

# Creating ready-made research datasets from national administrative registers

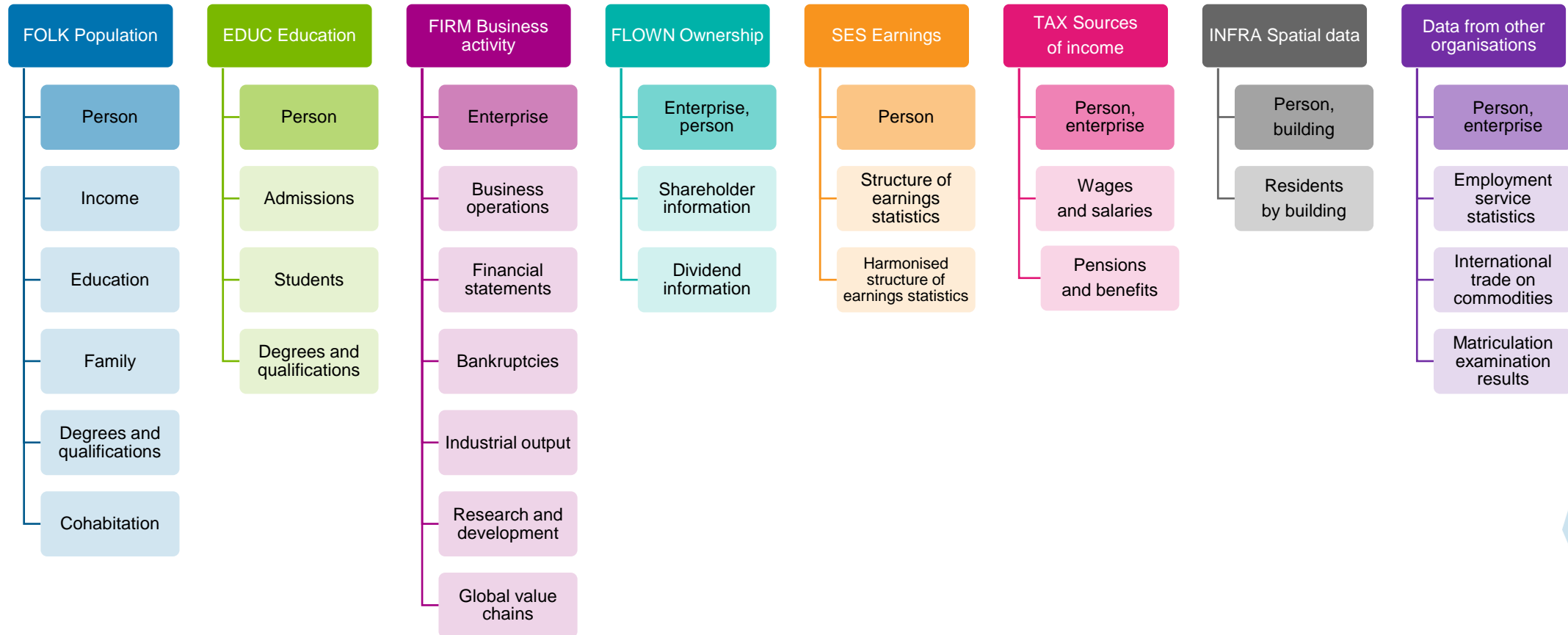
Päivi Kankaanranta and Aino Melakari

The 2021 Joint UNECE/Eurostat Expert Meeting on Statistical Data Confidentiality

# Innovative approach to providing access to microdata for research purposes

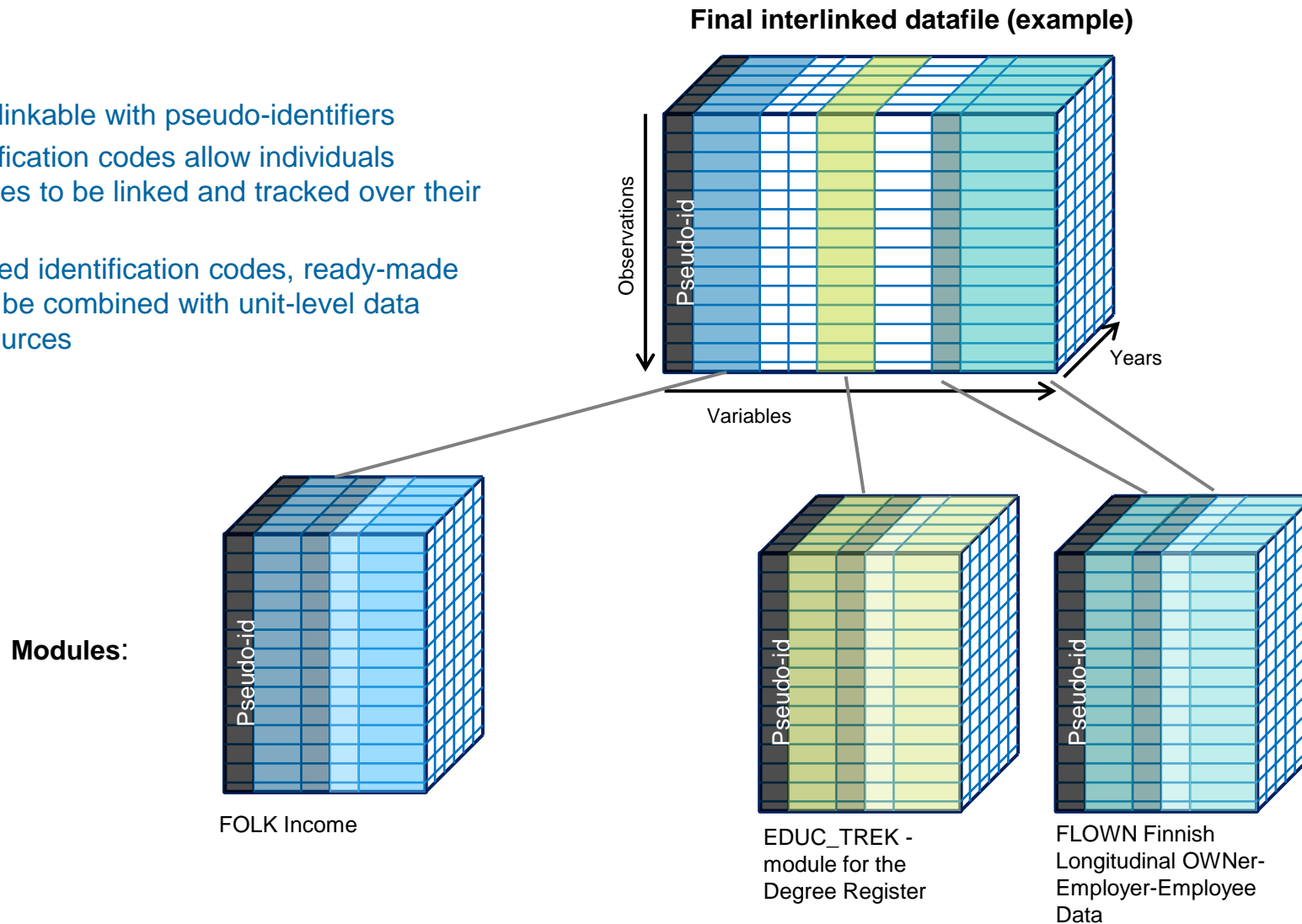
- Improving access to national administrative registers for research purposes is one of the strategic goals of Statistics Finland
- The Research Services at Statistics Finland produces ready-made data modules on the basis of researchers' needs from the in-house data reserves
  - These reserves consist mainly of register-based data used to produce official statistics.
- Ready-made data modules contain a standard set of variables with unit-level information on all enterprises and persons extracted mainly from administrative records.
  - Information on the whole target population as opposed to a sample of survey respondents
- The number of variables in a module is limited in order to accommodate GDPR demand for data minimisation
  - However, the number of statistical units in a module is not limited

# Thematic collections of ready-made research data modules, target populations and selected topics within themes



# Interlinked ready-made data modules create rich datasets

- Modules are linkable with pseudo-identifiers
- Unified identification codes allow individuals and businesses to be linked and tracked over their life span
- With the unified identification codes, ready-made modules can be combined with unit-level data from other sources



# Smooth authorisation of access to confidential microdata

- Licence applications for ready-made data modules are processed in less time compared to tailor-made research data
- Currently, the licence covers all future updates to data for as long as the licence is valid
  - There is no need to separately apply for a new licence to access the latest release of the research data
  - New licence application is required whenever new ready-made modules or other data are added to an existing research project
- When access to microdata on persons is requested, the applicant needs to provide a justification for the need to use information on the total population
  - Qualifying reasons include e.g. following families over time or producing control groups for the study population

# Legal and practical measures supporting the release and use of ready-made research data modules

- The EU-level and national legal arrangements define which unit-level information contained in registers can be released and how this microdata can be accessed
  - Direct identifiers are excluded and unique identification codes are replaced with pseudo-identifiers
  - Research projects use only the modules necessary for their research purposes
- The ready-made data can be used only via the remote access system FIONA
  - FIONA is a closed environment without external internet connection
  - All data transfer between FIONA and researchers goes through an output checking protocol
  - The researchers pledge to follow the data protection requirements (e.g. threshold value, dominance rule)

# Conclusion: a data release policy enhancing access to unit-level data

- The release policy does not compromise the confidentiality of the data
- Cost- and time-effective for both Statistics Finland and the research projects
- Ready-made data modules are valued by the researchers
- Ready-made data (modules) are expected to constitute the core of the microdata services provided for researchers by Statistics Finland
  - A separate team for maintaining and developing ready-made modules
  - The number of modules is likely to increase in the near future
  - COVID19 pandemic has brought about the need for quick up-to-date economic analysis and ready-made modules that are updated on a monthly basis

# Thank you!