

The potential of micro data in developing new research frontiers – a business statistics approach

CESS 2024

Kalle Emil Holst Hansen

Peter Bøegh Nielsen



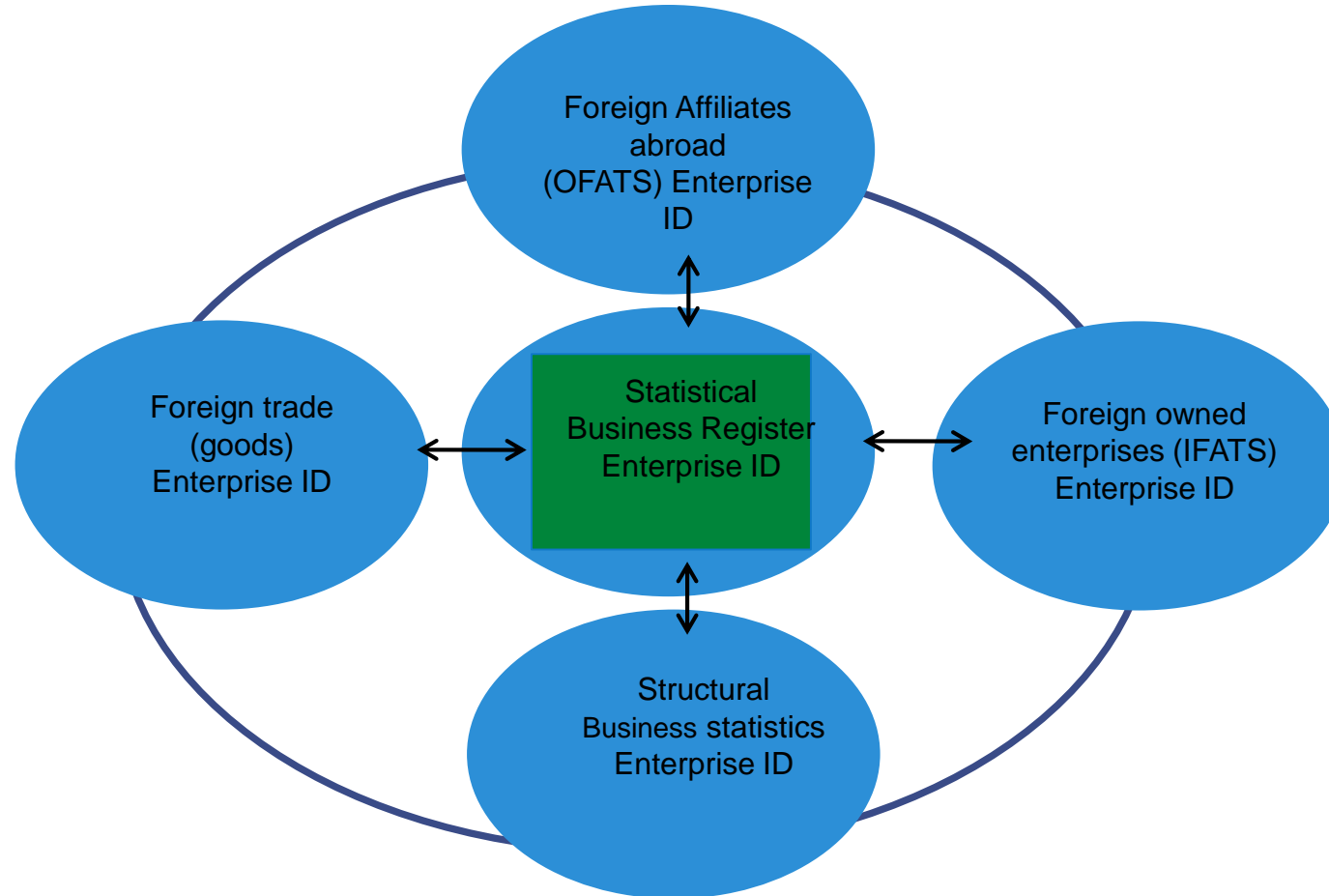
Presentation overview

- Why focus on Micro data linking (MDL)?
- MDL methodological issues and consistency
- New data initiatives
- The Nordic MDL approach and beyond

Use of Micro Data Linking at Statistics Denmark

- Production of statistics, especially concerning social statistics
- A means to develop new official statistics without launching new surveys
- Production of experimental statistics accommodating shifting analytical and policy agendas
- Access for external users: researchers and ministries

What is Micro data Linking within business statistics? Linking of statistical registers at unit level via the SBR



Main Methodological Issues

Fundament: Identical ID number used in the SBR and across statistical domains

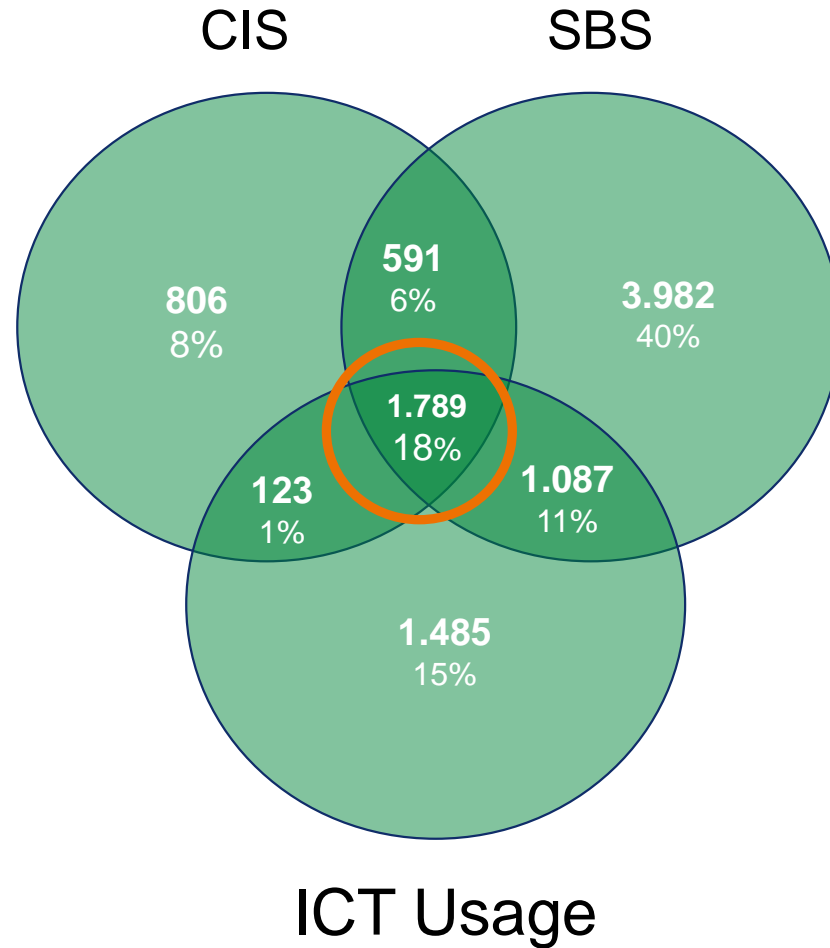
Unit consistency: businesses with complex legal structures (enterprise groups) might use different ID no.s for reporting to different statistics (example: SBS and ITGS)

Sampling frame and populations differ across the statistics in question due to scope in terms of activity and/or size class

Data consistency: Checks for reported values to different statistics. Inconsistency might be due to different reference periods, different definitions or simply reporting of incorrect values

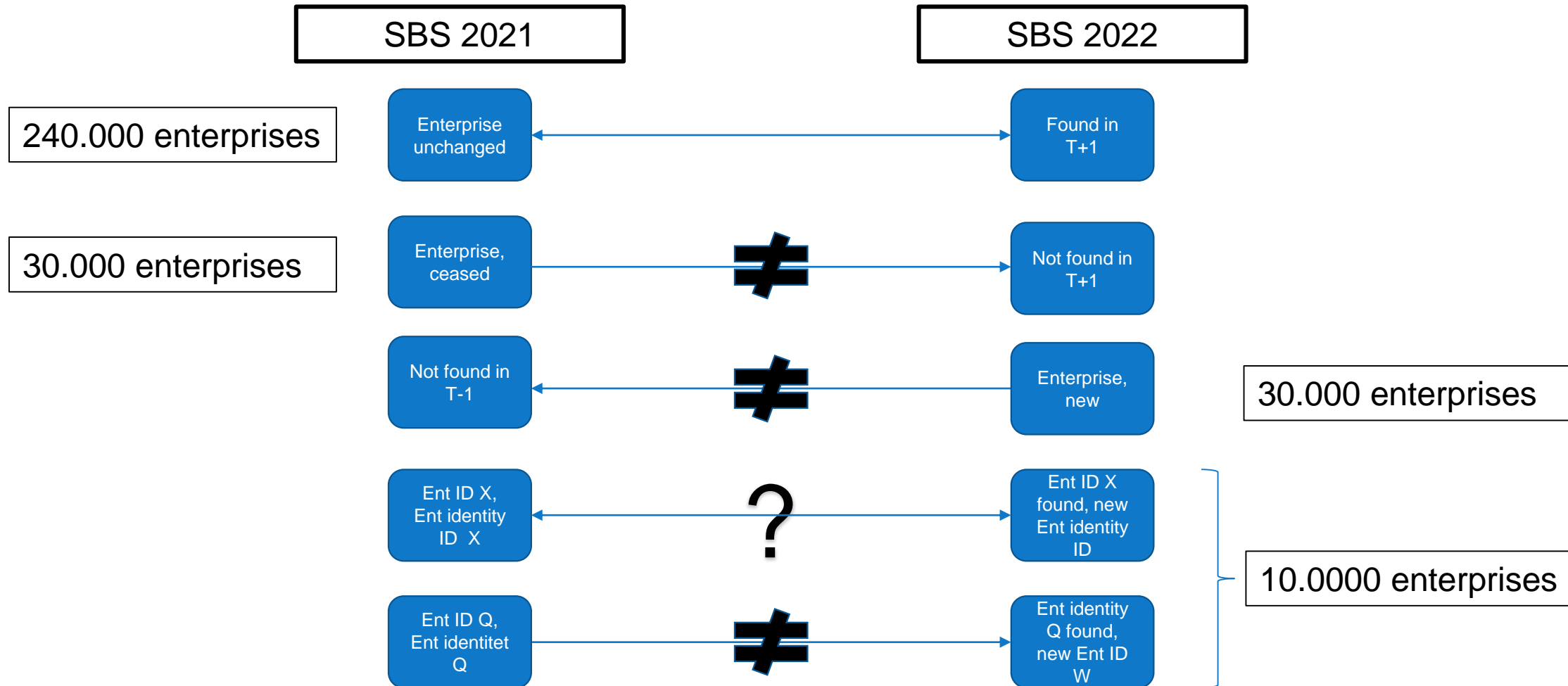
Official or experimental statistics? If official statistics, reweighting is an issue, if experimental statistics only using the observed units, representatively can be an issue

Match and no-match between the different survey samples Danish case



Consistency over time

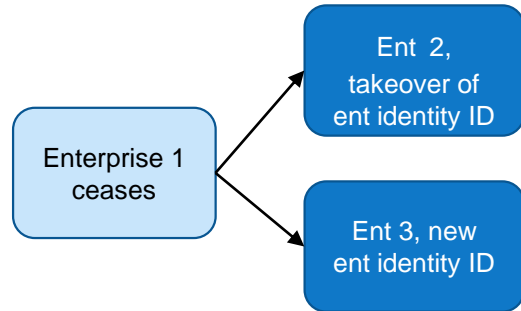
– What constitutes a good year-on-year match?



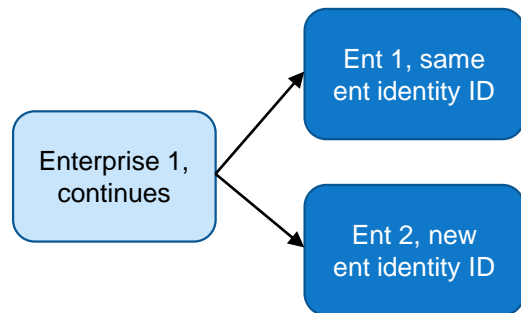
Consistency over time

– demographic event and enterprise ID changes

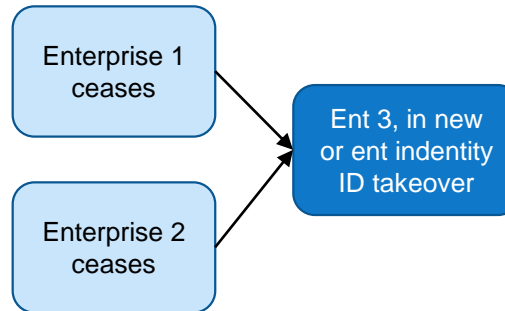
1. Split-ups



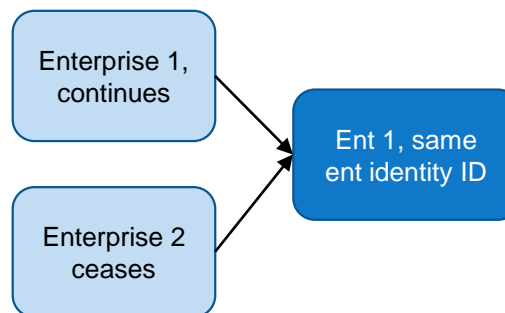
2. Spin-offs



3. Mergers



4. takeovers



Business demography

New enterprises, no demographic events

Ceased enterprises, no demographic events

Quality challenges to be continuously addressed

- Complex enterprise structures
 - Monitor the largest groups
 - Introduce a Group ID no.?
 - Eliminate intragroup transactions
 - How to determine activity of groups?
- Improve capturing of demographic events
- Increased data confrontation before final editing of data
- Addressing the conflicting goals of improving micro data research and minimise respondent burden

What's inside the new Statistics Denmark Data Warehouse

- 19 different business statistics registers
 - Of which 9 new
 - All microdata
 - Continuous updating
- Improved metadata and documentation
- We aim for adding
 - registers addressing the green economy
 - short term statistics



Connecting the dots – in new dimensions

Linking business statistics with other data sources

Privately held data:

- Electric meter data
- Scanner data
- Platform data

Social statistics registers:

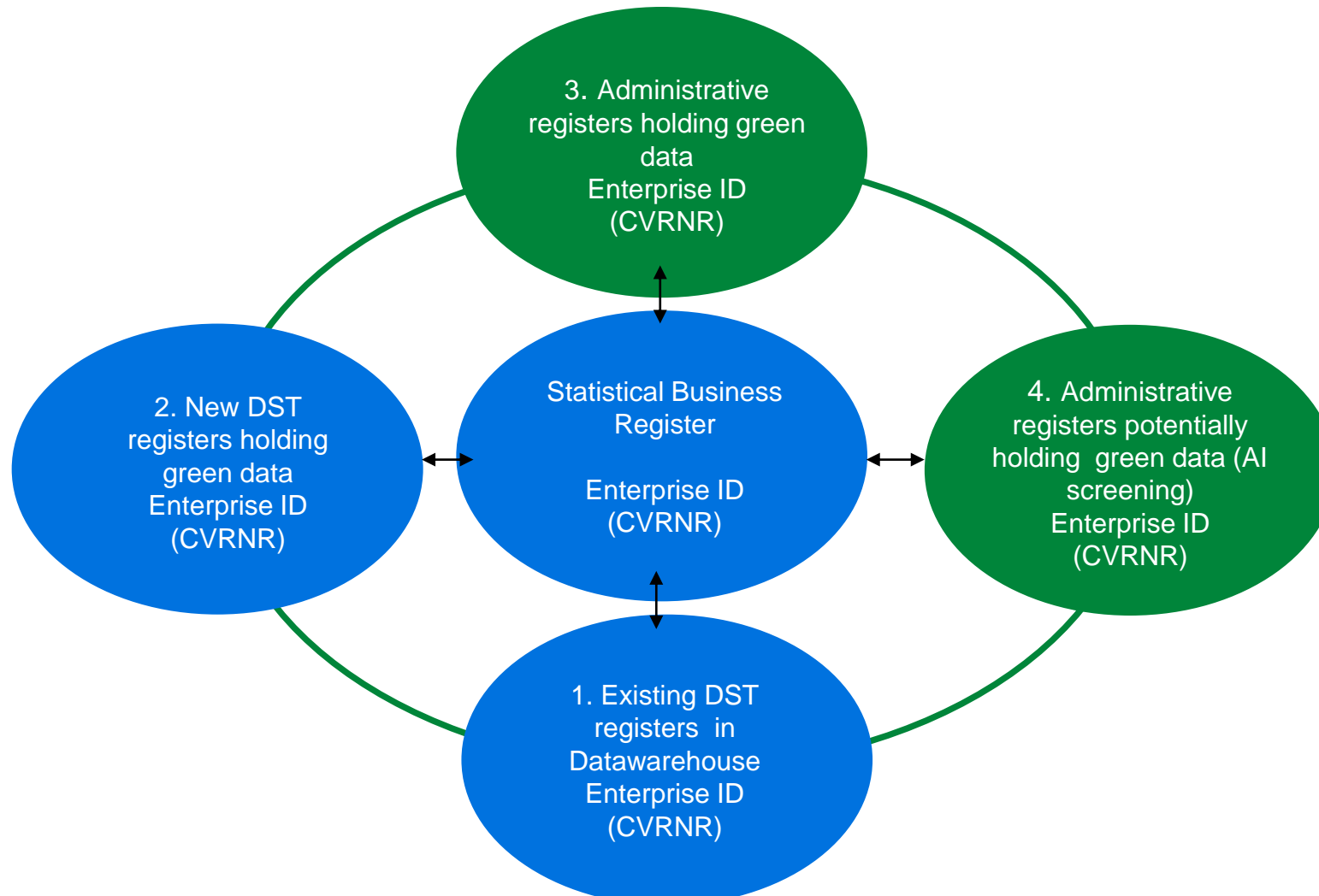
- Education
- Employment
- Income
- Health



New / future admin data:

- CSR-D (ESG) data
- Traffic toll data
- Climate Atlas

The possible road ahead: Enlarging the current Business Statistics Data Warehouse with administrative registers holding green data at micro data



The Nordic micro data infrastructure at a glance: Concepts of harmonised databases and shared syntaxes

1. Defining the scope and content

Which registers, variables and statistical years to be included?

2. Data validation across statistical registers

Focus on unit identification, identity over time, variable values, cross validation and correct links.

3. Establishing databases stored in each NSI

- Harmonise labelling of variables
- import and store the contents agreed in each NSI.
- Establish procedure for automatic update

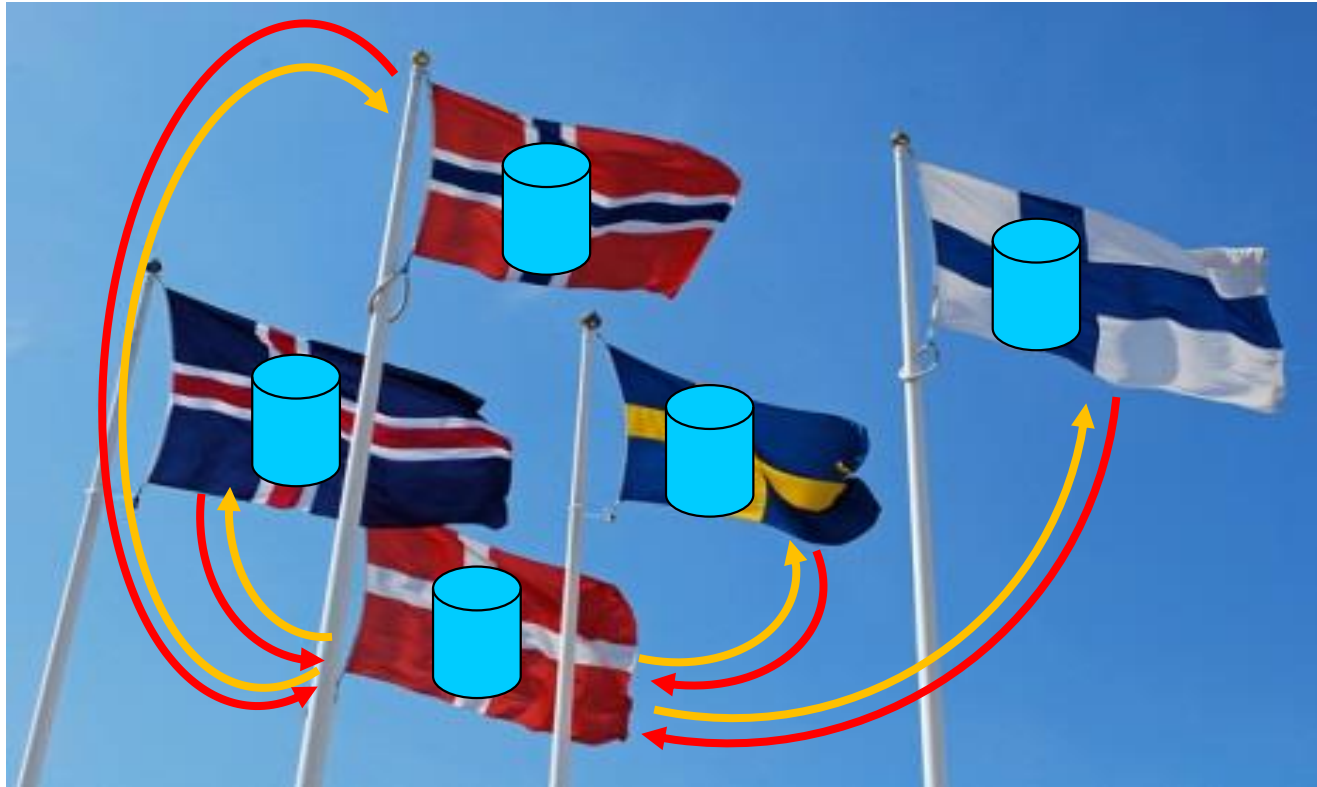
4. Define topic to be addressed

Centrally scripted syntax and decentralised code execution

5. Harmonised output tables

Each NSI deliver the requested tabular data to the project co-ordinator

Distributed micro data research in the Nordic project: How?



1) Guidelines for database 2) SAS code for execution

3) Standardised output tables for analysis

Database basic content: Harmonised variables based on EU regulations – allowing for expansion outside the Nordics!

- **Statistical Business Register:**
 - Activity/Group relation
- **Business Demography:**
 - Newly born/Dead enterprises/Age
- **Structural Business Statistics:**
 - Turnover/Gross Value Added/Full time equivalent no. of employees
- **International Trade in Goods Statistics**
 - Exports/imports of goods (values, destination, type of goods)
- **FATS statistics:**
 - Nationality of ownership/ No. of employees in foreign affiliates

Thank you!

Any questions?

