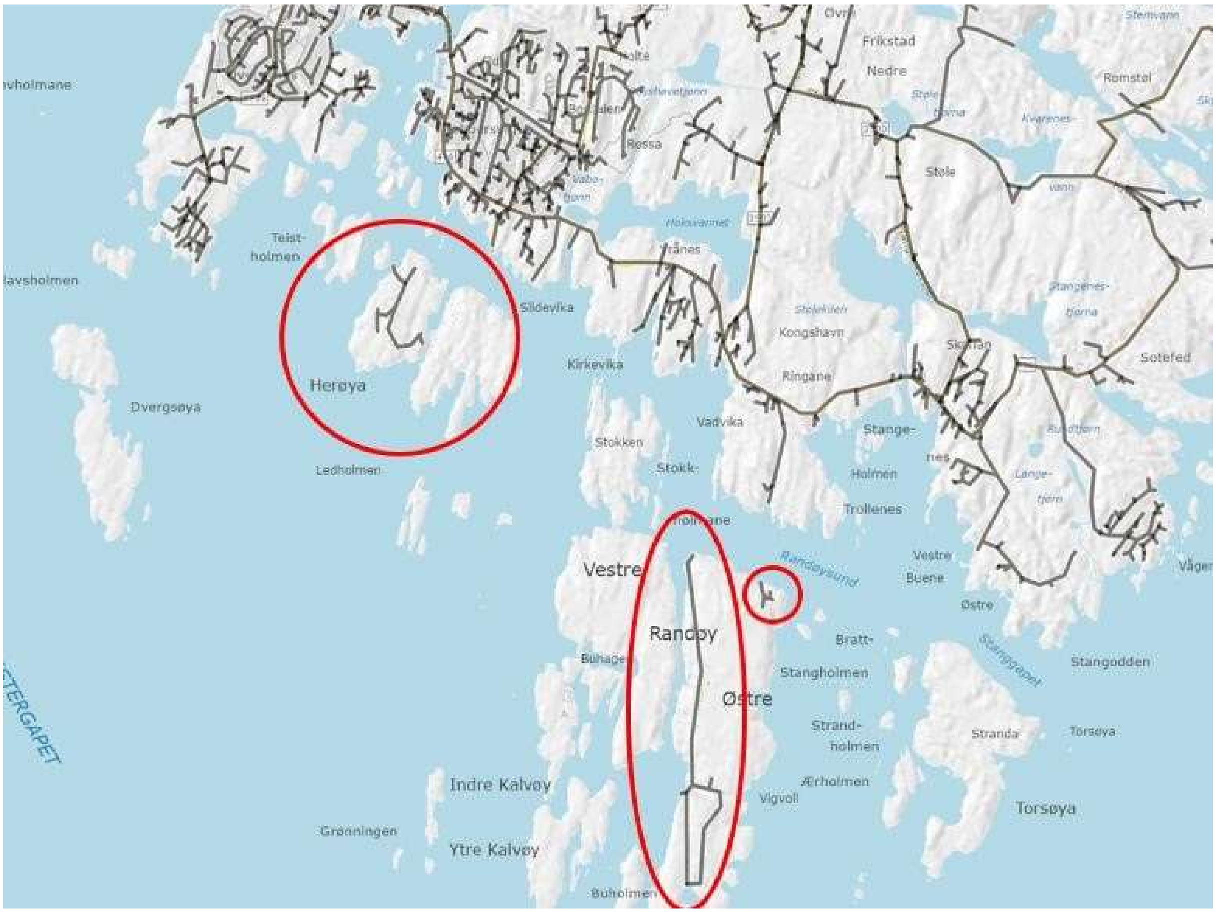
In the current version of BEACON, we have a complete mapping of Norway's entire road network as well as all available poles and masts. Although the geographical analyses are primarily based on road network data, the model has been constructed with the flexibility to be able to include poles for cable conveyance where appropriate.
The model's architecture is designed for seamless expansion with additional infrastructure data, such as submarine cables or alternative routing routes, without the need for substantial changes to the underlying geographic algorithms. This gives BEACON significant scalability – new data sources can be incorporated directly, which will increase the level of precision in the analyses without changing the fundamental methodology.
Now we will take a closer look at the characteristics of the road network, as well as how we have prepared this dataset for further analysis.

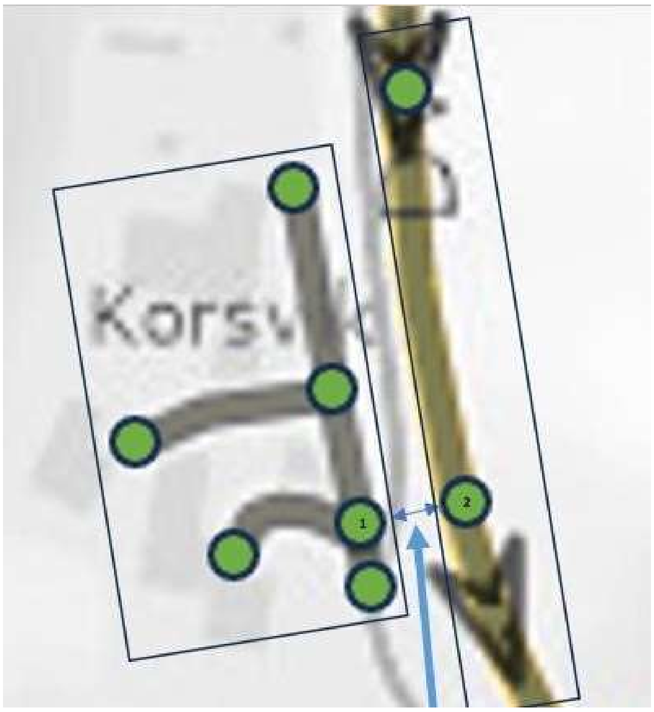
1. We retrieve the road network data as csv files from vegkart.atlas.vegvesen.no
    a. We select all drivable roads
    b. We also pick up bike paths that are possible to drive on (driving to properties, or transporting goods

The road network data consists of about 2 million road links that are defined with a start node, end node, road type and geometry (Linestring) in wkt format.

2. The first processing done is to clean the data of zero values (there are a very few of these) and connect the drivable road. In this process, we give a new "integer name" to the nodes, and make sure that we do not get duplicates. The nodes that are the start or end nodes of a road link are hereinafter referred to as k-nodes.
3. After processing, we identified that the road network consists of just over 3000 separate partitions – independent network segments with no direct connection to the main road network. To optimize the model's accuracy, we have implemented a solution where we establish virtual path links between partitions located within a distance of 50 meters from each other. We reduce the number of partitions to just under 1000 after this is done. This is probably pretty close to the correct number of actual partitions (typically islands without a bridge connection).

   This methodological approach is based on a pragmatic assumption that fibre development over such short distances normally represents minimal technical and economic challenges in practice. The virtual links ensure a more realistic network topology by reflecting actual development opportunities, while significantly increasing the model's ability to form continuous fibre segments in the geographic analysis. See the pictures on the next page for illustration.



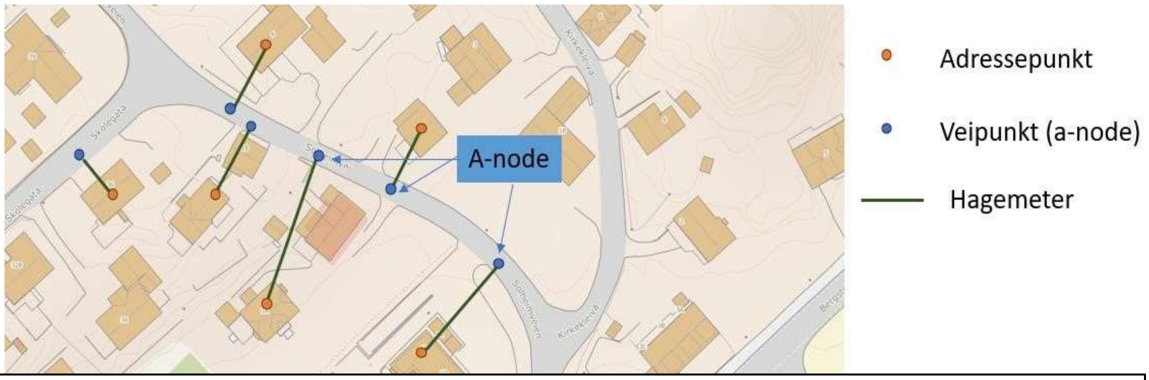


A new link will be modelled from k-node “1” to k-node “2” if distance is less than 50 m.

4. The next step is to integrate the address points with the road network. This is done by connecting each address to the nearest road using an orthogonal projection – i.e. a

line that represents the shortest distance from the address point to the road. We denote these projection points a-nodes. The result is a network of interconnected nodes, which

are divided into two types: k-nodes and a-nodes. In order to use algorithms from established graph theory (network theory) in the further calculations, we must now determine the distance between all connected nodes in the network.



Address point Road connection point Private property conveyance

The line from the address point to the road is called garden meter. We count an address as HP on fibre if the associated a-node is fibre covered. If there is also a drop cable from the road to the house, then the address is HC.

5. What we then do is connect the address data to the road network. We name all the anodes based on whether there are households, businesses or holiday homes associated with the a-node. All k-nodes now have integer names from 1 – 2,000,000, a-nodes with households are in the interval 10,000,000 – 20,000,000, businesses from 20 – 30 million, holiday homes from 30 – 40 million, and other addresses have names higher than 40 million. We also do a small analysis around each a-node to check how many addresses are in a circumference of 100 meters around the a-node. This will be a density concept that we can use at the single-node level later in the analyses.

3.5 Cost variables for deployment

Cost variables for fibre cable deployment are based on aggregated and standardized costs for different kinds of deployments: Aerial, buried, ducts/pipes etc., as well as other kinds of supporting infrastructure like cabinets.
Standardized values are also used for:

- which households can be expected to have commercial value and which the providers must be expected to develop without state aid, and

- which households are so expensive to develop that it must be expected that the households will not be developed without state aid.

For fibre based customer connection valuation, values of NOK 35,000 (for rural area connections) to 40,000 (for urban area connections) are used. The assumption is that a household or an enterprise has commercial value if the development cost is lower than the customer value. In this way, the customer value provides the basis for estimating both households of commercial value, and the need for state aid.

4 Geographic analysis

We mainly use the python API of igraph for network calculations in our geographical analyses, as this library is compiled in C++ and thus provides a significant performance advantage. In some cases, we also use networkx, as this python-based library offers several built-in algorithms that are not found in the igraph. This combination gives us both high performance and a wide range of algorithmic tools.
A significant part of the task going forward will be to find the most optimal – i.e. shortest – routes through the road network to reach our targeted a-nodes, for example the addresses that lack fibre. To achieve an economical and efficient route planning in an uncorrected graph with two node types, where we are only interested in the a-nodes, we use established theory from graph theory. This problem is known as the Steiner tree problem (https://en.wikipedia.org/wiki/Steiner_tree_problem).

In short, the Steiner tree problem is about finding a network (a tree) that connects all the necessary terminal nodes – in our case a-nodes – with a minimum total weight (shortest total length). This is done by also including extra nodes, so-called Steiner nodes, if this leads to a lower overall cost or shorter total distance. The implementation of such algorithms often involves an iterative approach, first estimating a preliminary route plan and then fine-tuning by including alternative routes and any additional nodes to minimize the overall distance. Although the Steiner tree problem is NP-hard, there are effective approximation algorithms that provide solutions with sufficient accuracy for practical use.
By using these methods, we can optimize the network design for fibre development, reduce material costs and ensure that the infrastructure is established in the most cost-effective way possible.

4.1 Preparing the network data

The first step in the preparation of the network is to construct a new network consisting only of potential terminal nodes, i.e. a-nodes. We call this network the anode graph. This is crucial in terms of performance, as we originally have a network of over 4 million nodes and even more edges.
To achieve this, we use an efficient algorithm – we call it the Steiner subgraph method – that removes all the k-nodes while establishing new edges between the remaining a-nodes, so that we retain the necessary association with updated distance measurements. The method is based on the following:

1. First, we create a subgraph consisting entirely of the k-nodes (Steiner nodes).
2. Next, we identify all partitions in this subgraph.
3. For each partition, we connect all the a-nodes that are neighbours to the k-node partition and calculate the shortest distance between them using the Dijkstra algorithm.

The result is a graph that contains only a-nodes but still retains all the original connections with correct distance measurements. Due to this partitioning, we are able to solve the problem in just a few minutes, even with a normal laptop.

4.2 Calculation of length of fibre cable segments required

We are approaching one of the core tasks of the BEACON method: to calculate the total fibre length that must be deployed to achieve coverage for the desired addresses, such as any that lack fibre coverage. This metric includes two types of coverage:

- **HP coverage:** Here, an address is assumed to be covered if the fibre cable reaches the associated a-node. In other words, all addresses associated with an a-node with a fibre cable connected will receive HP coverage.
- **HC coverage:** For an address to be considered HC covered, the fibre cable must be connected directly to the building. This means that we need to add the extra length, previously referred to as "yard meter", which makes up the distance from the a-node to the address point itself.

The process starts by making a list of all addresses which are not covered by fibre, and then we identify the corresponding a-nodes. This is where the hierarchical cluster structure we established earlier (e.g. c4, c5, c6, ..., c12) comes into use **Procedure for the fibre deployment:**

1. **Identification of fibre-free trees:**

The fibre is initially deployed from the nearest fibre-covered a-node. This gives rise to several "trees" that branch out from each fibre point. Our challenge is to find the nearest fibre-covered a-node for each of these trees.

2. **Utilization of the cluster structure:**
    - We start by finding the largest cluster (e.g. c6) that consists exclusively of anodes without fibre coverage.
    - Since this cluster contains no fibre points, we know that the closest fibrecovered a-node must be in the next, larger cluster – for example, c7, which includes the c6 cluster in question.
    - Using Dijkstra's algorithm, we calculate the distance from each a-node in the fibre-free cluster (c6) to the nearest fibre-covered point in the cluster above (c7).
3. **Steiner tree approximation**
   After we have identified for each a-node without fibre coverage the closest fibre point, we group these nodes together with the identified fibre points to calculate the total fibre length. This is done by:
    - To apply the a-node graph to the relevant a-nodes. o Run a minimal spanning tree (MST) algorithm on each "tree structure" to find the shortest route connecting all the nodes.
    - Finally, we integrate the original k-nodes back into the calculation to ensure that we get the correct and shortest route for the entire tree.

Through this process, we estimate both the fibre length needed for the HP coverage (up to the anode) and for the HC coverage (where we add "garden meters" from the a-node to the address). This combined approach ensures an accurate estimate of the physical costs of the fibre deployment and lays the foundation for further cost analysis in the BEACON method.

5 Cost modelling

Now we are focusing on estimating the actual investments related to the fibre development. Algorithmically, the calculation is relatively simple, but before we can get started, a number of assumptions and adjustments must be in place.
The biggest cost driver in the deployment is the establishment of the fibre cable path/conveyance. This performance consists of a mix of different installation methods, including:

- **Aerial installation:** By far the cheapest method, where fibre is laid in the air, often using existing masts and poles.

- **Digging in soil:** A method that involves laying fibre in trenches, which is usually less expensive than digging on paved areas.
- **Digging in paved areas:** This method is often more expensive, not only because of the digging itself, but also because the solution for crossing roads – especially when addresses are on both sides – requires extra measures and therefore a special adjustment factor.
- **Other methods:** Other approaches can be used depending on local conditions and existing infrastructure. In Norway, rocky surfaces along roads are not uncommon.

For the total cost calculation, it is essential that we set an exact metre price for each type of deployment, as the distribution of these will have a major impact on the final estimate. In addition, we need to calculate the number of new fibre nodes that need to be established to reach all addresses. These nodes, which cost approx. NOK 500,000 each, are a significant cost item that must be included.

It is also important to emphasise that we calculate the costs based on the most optimal deployment method – an ideal solution that is often not feasible in practice. For example, a provider may face higher costs if he does not own the nearest fibre node, forcing alternative and more expensive solutions.

We have therefore implemented calculations for several scenarios, including:

- **Rural areas:** Estimates the cost when using only a particular method, such as aerial stretching only or digging only, for the more remote areas.
- **Mixed method:** A combination of the different types of deployment - which is considered the most likely in practice.

Furthermore, there is considerable uncertainty associated with the price of the various network routes. The prices we use in the model will be cross-checked with relevant providers to ensure that the estimates are close to the actual cost levels. A particular challenge is the cost difference on paved roads – where a simple ditch in the middle of the road is not possible, and development on both sides of the road entails additional costs. To deal with this, we use a correction factor that reflects the increased costs.

As part of the quality assurance, we have also agreed with several providers that we will send our estimates for selected areas, so that they can compare with real prices and provide feedback. This approach ensures that our models are continuously improved and that the practical conditions are reflected as accurately as possible.

In summary, the final cost calculation forms the basis for a model that balances theoretical optimality with practical realities, and which provides us with a solid basis for assessing both investment costs and financial consequence