



Creating a National Remote Access System for Register-based Research

Marianne Johnson, Statistics Finland
Statistical Data Confidentiality Work Session Oct 2015

Finnish administrative registers

- several comprehensive national registers
- contain unit level data on **individuals, families, housing, enterprises**
- compiled and maintained for administrative or statistical purposes, e.g.
 - ***Population Register Centre (VRK)***
 - Population information system
 - ***Social Insurance Institution (KELA)***
 - Registers on obtained social benefits
 - ***National Institute for Health and Welfare (THL)***
 - Medical Birth Register,
 - Care Registers for Social Welfare and Health Care (HILMO),
 - Finnish Cancer Register
 - ***Ministry of Labour (TEM)***
 - Register over job seekers
 - ***Statistics Finland (Tilastokeskus)***

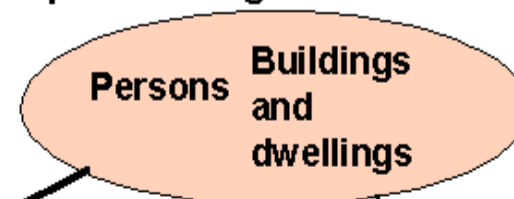
Secondary usage of administrative registers

- Production of official statistics is to a large extent based on registers in Finland
 - the population and housing census has been based totally on register sources since 1990
 - Handbook: Use of Registers and Administrative Data Sources for Statistical Purposes – Best Practices of Statistics Finland
- Register-based research
 - 20 % of doctoral thesis' within medicine in Finland include data from national registers

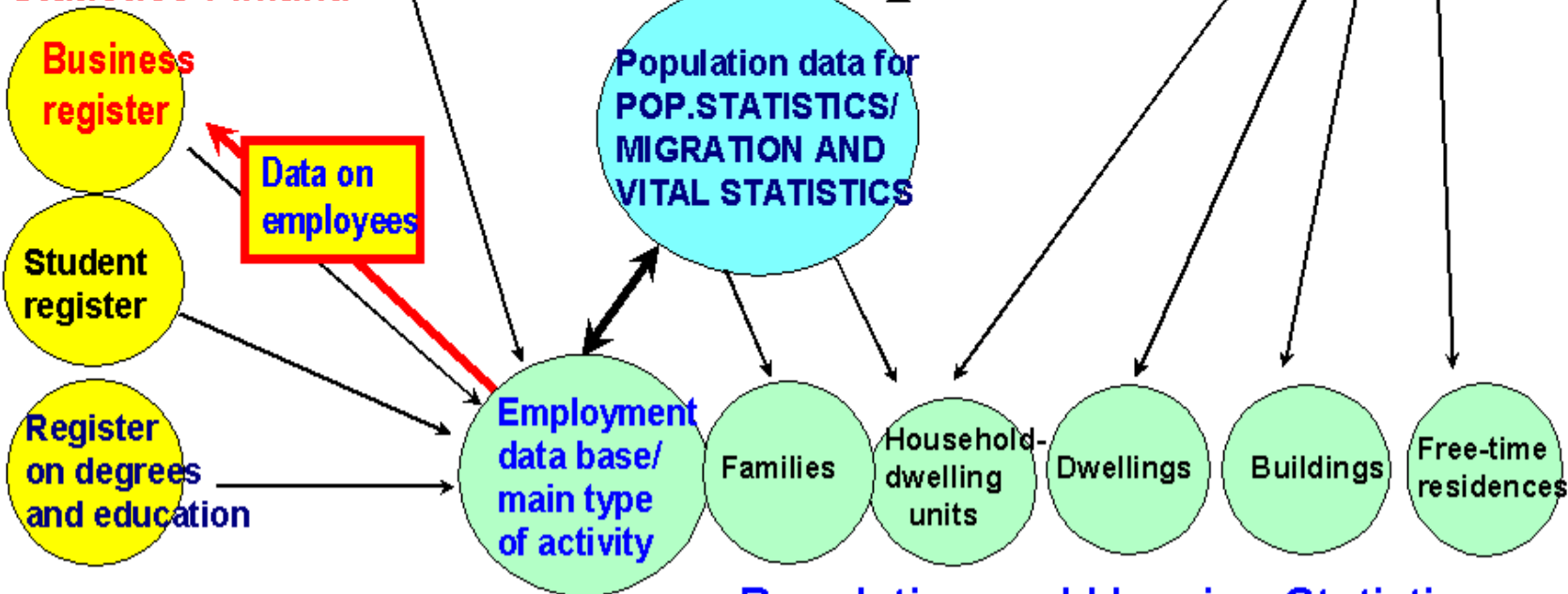
Administrative registers



Population Information System of the Population Register Centre



Statistics Finland



Population and Housing Statistics

Prerequisites for register-based research

- Common personal identification number in all registers
 - first used in 1964 (between 1964-1970 two different systems)
 - since 1971 a digital population register
 - all Finns have a PIN

→ data from different registers can be linked by PIN e.g. for research purposes
- Legislation that allows the use of confidential personal data for scientific research
 - Trust in register keepers and researchers
- Comprehensive, well documented registers

Legislative basis for research use of data from Statistics Finland

- Statistics Act (280/2004)
- In 2013 the Statistics Act was amended to better facilitate the use of data gathered at Statistics Finland for research purposes.
- New objective of the Act
 - **To extend the use of the data collected for statistical purposes in scientific studies and statistical surveys on social conditions.**
- Possibility for researchers to **gain access** to confidential data from which **only the direct identifiers** have been removed.
 - *Before 2013* statistical authorities **could not** give permission to such confidential data from which the statistical unit could be **indirectly identified**.
 - *Gain access = see and analyze data by a remote access - system*

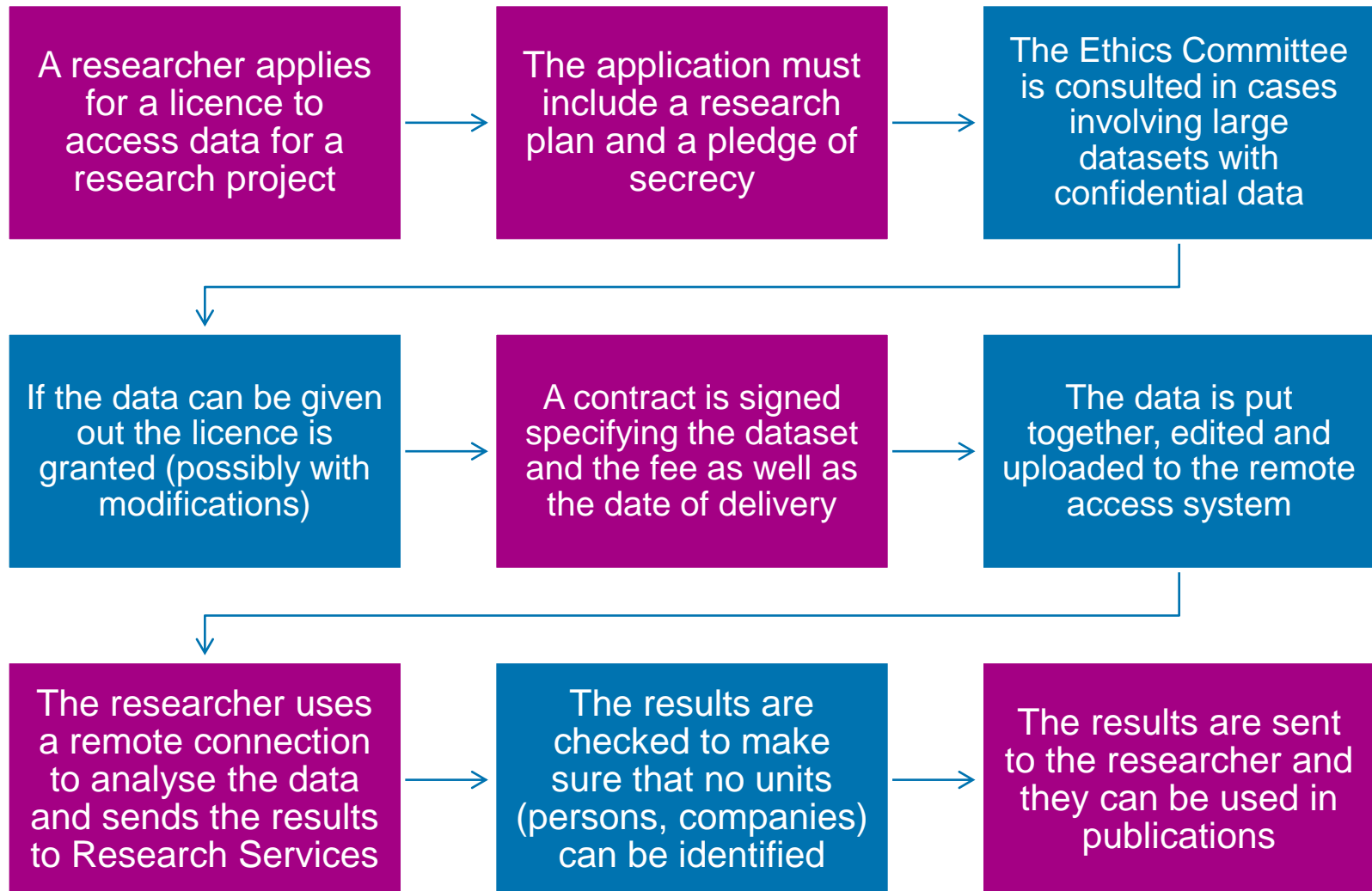
Remote access system (FIONA)

- In use at Statistics Finland since 2009, development project 2014-2015
- Model taken from Sweden, Denmark and the Netherlands
- Researchers use data on Statistics Finland's server at their own workplace via a secured Internet connection, data remains at SF
- Researchers use a Windows remote desktop, and have access to the data they have obtained permission to as well as to metadata
- The researchers have access to wide range of statistical programs : STATA, SPSS, R, SAS, Python Anaconda, ...
- Each research project has its dedicated folders and storage space in the system
- Technical maintenance of the FIONA-system transferred to CSC-It Centre for Science in 2015
- Number of users and data sets in the remote access system is growing steadily, currently about 150 active users

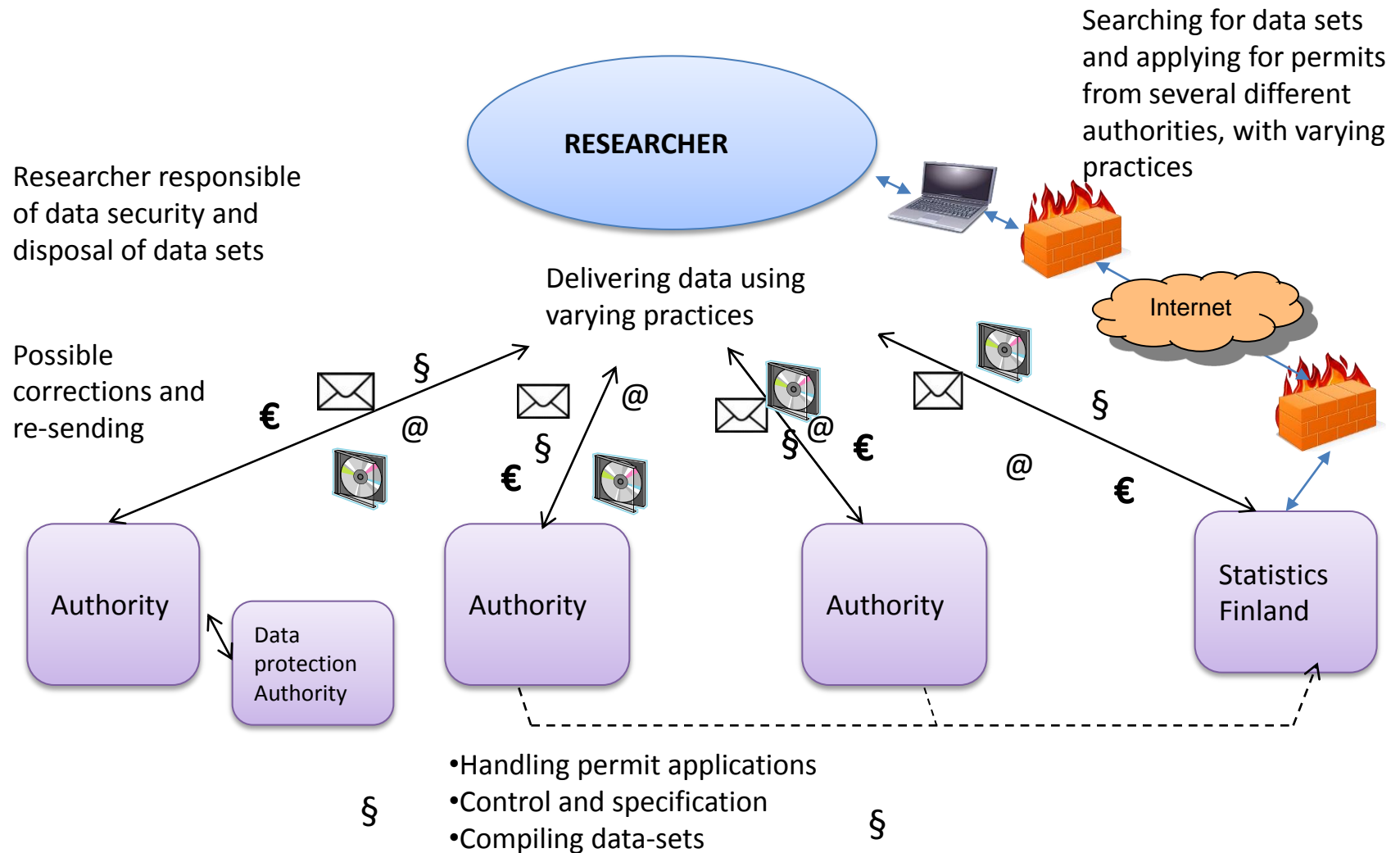
Confidentiality

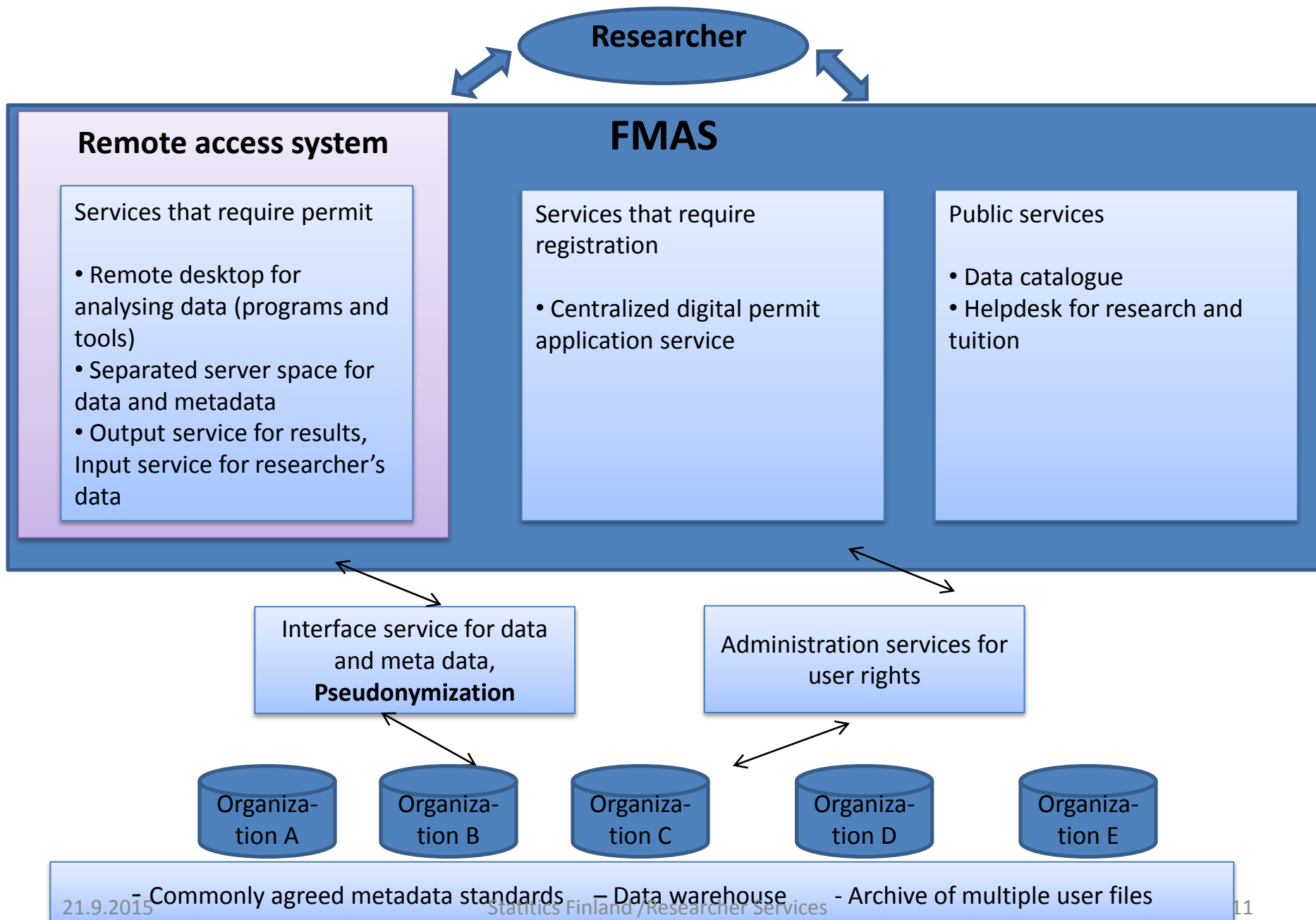
- Research data sets are stored on Statistics Finland's /CSC's servers
- Only mouse, keyboard and graphic signals are transferred
- Access to the system only from preapproved IP-addresses
- A disposable SMS password is sent each time the researcher logs in to FIONA
- All data transfers from and to FIONA are handled by personnel at the Researcher Services of SF
 - Outputs are checked so that direct or indirect identification is not possible and files are saved for possible future reference
- Access to data is terminated when the permit for the project expires
- FIONA environment is separated from the production network
- The system will be audited in fall 2015 after being transferred to CSC

A typical process in applying for sensitive research data



Present process for obtaining register data for research

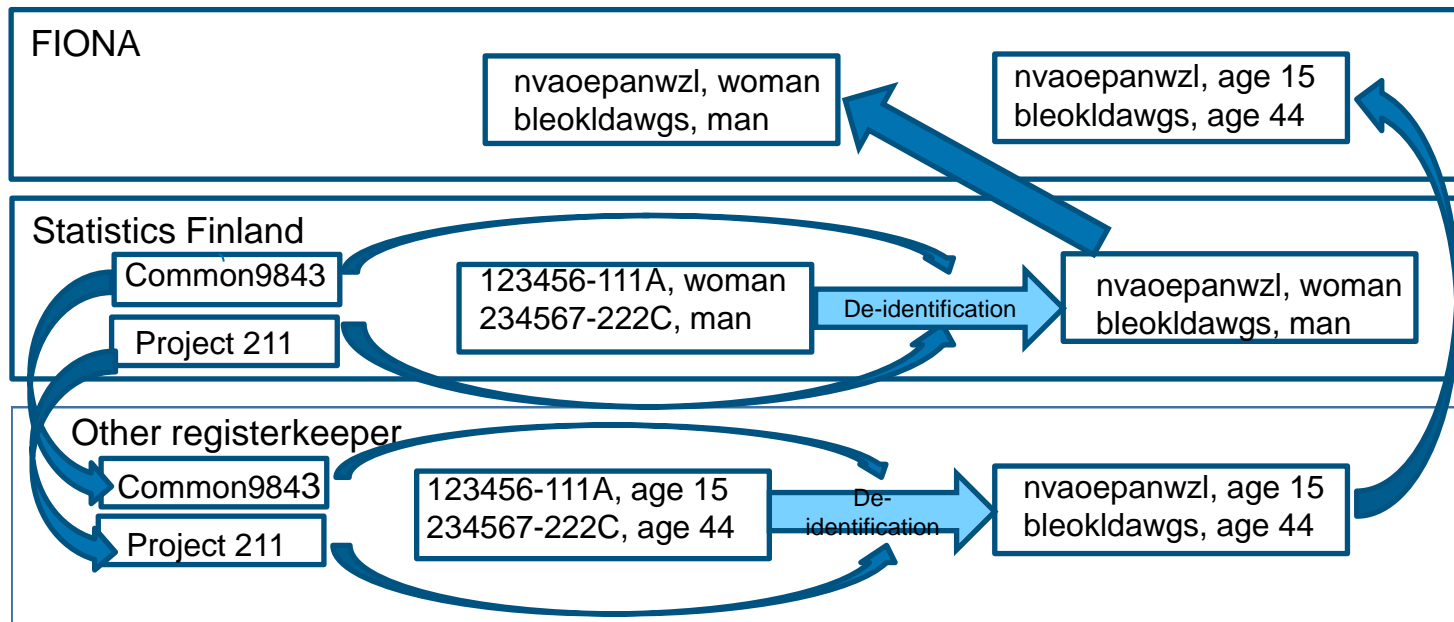




Linking data from different sources

- Present method
 - Register keepers send the data requested by the researcher over a secure connection , by recommended mail, with courier services etc. to Statistics Finland
 - The data includes the Finnish PIN or BIN (or a pseudocode created by the register keeper and the key is sent separately)
 - Statistics Finland creates a project specific pseudocode, changes the PIN (BIN) in the research data sets and uploads the data in the remote access system
- Aim
 - Pseudocodes should be used in all data deliveries
 - Register keepers should be able to upload their data direct to the remote access system using a standard pseudonymization method

Pseudonymization –project specific



To be developed....

- We see a problem with the set pseudocodes of the 'ready-made' data files
 - Solution 1: Create project specific pseudocode also for projects that use the 'ready made'
 - Problem: A copy of 'ready made' data sets has to be made for each project -> much excessive disc space is needed
 - Solution 2: Send the seed code that has been used for the 'ready made' files to the other register keepers
 - Problem: The key PIN /BIN - pseudocode used by Statistics Finland will be widely known