국외출장 결과보고서

사회정책 통계 인프라 구축 및 활용 관련
NSM(Nordic Statistical Meeting) 2019 주요 이슈 및 연구 동향

2019. 8.

KIHASA 한국보건사회연구원
Korea Institute for Health and Social Affairs

1 출장 개요

□ 출장목적

○ 데이터 연계 및 자료수집 방법론, 통계 인프라 구축 및 분석을 통한 정책적 활용, 정책 통계지표 개발 및 관리 등과 관련하여 스칸디나비아지역 국가의 통계청이 공동으로 3년마다 주최하는 통계 및 데이터 분야 전문가 Workshop(NSM, Nordic Statistical Meeting 2019)에 참석하여, 다양한 방법을 활용한 통계 인프라 구축 및 생산, 데이터 분석 및 정책적 활용사례 관련 최근 동향 발표자료 수집 및 인터뷰를 통해 포용국가 비전 및 전략 기반 사회정책 통계지표 개발 및 관리 연구수행에 활용함

□ 연구과제

○ 포용국가 비전 전략 기반 사회정책 통계지표 개발 연구
(보건복지부, 2019. 1 ~ 2019. 9)

□ 출장자 : 최 현 수 연구위원 (사회보장통계센터장)

□ 출장지역 및 기관 : 핀란드 Helsinki / Statistics Finland (Helsinki Marina Congress Center)

○ NSM(Nordic Statistical Meeting) 2019 주요 참여국가

☞ Statistics Finland (2019년 주최 기관), Statistics Denmark Statistics Iceland, Statistics Norway, Statistics Sweden, Statistics Estonia (신규 참여국가)

□ 출장기간 및 일정 : 2019. 8. 26 (월) ~ 8. 29 (목) (2박 4일)

출장일	행선지 (국가/도시)	방문기관/면담자	세부 활동내용
2019.8.26.(월)	인천 - 핀란드 Helsinki	Helsinki Marina Congress Center / Heli Korhonen, Kim Molin 외 NSM 발표자 및 참석자	출국(암스테르담 경유) / 스칸디나비아 국가 통계청이 공동으로 3년마다 주최하는 통계지표 및 데이터 전문가 Workshop (Nordic Statistical Meeting, NSM 2019) 참석 (1일차)
2019.8.27.(화)	핀란드 Helsinki	Helsinki Marina Congress Center / Marjo Bruun, Jaana Majalahti, Margret Vala Gylfadottir 외 NSM 발표자 및 참석자	스칸디나비아 국가 통계청이 공동으로 3년마다 주최하는 통계지표 및 데이터 전문가 Workshop (Nordic Statistical Meeting, NSM 2019) 참석 (2일차)
2019.8.28.(수)	핀란드 Helsinki	Helsinki Marina Congress Center / Indrek Onnik, Daniele Shamilowiic 외 NSM 발표자 및 참석자	스칸디나비아 국가 통계청이 공동으로 3년마다 주최하는 통계지표 및 데이터 전문가 Workshop (Nordic Statistical Meeting, NSM 2019) 참석 (3일차)
2019.8.29.(목)	핀란드 Helsinki - 인천		28일 저녁 귀국 / 29일 인천 도착 (프라하 경유)

2 주요 내용

가. NSM(Nordic Statistical Meeting) 2019 개요 및 프로그램주요 이슈 및 연구 동향개요 및 주요 프로그램

□ **데이터 연계 및 자료수집 방법론, 통계 인프라 구축 및 분석을 통한 정책 활용, 정책 통계지표 개발 및 관리 등과 관련하여 북유럽 국가 통계청이 공동으로 3년마다 개최하는 통계 및 데이터 분야 전문가 Workshop(NSM, Nordic Statistical Meeting)으로 2019년은 핀란드 통계청이 “Facts for Future"와 관련된 주제로 주최함**

□ **전체 20개 영역으로 구성되어 총 70여개 주제가 발표되었으며, 약 500명의 스칸디나비아 국가 통계청 및 다양한 정책 영역별 통계 관련 전문가가 참여하여 최근 이슈 및 연구 동향, 다양한 사례 등을 발표 및 토론함**

□ **Key Note & Panel Discussion**

○ OPENING WORDS ☞ Marjo Bruun (Director General of Statistics Finland)

○ KEYNOTE ☞ Tarja Halonen (President)

○ PANEL DISCUSSION – DATA, STATISTICAL LITERACY AND FACTS

☞ Esa Mäkinen (Data journalist, Helsingin Sanomat),
Kimmo Vehkalahti (University lecturer, University of Helsinki),
Tuuli Kaskinen (Executive advisor, Demos Helsinki)

○ KEYNOTE – DIGITALISATION IN THE PUBLIC SECTOR

☞ Siim Sikkut (Government CIO of Estonia)

○ KEYNOTE – AI AND MACHINE LEARNING IN OFFICIAL STATISTICS

☞ Samuel Kaski (Academy professor, Aalto University)


Facts for Future
nsm2019
Nordic Statistical Meeting
26.–28.8.2019

Monday, 26 August, 2019 / Lunch at 12.00–13.00*

*for short course participants only

AT GRAND MARINA / 11.00–12.00	AT GRAND MARINA / 15.30–16.30
REGISTRATION AND WELCOME	REGISTRATION AND WELCOME

13.00–16.30

SHORT COURSES

(Requires separate registration)

Machine Learning
GIS - Geographical Information Systems
Video Communication
Nowcasting

16.30–18.30

SMALL OFFICES SEMINAR

19.00–20.30

COCKTAIL RECEPTION AT HELSINKI CITY HALL

A warm welcome to the Nordic Statistical Meeting!

The theme of the conference is Facts for Future. Our statistical community must promote fact-based reasoning, debate and decision-making. Reliable statistical information and indicators can guide individuals and societies in building a better and sustainable world.

Our program is full of these relevant topics and we are excited to have high-level professionals as speakers. Academy professor Samuel Kaski will hold a keynote about AI and Machine Learning in Official Statistics, and Government CIO of Estonia Siim Sikkut about Digitalisation in the Public Sector Government. President Tarja Halonen will share her experience of the Sustainable Development Goals and the importance of measuring them. During parallel sessions we dig deeper into themes like new statistical methods and tools, data quality and the changing roles of National Statistical Institutions.

I truly hope you enjoy the conference and program, participate actively and leave Helsinki with new ideas, memorable experiences and a broader network.

Marjo Bruun
Director General of Statistics Finland

Tuesday, 27 August, 2019 / Lunch at 12.00–13.00

PLENARY / 9.15–12.00

OPENING WORDS

Annika Damström, Moderator
Marjo Bruun, Director General of Statistics Finland

KEYNOTE – AI AND MACHINE LEARNING IN OFFICIAL STATISTICS

Academy professor **Samuel Kaski**, Aalto University

PANEL DISCUSSION – DATA, STATISTICAL LITERACY AND FACTS

Executive advisor **Tuuli Kaskinen**, Demos Helsinki
Data journalist **Esa Mäkinen**, Helsingin Sanomat
University lecturer **Kimmo Vehkalahti**, University of Helsinki

> “I hope to learn something completely new and surprising. As an aspiring amateur athlete, I will also take part in the morning run!
>
> AURA PASILA, SOCIAL STATISTICS

FENNIA / 13.00–13.50

METHODS TO IMPROVE QUALITY OF QUESTIONNAIRE-BASED DATA COLLECTION

Vesa Virtanen & Juhani Saari (Statistics Finland)
Surveys are like a box of chocolates – you never know what you're going to get? Assessing the impact of a digital chocolate voucher on the Finnish ESS9 on response rate and data quality

Øyvind Bruun Skarsbø & Kristin Holland (SSB)
Could "fresh" register data improve answers in surveys?

Coen Hendriks, Per Ola Haugen, Dag Gravem & Katharina Rosenbach (SSB)
Evidence-based quality improvement in questionnaire-based data collection

NORDIA / 13.00–13.50

PROCESSING SOCIAL STATISTICS

Margrét Gylfadóttir (Statistics Iceland)
Wage Price Index – 30 years of experience and questioning

Daniel Widegren (SCB) **& John Ekberg** (Swedish National Mediation Office)
The wage effect of commuting – An analysis of the gender wage gap in local labor markets

Jouko Kinnunen & Kenth Häggblom (ÅSUB)
Producing population scenarios within a regional macroeconomic model

Helle Visk & Kristi Lehto (Statistics Estonia)
Constructing families using administrative registers

EUROPAEA / 13.00–13.50

IMPROVING DIRECT COMMUNICATION WITH USERS

Katri Haverinen (Statistics Finland)
Automatization of Customer Service
Could a Chatbot Answer Statistical Questions?

Simon Sætre, Evy Esk, & Aud Kari Jorunn Holt (SSB)
Experiences from active dialogue with users (or collaborating with external actors)

Minja Tea Dzamarija (SSB)
Connecting data producers and policymakers – the role of the Coordinator of migration related statistics and analyses at Statistics Norway (SSB)

NAUTICA / 13.00–13.50

THE CHANGING ENVIRONMENT FOR OFFICIAL STATISTICS

Annie Stahel (DST)
The implications of data sharing in the public sector – a case for the universities

Hans Viggo Sæbø, Richard Rognerud & Trine Westvold (SSB)
Official statistics as a safeguard against fake news

Joakim Malmdin (SCB)
A Quality Framework for Official Statistics of Sweden

FENNIA / 14.00–14.50

MIXED-MODE SURVEYS

Tine Cordes (DST)
The impact of using web in the Danish LFS

Laura Saali (Statistics Finland)
The key to good quality mixed mode surveys: cooperation, hard work and patience

Kari-Anne Lund (SSB)
Forward-looking or backward-looking? A descriptive analytical case study of web survey mode embracement

Dag F. Gravem (SSB)
Mode-Mixers of the North Unite? Trends in Nordic and European data collection for social surveys

NORDIA / 14.00–14.50

FRUITFULL MULTINATIONAL COOPERATION

Arto Miettinen (Statistics Finland) **& Arne Jensen** (SSB)
Differences in accidents at work statistics between the Nordic countries

Jelle van der Kamp, Kalle Emil Holst Hansen, Peter Bøegh Nielsen (DST), **Henri Luomaranta** (Statistics Finland) **& Andreas Poldahl** (SCB)
Establishing Harmonised Business Statistics Databases in the Nordic NSIs – challenges and achievements

Vebjørn Aalandslid (SSB), **Guðjón Hauksson** (Statistics Iceland), **Marja-Liisa Helminen** (Statistics Finland), **Karin Lundström, Petter Wikström** (SCB) **& Jens Bjerre** (DST)
New comparative data on migrants and their integration in the Nordic countries

EUROPAEA / 14.00–14.50

CHANELLING EXPERT USERS THROUGH GOOD METADATA AND APIS

Piret Pihlaja & Taavi Mei *(Statistics Estonia)*
Harmonized structural metadata for faster and more coherent output.

Erik Pietinoverk, Daniel Davis & Emilia Saarinen *(Statistics Finland)*
Toward standardized statistical tables: Metadata driven PX file production.

Tryggvi Björgvinsson *(Statistics Iceland)*
A Leadership Lost: Reclaiming the Nordic's Place in the Open Data Space.

NAUTICA / 14.00–14.50

DATA SCIENCE AND TECHNOLOGY PARTNERSHIPS

Lasel Hess Mortensen *(DST)*
Why is Statistics Denmark interested in "Big Data in Biomedicine"?

Diana Beltadze *(Statistics Estonia)*
Creating Partnership with Census Data Users.

Svein Tore Bø *(SAS Institute)*
Succeeding with Analytics.

EUROPAEA / 15.10–16.00

FINDING – UNDERSTANDING – ACCOMPLISHING: EFFORTS IN MORE USER-FRIENDLY DISSEMINATION

Caroline Bo & Casper Winther *(DST)*
Explaining and analysing our data.

Maija Metsä *(Statistics Finland)*
Experiences of articles and blogging service based on new mobile first publishing concept.

Ann-Marie Karlsson *(Swedish Board of Agriculture)*
Ethos, pathos and logos for building trust in official statistics.

NAUTICA / 15.10–16.00

NEW ROADS

Jon Mortensen & Annette Thomsen *(DST)*
Assessing the MNE Pilot Exercise.

Hjörvar Pétursson, Svanhvít Eggertsdóttir, Magnús Kári Bergmann & Böðvar Þórisson *(Statistics Iceland)*
Estimating the residential accommodation component in a growing tourism market: New landscape, new methods.

Fredrik Granström, Sara Nilsson Hörnell & Ronny Kanali *(The Swedish National Council for Crime Prevention)*
Developing new statistics describing the legal process of criminal justice.

EUROPAEA / 15.10–16.00

MAKING AN IMPACT IN THE NEWS

Magnus Nørtoft & Helle Harbo Holm *(DST)*
Proactive presswork – during national election and in general.

Joona Majalahti *(Statistics Finland)*
What media wants from Statistics Finland experts? A study on how to increase the visibility of our experts.

Kristin Fredriksson *(SSB)*
Use your own statistical experts to set the agenda and build reputation.

NAUTICA / 15.10–16.00

USER DRIVEN INNOVATION

Mikael Skovbo & Dagmar Thomsen *(DST)*
The core business of statistical organisations – the case of tailor made data solutions.

Laura Mäkelä *(Statistics Finland)*
Design thinking in developing new concepts for statistical content.

Magnus Jonsson & Emma Schiro *(SSB)*
Modernization of statistics: How about modernizing the development process itself? – A conversation between an end user and a coder.

FENNIA / 15.10–16.00

COMBINING DIFFERENT SOURCES FOR COMPILING STATISTICS

Ossi Nurmi, Katja Löytynoja, Matti Kokkonen & Johannes Kola *(Statistics Finland)*
Method for compiling statistics on rented holiday homes.

Anna Korlyak, Marius Andersen & Camilla Rochlenge *(SSB)*
Data collection from different sources in price indices for passenger transport by air.

Lars Wiklund & Louise Sörme *(SCB)*
Municipal waste – existing data ready to be utilised.

NORDIA / 15.10–16.00

NEW ROADS

Jon Mortensen & Annette Thomsen *(DST)*
Assessing the MNE Pilot Exercise.

Hjörvar Pétursson, Svanhvít Eggertsdóttir, Magnús Kári Bergmann & Böðvar Þórisson *(Statistics Iceland)*
Estimating the residential accommodation component in a growing tourism market: New landscape, new methods.

Fredrik Granström, Sara Nilsson Hörnell & Ronny Kanali *(The Swedish National Council for Crime Prevention)*
Developing new statistics describing the legal process of criminal justice.

FENNIA / 16.10–17.00

DATA MANAGEMENT AND SHARING

Veiko Berendsen *(Statistics Estonia)*
How state data governance can contribute to better statistics.

Alvi Saar & Liisi Lillipuu *(Statistics Estonia)*
The role of administrative data manager in getting access to data and metadata.

Kaija Ruotsalainen *(Statistics Finland)* **& Ulla Agerskov** *(Nordic Council of Ministers)*
Nordic Mobility Statistics.

NORDIA / 16.10–17.00

STATISTICAL COMPUTING

Nina Gunnes *(SSB)*
Predicting the Contractual Full Time Equivalent Percentage Using XGBoost.

Kristian Harald Myklatun & Silje Skjølden Thorsnes *(SSB)*
Utilizing machine learning in the Consumer Price Index.

Jan Olav Rørhus *(SSB)*
Validation of international trade in goods data through machine learning.

Wednesday, 28 August, 2019 / Lunch at 12.30–13.30

PLENARY / 9.30–10.30

OPENING WORDS

Annika Damström, Moderator

KEYNOTE – DIGITALISATION IN THE PUBLIC SECTOR

Government CIO of Estonia, **Siim Sikkut**

PLENARY / 13.30–15.30

NORDSTAT GOLDEN DUCK AWARD CEREMONY

KEYNOTE

President **Tarja Halonen**

OPENING WORDS

Marjo Bruun, Director General of Statistics Finland

FENNIA / 10.40–11.30

NEW DATA, MACHINE LEARNING

Gustav Haraldsen & Arild Langseth (SSB)
A statistical approach to Big Data.

Thomas Aanensen (SSB)
Electricity data hub – a new data source in statistics.

Kristin Foldal Haugen & Thivyesh Ahilathasan (SSB)
Machine Learning on the classification of economic activities (SIC).

NORDIA / 10.40–11.30

LINKING MULTISOURCE DATA

Daniel Djerf, Tuuli Pihlajamaa, Essi Kaukonen, Riina Tammisto, Jonnika Leino, Sanna-Mari Pelkonen *(Statistics Finland)*, **Esa Hakkinen & Kai Koistinen** *(National Land Survey of Finland)*
Integrating areal classifications and geospatial data: Case KM (?).

Anders Grönvall *(Swedish Board of Agriculture)*
Record linkage in agricultural statistics.

Erik Fjærli (SSB)
Saving and investments in the private sector: Combining business statistics and social statistics can provide new insights.

EUROPAEA / 10.40–11.30

IMPLEMENTING NEW TOOLS FOR DISSEMINATING STATISTICS AND MICRODATA

Joonas Valve & Maria Yli-Heikkilä *(Natural Resources Institute Finland)*
Fishing dashboard: Interactive visualization of Fleet Economic Performance.

Annika Bruser *(Statistics Estonia)*
Building digital channels for fully personalized data experience.

Johan Sjøberg & Andreas Skogseo Rønning (SSB)
Microdata.no – instant access to research data with no applications.

NAUTICA / 10.40–11.30

MODERNISATION OF PROCESS, SYSTEM AND QUALITY MANAGEMENT

Essi Kaukonen & Mikko Salolla *(Statistics Finland)*
The expanding role of metadata supporting modern statistical production process.

Trygve Falch, Rune Gløersen & Jon Fallsdal (SSB)
Creating Partnership with Census Data Users.

Kristin Glomsås (SSB)
Hackweek in Statistic Norway.

FENNIA / 11.40–12.30

DIFFERENT FACES OF VALIDATION

Anette Mouglin Hertz & Katja Overgaard (DST)
Re-thinking data validation.

Aura Pasila, Meeri Seppä & Lea Löf *(Statistics Finland)*
Automatized Validation and Pseudonymization of Data Sets.

Auður Ólína Svavarsdóttir, Bryndís Bjarnadóttir, Garðar Páll Gíslason, Kolbrún Ýr Jóhannsdóttir & Violeta Calian *(Statistics Iceland)*
Error detection for the statistics of external trade in goods.

NORDIA / 11.40–12.30

ENSURING QUALITY IN OFFICIAL STATISTICS

Sanne Nielsen (DST)
Concepts – making the complex simple.

Gró Einarsdóttir *(Statistics Iceland)*
(Don't) be afraid of the conceptual validity.

Heather Bergdahl & Mats Haglund (SCB)
Reporting on Accuracy for the National Accounts.

EUROPAEA / 11.40–12.30

PROMOTING STATISTICAL LEARNING IN A TIME OF INFORMATION OVERLOAD

Marika Jokinen *(Statistics Finland)*
Fact-Based Noise in Social Media: Effects of Statistical education in Marketing.

Violeta Calian *(Statistics Iceland)*, **Anna Helga Jonsdottir, Gunnar Stefansson** *(University of Iceland)* **& Jamie Lentin** *(Shuttle Thread)*
SMLE: an e-learning system using data of official statistics.

Camilla Stabell & Kathrine Remen Hansen (SSB)
Reaching the young population.

NAUTICA / 11.40–12.30

THE GLOBAL PICTURE – NSIS AND THE SDGS

Nina Gunnes (SSB)
Predicting the Contractual Full-Time Equivalent Percentage Using XGBoost.

Kristian Harald Myklatun & Silje Skjelden Thorsnes (SSB)
Utilizing machine learning in the Consumer Price Index.

Jan Olav Barkas (SSB)
Validation of international trade in goods data through machine learning.

Speakers

Marjo Bruun

Director General
of Statistics Finland

Tarja Halonen

President

Esa Mäkinen

Data journalist,
Helsingin Sanomat

Siim Sikkut

Government CIO
of Estonia

Tuuli Kaskinen

Executive advisor,
Demos Helsinki

Kimmo Vehkalahti

University lecturer,
University of Helsinki

Samuel Kaski

Academy professor,
Aalto University

"I hope people have some time to roam and wander around the city and find themselves in a karaoke bar or a seaside sauna by accident."

ARTO MUTTUNEN, SOCIAL STATISTICS

나. NSM 2019 영역별 주요 발표자료 Abstract

□ PROCESSING SOCIAL STATISTICS

○ Margrét Gylfadóttir (Statistics Iceland)

Wage Price Index – 30 years of experience and questioning

<Abstract>

The Icelandic wage price index (IWPI) grounds on the act on wage index from 1989. The intention of the legislator was to get a price measure on the monetary compensation for an hour worked.

The IWPI aggregates employee wage information at the first stage by calculating Törnqvist month-to-month indices of occupation and economic activity "cells". These cell indices aggregates into higher-level numbers for economic sectors by using a Laspeyres type formula and then chained over time.

The index has become widely used in the Icelandic economy as the main indicator of changes in wages and wage growth, especially in collective negotiations by the social partners, who consequently started to question the soundness of the methodology of the index.

External and internal methodologist made comprehensive review of the IWPI in 2018 and 2019. The review highlighted among other the issues of:

- method
- coverage and sampling
- item life cycle
- quality changes
- chain drift

The reviews concluded that:

- The choice of index calculation method is sound.
- The coverage is sufficient.
- The effect of the item live cycle is small and not significant.
- The effect of quality changes (determined by length of service and education) should be examined.
- The chain drift is not significant since its direction is indeterminate at each time step and its values very small.

Keywords: Wages, index, price

○ Daniel Widegren (SCB) & John Ekberg (Swedisch National Mediation Office)

The wage effect of commuting - An analysis of the gender wage gap in local labor markets.

<Abstract>

Using a matched dataset on Swedish wage statistics, where information on both the local labor market and the workers commuting distance is included, the gender wage gap is investigated. The purpose of studying local labor markets is to describe the functioning of the labor market for geographical areas that are relatively independent of the outside world in terms of supply and demand of labor. The second part of the paper examines aspects of gender differences in commuting distance. By calculating, the distance between the work place and the workers registration address, using GPS-coordinates, gender differences of commuting in local labor markets are analyzed. Finally, the paper examine the effect of commuting on the gender pay gap.

The results show that the wage levels and the gender wage gap varies considerably across local labor markets. The analysis also reveals large gender differences in commuting distance within and across the local labor markets. It turns out that commuting distances seem to covariate, with wage and gender at sector-, occupational- and individual level. Although the covariation seams quite weak. When comparing short and long commuting regions it is found that commuting and wages co-vary - mostly for men. Analysis also show that longer commuting in occupations with a large wage dispersion is associated with higher wage for men.

Keywords: Gender wage gap, local labour market, commuting distance, GPS coordinates, data visualization

○ Helle Visk & Kristi Lehto (Statistics Estonia)

Constructing families using administrative registers.

<Abstract>

In the next census round in 2020-21, Estonia will conduct its first register-based census. In the register-based setting, households are formed of people who share a place of residence. The Estonian Population Register is lacking correct data on the residence of about 20% of people. The pilot census in 2016 revealed that relying on this data produces heavily biased statistics of households and families, e.g. share of lone parent families was 41% in the pilot census, but only 24% in 2011 census. This overestimation results from family members registered in different dwellings.

The key problem in reconstructing the families from the data is to detect partners if their registered home addresses do not coincide. To solve this problem, we collected additional information on 'signs of partnership' (SOPs): marriage, housing loans, mutual children, co-owning property and other administrative data that connects two persons and indicates potential partnership. Data on 17 SOPs from 9 registers was used to predict partnership status. The Estonian Social Survey and Estonian Labour Force Survey provided us data on actual partnership status.

A model based on logistic regression and stable marriage matching on 2018 SOP data was used to predict partnership. Predicted partnership was used as an input to construct register-based nuclear families of Comparative Survey of Household and Place of Residence respondents. The distributions of family characteristics improved substantially. Share of lone parents in the sample was 22% according to self-reported data and 24% by register data. In greater detail, however, the proportion of families of adult children is systematically overestimated – a result of particularly widespread mismatch between actual and registered place of residence among young adults.

Keywords: Couples, families, place of residence, population register, register-based census

□ THE CHANGING ENVIRONMENT FOR OFFICIAL STATISTICS

○ Annie Stahel (DST)

The implications of data sharing in the public sector – case for the universities

<Abstract>

In Denmark, the legal basis for collecting data for official statistics is ensured by the law on Statistics Denmark. Data that has been collected for the purpose of official statistics may not be used for other purposes, e.g. administrative purposes on a micro level where it is possible to identify the individual. The initial owner of administrative data (i.e. other public authorities) can, however, authorize and instruct Statistics Denmark in the role of data processor in different ways.

This paper investigates data sharing from a legal, organizational, and technical viewpoint and shares an example of concrete data sharing with the universities in Denmark. The paper will clarify what Statistics Denmark considers data sharing to be and describe the considerations that have been taken in connection with the sharing of data. We realize that the topic of data sharing has the potential to cause a certain amount of concern in NSI's as it did at Statistics Denmark and this aspect is also considered. The end result is a new way of utilizing competences concerning statistical micro data processing methods for non-statistical purposes in an increasingly data driven public sector.

Keywords: Facing the future, new roles and opportunities for the NSI's, data sharing, data processing, administrative purpose

※ Strategy and core values at Statistics Denmark

For some time, Statistics Denmark has had thoughts about how to facilitate a new and useful use of data and the data processing capabilities that we possess. With data being one of the world's most valuable resources, it is only fair that Statistics Denmark contributes to society in this field. Therefore, data sharing is an integral part of our Strategy 2022.

The core values of Statistics Denmark are independence, trustworthiness, data protection and user orientation. But as an old institution with its roots stemming from the birth of the Danish democracy we also need to develop in order to remain relevant in a continuously evolving world. Therefore, we want to strengthen our competencies and activities, especially in areas that can be characterised by our 3 new values, namely adaptability, a holistic approach and openness.

○ Joakim Malmdin (SCB)

A Quality Framework for Official Statistics of Sweden.

<Abstract>

Together with the 27 national authorities responsible for official statistics (SAM), Statistics Sweden has during 2016-2018 created a national quality framework for official statistics in Sweden which comprises

Legislation: definition of the quality concept for the development, production and dissemination of official statistics

Legislation: what is meant with the official release of official statistics (quality declaration as a necessary part of the official release)

Legislation: evaluation of the quality of official statistics

Guidelines: what constitutes official statistics

When laying out these four puzzle pieces a more comprehensive picture appears where Official statistics and European statistics become more and more integrated

The Swedish quality concept continues to closely relate to the European quality concept as defined in the ESS Code of Practice

User information needs and Purpose of the statistics are given a prominent place in the quality concept, quality declaration, quality evaluation and in the work to clarify what comprises official statistics

In this paper we will share our recent experiences in Sweden to create a common vocabulary for quality in statistics, clarify the purpose of the statistics in relation to user information needs, support the dialogue with users, and ultimately to strengthen the trademark of official statistics and stimulate quality improvements towards the goal that statistics be fit for their purpose.

Keywords: quality framework, quality concept, official statistics, purpose

□ DATA MANAGEMENT AND SHARING

○ Aivi Saar & Liisi Lillipuu (Statistics Estonia)

The role of administrative data manager in getting access to data & metadata.

<Abstract>

According to Statistics Estonia' s strategy, our goal is to produce high quality statistics with as low administrative burden and as high efficiency as possible. In order to achieve this, we need to improve the use of administrative data and describe the related metadata in our metadata management system.

Statistics Estonia (SE) has centralised its administrative data management. The process goes through different departments and teams, where central administrative data managers work in different roles and process phases.

In 2017, the Statistics Design Department was created, where administrative data and metadata are managed centrally. The central role of administrative data managers in this department includes creating and developing partnerships with data owners, coordinating information exchange and consulting subject matter experts in Statistics Estonia, and managing data delivery contracts and data requests. The methodologists of the Statistics Design Department are also responsible for getting access to metadata of administrative data, negotiating with data owners and agreeing on the deadlines, data delivery formats and structures.

At the moment, SE uses over 100 different administrative data sources (state registries) in the statistical production process. Managing, describing and improving the related information and metadata of those sources is a challenging and ongoing process.

Keywords: administrative data, metadata, data management

□ NEW DATA, MACHINE LEARNING

○ Gustav Haraldsen & Arild Langseth (SSB)

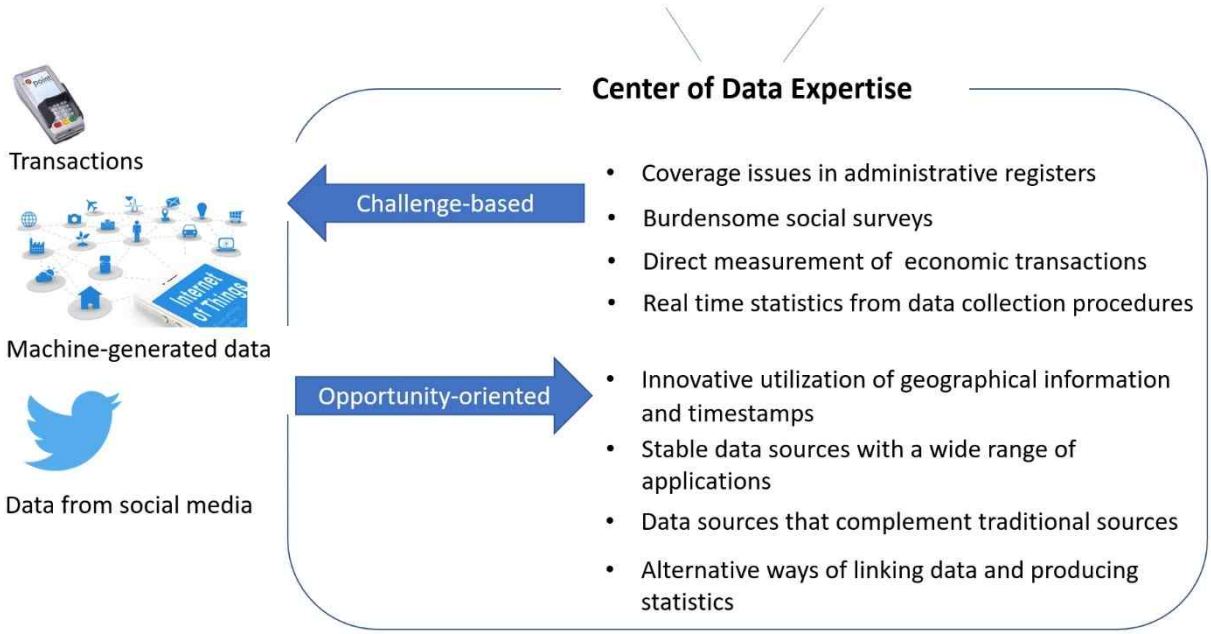
A statistical approach to Big Data

<Abstract>

Statistics Norway is presently working on a strategy on how to utilize new data sources. We argue for a strategy which combines what we name a challenge-based approach and an opportunity-oriented approach. The challenge-based approach takes quality issues recognized by the statistical producers as a starting point, and we report from three focus groups that we have conducted addressing this. The opportunity-oriented approach takes the characteristics of Big Data as a starting point and explores what kind of statistics which can be produced from them. We discuss three main ways to produce statistics from Big Data sources.

Keywords: Statistics, Big Data, new data sources

<Challenge-based and Opportunity-oriented approach>



○ Thomas Aanensen (SSB)

Electricity data hub – a new data source in statistics

<Abstract>

An electricity market data hub was launched in Norway in the first quarter of 2019 and provides an exciting new and extensive data source which can be used for many different statistical purposes.

Other Nordic countries like Sweden and Finland are in the processes of developing similar electricity market hubs and in Denmark and Estonia these types of hubs are already operational.

The electricity data hub is a centralized information exchange system for the electricity retail market which contains data on consumption and production from remotely (and traditional) readable electricity metering points.

The information stored in the data hub will primarily be used by electricity suppliers and distribution network companies serving electricity consumers more efficiently, however this very rich data source can also be a very useful source of information for us statisticians.

In the paper we want to share our experiences from the process of getting hold of the data from the hub and discuss how this new and rich data source can be used to produce better and more timely electricity statistics and briefly look at other statistical purposes as for instance improving the quality of our statistical registers, identifying empty dwellings and use of the data for research purposes.

We will also discuss some of the limitations in the data from the hub and how we plan to approach the regulating authority in the Norwegian electricity market with recommendations for improving the quality of the data and adding additional information to make the electricity hub more applicable for statistical purposes.

Keywords: data hub electricity big data

○ Kristin Foldal Haugen & Thivyesh Ahilathasan (SSB)

Machine Learning on the classification of economic activities(SIC)

<Abstract>

Today, an executive officer at the Central Coordinating Register for Legal Entities classifies the entities according to the activity the entities carry out. This classification is a time-consuming process and often leads to the wrong result. By these reasons a new system, which would provide automation of the classification process, is required. Statistics Norway together with other agencies in Norway has an ongoing project on developing a new system for determining standard industrial classification (SIC) codes. The new system will let the entities themselves classify their activities with the help of a trained model. In the paper, several models are presented, being developed with the help of machine learning techniques to suggest to entities SIC codes based on the companies' business descriptions. The models tested are fastText, CNN and a two-level modelling. Additionally, there are outlighted ways to handle an imbalanced dataset, which contains a lot of mistakes.

Keywords: classification of SIC, machine learning, text classification, quality improvement

다. 연구 수행 관련 시사점 및 제안사항

□ 데이터 연계 및 자료수집 방법론, 통계 인프라 구축 및 분석을 통한 정책 활용, 정책 통계지표 개발 및 관리 등과 관련하여 북유럽 국가 통계청이 공동으로 3년마다 개최하며, 2019년 “Facts for Future"를 주제로 통계 및 데이터 분야 전문가들이 참가하여 개최된 Workshop(NSM, Nordic Statistical Meeting)은 연구과제(포용국가 기반 사회정책 통계 인프라와 지표 개발) 수행과 관련하여 다음과 같은 시사점 및 정책적 제안사항을 도출함

□ 사회정책 통계 인프라 및 지표 개발, 다양한 마이크로데이터 생산 및 활용, 미래 예측을 위한 다양한 방법론의 활용 등 기존 통계 인프라 생산 및 관리에서의 한계 극복 및 혁신을 위하여 북유럽 통계청 및 관계 전문가 역시 위기의식과 함께 선제적인 노력을 통해서 함께 고민하고 대안을 모색하고 있음

☞ “근거 기반 정책(evidence-based policy)” 에서 “데이터 주도 정책(data-driven policy)” 으로 강화하고 있음. 또한, 기존의 조사 방법론의 혁신 및 Register-based Survey를 통한 데이터 생산 및 활용, 통계 지표 생산 방법 관련 대안 모색을 강조함

☞ 데이터 주도의 통계 혁신은 지능정보기술을 활용한 미래 예측 및 최적화(optimization)된 대안 제시를 강조하고 있으며, 기존 통계 인프라 영역에서도 빅데이터와 다양한 방법론을 활용한 데이터 주도의 통계 혁신을 통해 정책을 수립하고 평가하는데 활용함으로써 새로운 가치를 제고하는데 주목하고 있음

□ 정책 산출(Output)은 공공 재화나 서비스의 형태로 만들어진 정책 집행의 직접적 산물이며, 정책성과 또는 결과(Outcome)는 이러한

산출(Output)이 정책대상(Policy Target Group) 집단에 가져오는 최종적 영향(Final Effect)으로서, 우리나라의 경우, 정책성과 또는 결과(Outcome)를 반영할 수 있고, 생애주기별 삶의 영역과 생활 기반에 따라 정책과 연계될 수 있는 사회정책 통계 인프라 구축 및 통계 지표 발굴 필요

☞ 포용국가 비전 전략 기반 사회정책 영역 및 생애주기별 분류에 따라 기존 통계 지표의 한계를 개선하고 새로운 분류 프레임에 따른 통계지표 조정 및 개선 작업 필요

- 이러한 동향을 반영하여 포용과 혁신의 사회정책 주요 영역 및 생애주기별 생활 영역과 삶의 기반 등 분류 프레임에 따른 기존 통계지표 조정 및 메타 데이터를 제시하기 위하여 신규 통계지표 발굴 및 메타 데이터 정의 및 산출 방법론 도출

□ 현실에 대한 진단과 미래 예측을 위한 북유럽 통계 개선 노력에 대한 최근 동향을 참조하여, 우리나라 정책 통계의 한계 및 수요 변화 등을 고려하여 포용과 혁신의 사회정책 영역 및 생애주기별 통계지표 분류에 따른 신규 통계지표 생산 및 관리 방안 제안

- 사회정책 통계지표 생산 및 관리를 위한 부처 간 협업체계 구축 및 행정통계 연계 기반 실태조사 수행을 제안함
- 포용과 혁신의 사회정책 영역 및 생애주기별로 통계지표 분류 프레임을 제시하고, 이에 따른 조정 및 신규 통계지표 발굴 및 생산, 관리 및 활용 등 정책적 활용방안을 제시함

□ 사회정책 통계지표 분류 및 재구성, 신규 통계지표의 발굴 등을 통하여 사회정책의 변화에 따라 정책 효과성을 평가할 수 있는 통계지표 인프라 구축방향과 함께 새로운 방법론을 벤치마킹하여 설계 및 생산할 수 있는 대안 모색이 지속적으로 필요함