



Pekka Kahri, Director of Information Services, THL
Risto Kaikkonen, Head of Unit, THL

26th January 2016

THL'S DATA POLICY

Making efficient use of THL's data

One of THL's strategy goals is to bring efficiency to the use of our datasets and to be a forerunner in the production, distribution and integration of data (*THL's strategy 2015*). THL's data policy, "Making efficient use of THL's data", is a tool promoting the efficient use of our datasets, designed for use by both the management and the people working with data. The data policy defines the priorities and measures that THL will employ for five years as of 2016 to promote the efficient use of our information resources.

The data policy was drawn up in collaboration between THL's staff and Executive Board in 2015. The work was co-ordinated by the Information Resource Services Unit.

The data policy consists of five outcome measures and one overarching measure. THL's data policy measures are:

1. We will make our datasets visible

The first condition for efficient use of data is that there is information, or metadata, available about the existing data. Having up-to-date metadata, variable data and availability data makes it easier to find, utilise and collect datasets and develop their quality.

We will launch a project where we produce descriptive data about our most important register- and population-study-based datasets. We will release metadata, variable data and availability data as continuously updated open data. We will adopt a nationally and internationally compatible data model, and we will link our dataset metadata, including all language versions, to electronic search engines.

We will gradually make the systematic description of our datasets and the release of our metadata an established practice at THL.

2. We will make our datasets more timely and relevant and improve our custom reporting. We will boost our data-collection co-ordination and avoid collecting overlapping data.

We will review the timeliness and relevance of our datasets with a bold and critical eye. We will make the most use of the data we have already collected and avoid collecting overlapping data. We will launch an information resources programme to develop our key data collections.

We will boldly adopt new methods in data collection, analysis and custom reporting.



We aim to achieve a high level of co-ordination in the data collections for population studies. We also want to make sure that most of the register data we use are retrieved directly from the National Archive of Health Information, from client and patient information systems and from other public authorities.

3. We will make open data and open publishing our priority

We will invest in open data, provide actively our datasets for use by others as well as interact with data users and developers.

We will favour open publishing where possible, and we will support and encourage THL's researchers and experts to use open publishing.

Our target is that the metadata, statistical data and research results from THL's key datasets are released as continuously updated open data. Processed data products designed for use by businesses and other actors in society will be subject to a fee, where appropriate.

4. We will increase the availability of restricted data

We feel that even confidential data and other restricted data should be available for use in appropriate settings in appropriate time, taking into account information security and data protection. We will adopt clear rules on the disclosure and ownership of datasets and create efficient support services to further the availability of restricted data, such as biobank collections and data collected from registers, surveys and population studies.

We aim to create a national-level one-stop shop for access authorisation and data disclosure, covering different kinds of datasets. We also aim to make our authorisation processes transparent as well as enable easy linking and sharing of datasets from various authorities through open interfaces.

5. We will develop electronic data lifecycle management and electronic storage solutions

We aim to safeguard the availability, usability and retention of THL's datasets by providing our staff with guidance and counselling on how to store datasets and define retention periods. We will emphasise the necessity of drawing up a data management plan at the start of each data lifecycle and support our staff in this task.

We will design an electronic data warehouse and an electronic archiving solution specifically tailored for THL's needs in our quest to safeguard the usability of our data even in future. We will archive our datasets for the prescribed retention period.

Overarching measure: We will co-operate

We will promote the efficient use of our datasets by increasing co-operation both internally in THL and externally with national and international actors. We will also reinforce our partnerships with public authorities, research institutes, universities, private enterprises, and research support infrastructures. We will make use of existing solutions and services when they meet our needs.