Activity 1.0 – Part A
Administrative data

26-29 September 2022, Amman, Jordan

Statistics Denmark's use of administrative data for the production of official statistics

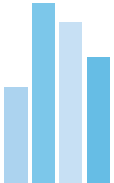
Peter Bohnstedt Anan Hansen, Statistics Denmark, pah@dst.dk

STATISTICS DENMARK

Outline of presentation:

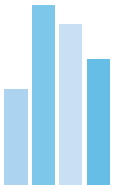
- *The value of administrative data for statistics*
- *Approach to working with administrative data*
- *How to develop a Memorandum of Understanding (MoU)*
- *Preconditions, challenges and pitfalls*

STATISTICS DENMARK



The value of administrative data for statistics

STATISTICS DENMARK



The role of NSIs: Why are we here?

- Part of the foundation for democracy and the economy – serving the country
  - Policy makers, media, scientists, businesses and citizens
- Describe and explain society's structures and development in figures – common reference points
- Enable evidence based policy making
  - Build knowledge, add value and reduce risks

What are the basic principles?

- Independence of NSI specified in law
- Statistics must be:
  - Impartial and objective: “Policy relevant – politically neutral”
  - Available to all users at the same time
  - Reliable - when errors occur, corrections are published asap
- Confidentiality of persons and businesses is guaranteed

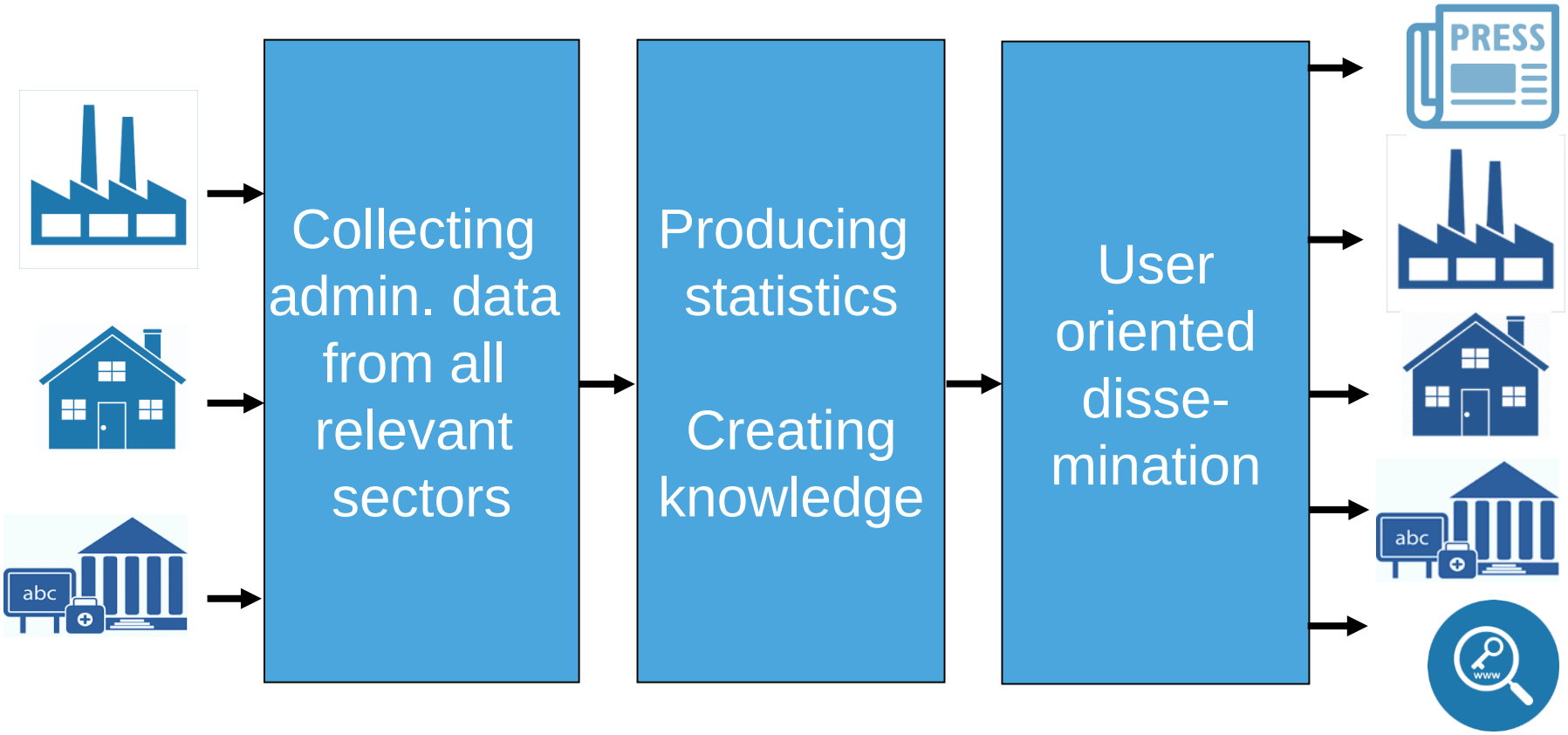
What is ‘admin data’ in this context?

- Structured information which is originally …
- collected by other authorities for their own purposes ...
- re-used by the NSI for official statistics.

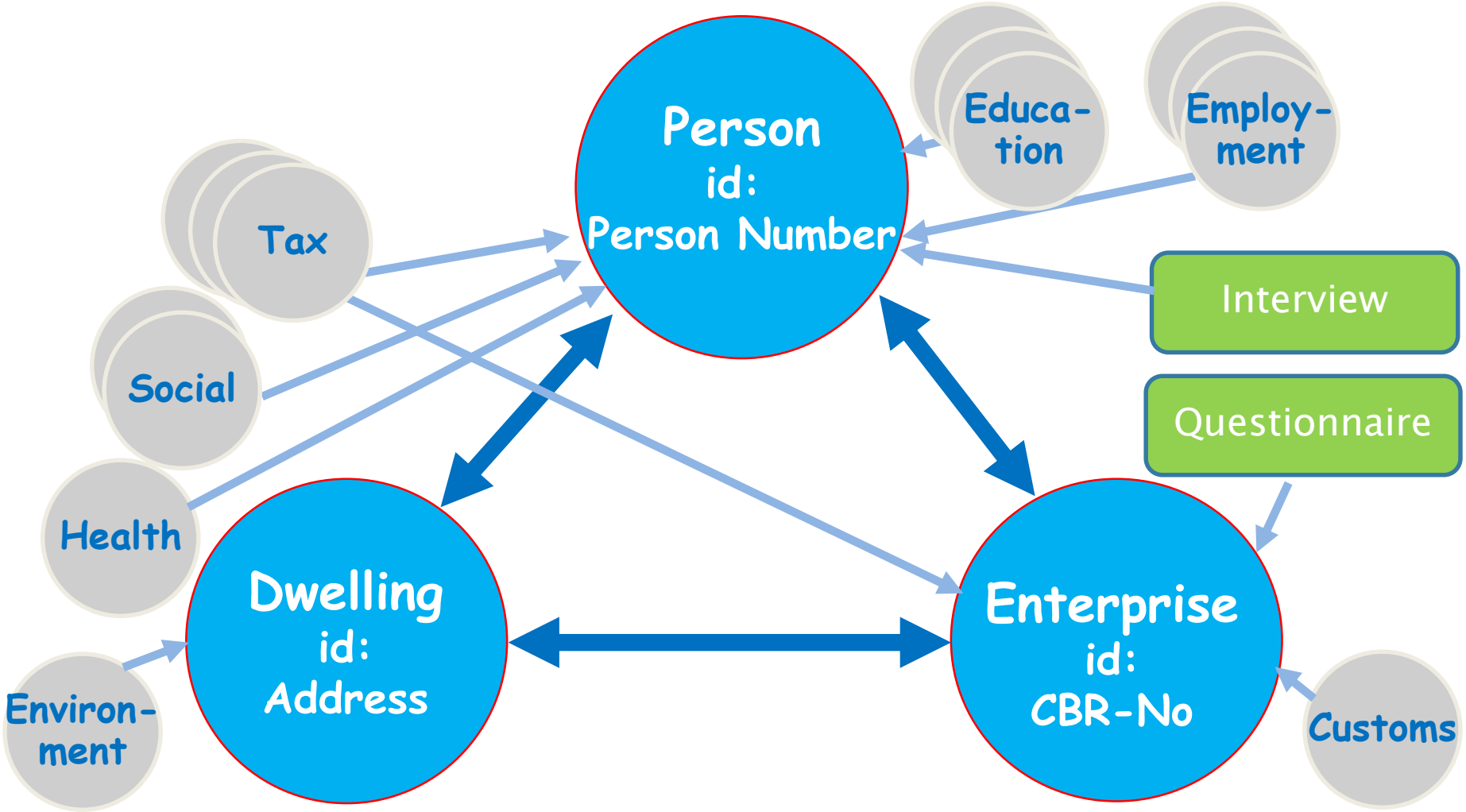
What are the advantages?

- Coverage
- Structure
- Coherence
- Accuracy
- Combining & linking
- Longitudinal studies
- Cost-effectiveness

How does the process work?



Which data – and how to link them?



How is data managed?

- Admin. registers are received, edited and organised in statistical registers
- Data is integrated across subject areas and time
- Statistical registers are combined to give comprehensive and new information
- Big investments – but even bigger returns!

How does data sharing materialize?

- Cooperation is key: Thorough discussions at strategic as well as practical level
- Focus on common interests / solutions – win-win situations and 'enabling incentives'
- Data sharing agreements (MoU) with data owners regarding content and conditions
- Raise awareness of the importance of high quality data registration
- Ensure adequate data protection

How is public trust maintained?

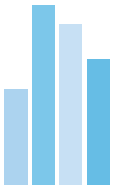
- Data protection legislation must be in place and complied with
- Data security rules and systems must be in place and used
- Individual data records/items (‘micro data’) are confidential
- Data on physical persons are given random identifiers
- No info about individual persons or enterprises is sent back (‘one way traffic’)

What were the key messages?

- A strategic national digital infrastructure is possible – and holds great returns
- Re-use of data brings benefits for both the NSIs and admin. authorities – it is a ‘win-win’
- ‘One-way data traffic; two-way cooperation’
- Roles of the NSI:
  - Use admin. data; contribute to ‘systematisation’
  - Provide basis for evidence based policy making
  - Adhere to international standards and methods – it helps both users and producers of statistics

Approach to working with administrative data

STATISTICS DENMARK



Parallel dimensions

Essential transitions within the NSI

External cooperation

Essential transitions within the NSI

Building a register based statistical model

- Aim for a statistical model based on registers using unique identifiers
- Think big – start small!
- Ensure that unique identifiers are recognized at highest national level and implemented in national law
- Develop specifications and methodology
- Optimize production processes, and identify new ways to meet users needs taking advantage of the register based data model

Organization of the NSI

- Prepare your NSI for a digital transformation
- Make a strategy for the transformation proces
- Upgrade staff skills to be prepared for processing and recoding of large data sets
- Prepare budgets for possible investments in hardware, software, it-infrastructure, data security and staff capacity

Strengthening IT-capacity

- Make a strategy for upgrading the it-capacity
- Plan for reaching the necessary level of capacity in terms of power- and internet supply, servers, hardware, software, it-infrastructure, data security and staff
- Build infrastructure to receive, store, proces and disseminate statistics based on administrative registers
- Specifications and design of databases must be aligned with the specifications of MoUs on data deliverances
- Data security and data confidentiality is a high priority!

External cooperation

Awareness

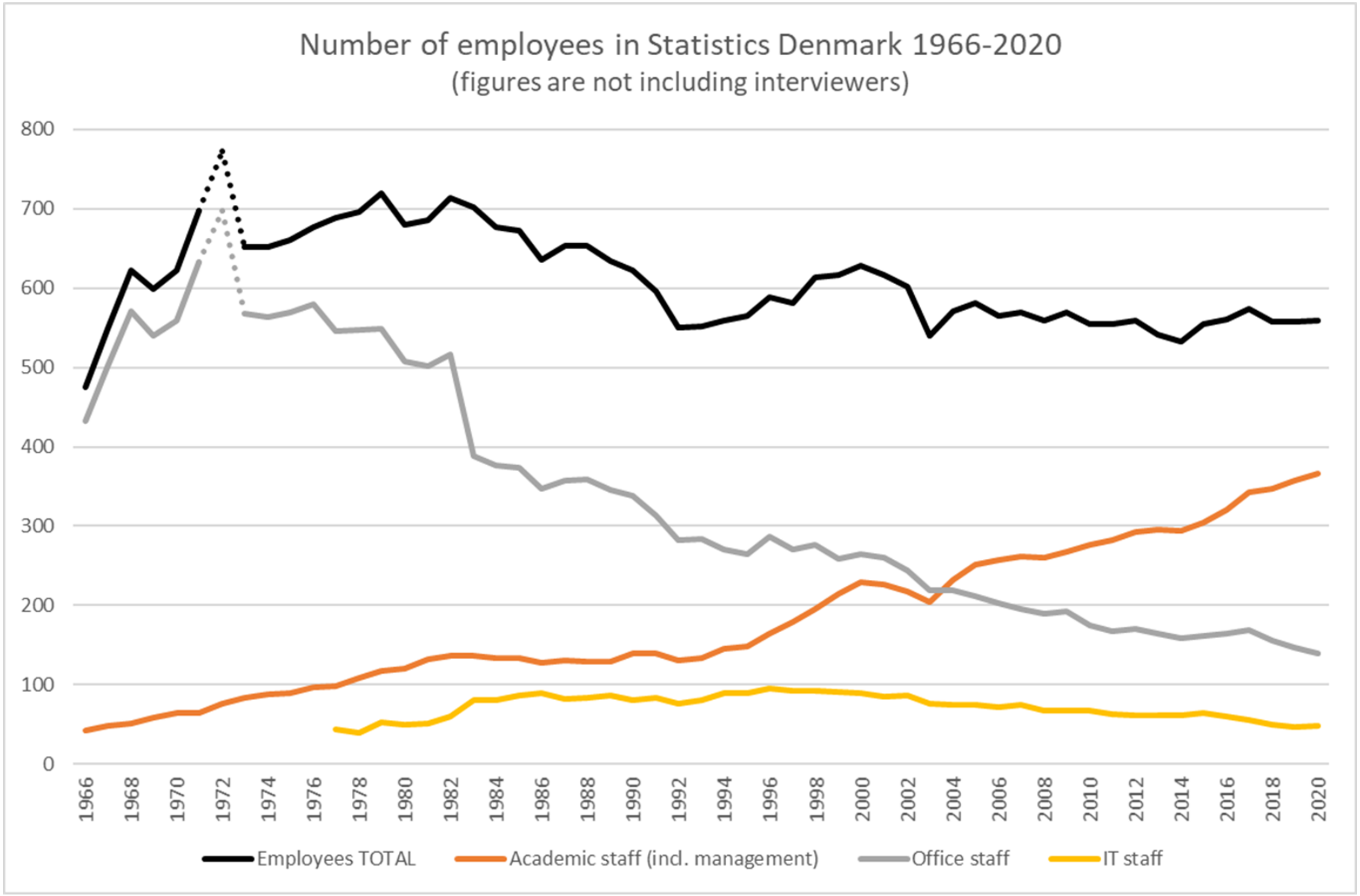
- Make a national demand for timely high quality statistics based on reuse of already existing data sources
- Benefits for the data suppliers to receive value added data output and statistics based on their own administrative data input
- Reward of investments in terms of rapid low-cost data collections
- Success stories: How can national statistics based on administrative data support evidence based policy making?

Legal mandate

- Regulations on exchange of data from public authorities to the NSI (and vice versa) should be implemented in national law
- Regulations on data security and protection of data confidentiality is an absolute need
- Who is financing the developments of data deliverances?
  - *The NSI develops receiving system and databases*
  - *The data supplier develops formatted dataset*
- Strive for a national agreement of unique identifiers to link data from administrative sources

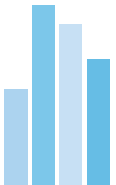
Data exchange agreements

- Make MoUs with each data supplier (think big – start small)
- Stakeholder meetings
- Identify what data is available, what can be created, what is 'need-to-have'?
- Specify variables and content of records and data
- Get access to raw data to be transformed to statistics by the NSI
- Identify sustainable data sources in order to produce time series
- Evaluate existing MoUs frequently


Number of employees in Statistics Denmark 1966-2020
(figures are not including interviewers)
800
700
600
500
400
300
200
100
0
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
Employees TOTAL
Academic staff (incl. management)
Office staff
IT staff

Development of a Memorandum of Understanding (MoU)

STATISTICS DENMARK



Generic proces for drafting a data provider agreement (MoU)

What are the purposes:

- Documentation of agreements
- Description af the overall purpose
- Description of data: data available, limitations and points of attention
- The legal basis for data transfer
- Responsibility for data processing
- Financial aspects: development / ongoing operation
- Technical aspects of data transfer
- Schedules and frequency

Steps in implementing the process

Step 1

- **Needs clarifications**
  Clarification of data available in relation to needs

Step 2

- **The formal basis**
  Clarify the formel basis of data exchange

Step 3

- **Data description in detail**
  Data description in detail and practical issues related to exchange of data

Step 4

- **Receiving of data**
  Determine methods of data transfer

Step 1

• Needs clarifications

Clarification of data available in relation to needs

Mapping of organisations with potentials as data sources :

- The origin of data: Administrative decisions /administration / use / traffic etc.
- Data quality: missing observations and bias in relation to statistical use
- Data availability / Data transfering

Acknowledge the data suppliers approach :

- Acknowledge that the supplier can have good reasons for not wanting data to be shared for statistical purposes.

DST's experience:

- Data are not fully comprehensive
  - Data has bias,
  - Data does not show what you expect
  - It is difficult to modify it-systems for export of data
  - Who is going to pay ?
  - And what is the legal basis for transferring data ?

It is very much a question of trust.
Go to the supplier with an open attitude to find solutions that address the concerns.
Find the balance between "best practice" and what is possible.

Step 2

The formal basis

Clarify the formel basis of data exchange

This step is based on the assumption that the supplier have potentially useful data.

The legal basis: How can data be used?
Financial aspects: Development / operating costs
Schedules and frequency: Development tasks / ongoing operation

- The supplier expects that the NSI has the legal basis and pays the bill !

The necessary tasks in the supplier's organisation are not the tasks with the highest priority.
It is not core business for the supplier.
If possible, find examples showing that the use of data for statistical purposes can be an advantage.

DST's experience:

- Go step by step – a small step is better than no step.
- Success is about trust.

Step 3

• Data description in detail

Data description in detail and practical issues related to exchange of data

Prepare an interface description of data with the following content:

- Overall description af data
- Schedules and frequency
- Periods covered by data (daily / weekly / monthly /yearly)
- Detailed description of variables: Name of variables / type of variables (char / num) / length / examples of variables / description

DSTs experience :

- A big task that takes time

Step 4

• Receiving of data

Determine methods of data transfer

Describe methods of data transfer:

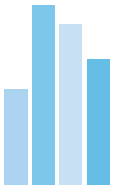
- ”File transfer” or ”system integrated” transfer
- Receiving enviroment: Standard systems, it-applications, machinery etc.

DSTs experience :

- Keep an ongoing dialogue with data suppliers about upcoming changes in it-applications, that can have an impact on the data transfer

Preconditions, challenges and pitfalls

STATISTICS DENMARK



Issues and obstacles

- Statistics Denmark has over the past many years worked with projects, which to a more or lesser extent have aimed to increase the use of administrative data.

- The issues and challenges presented in here are based on our experience from several countries.

Challenge #1:
Lack of support from other public authorities

- **A system-wide approach**
  - Building up a system of using administrative data is a long and iterative process (many meetings, discussions, and institutions involved)
  - There will however be a potential “ketchup”-effect, which will get even stronger if the message about data cooperation comes from not only the NSI but also from the national government and key government institutions
- **Legal issues**
  - Legal obstacles for sharing data are not unusual
  - Even though the legal obstacles are removed by introducing a new legal set-up, other arguments can be used to prevent data from being transmitted to the NSI, especially from those institutions who do not want to share their data
- **Trust**
  - Lack of trust towards the NSI in how they handle data and their political independence (or not) can be a burden
- **Power over data**
  - Sharing of administrative data can be seen as giving up power
    Some institutions are therefore not interested in sharing their data with the NSI
- **Costs**
  - It costs to connect an administrative database to the NSI in terms of defining data format, selecting relevant data, and developing a system for regular data transmissions
  - If there are cases of conflicting interests, lack of understanding or disagreements, the issue of who is going to pay can be a huge obstacle for data transmission

Challenge #2:
Quality of the administrative data

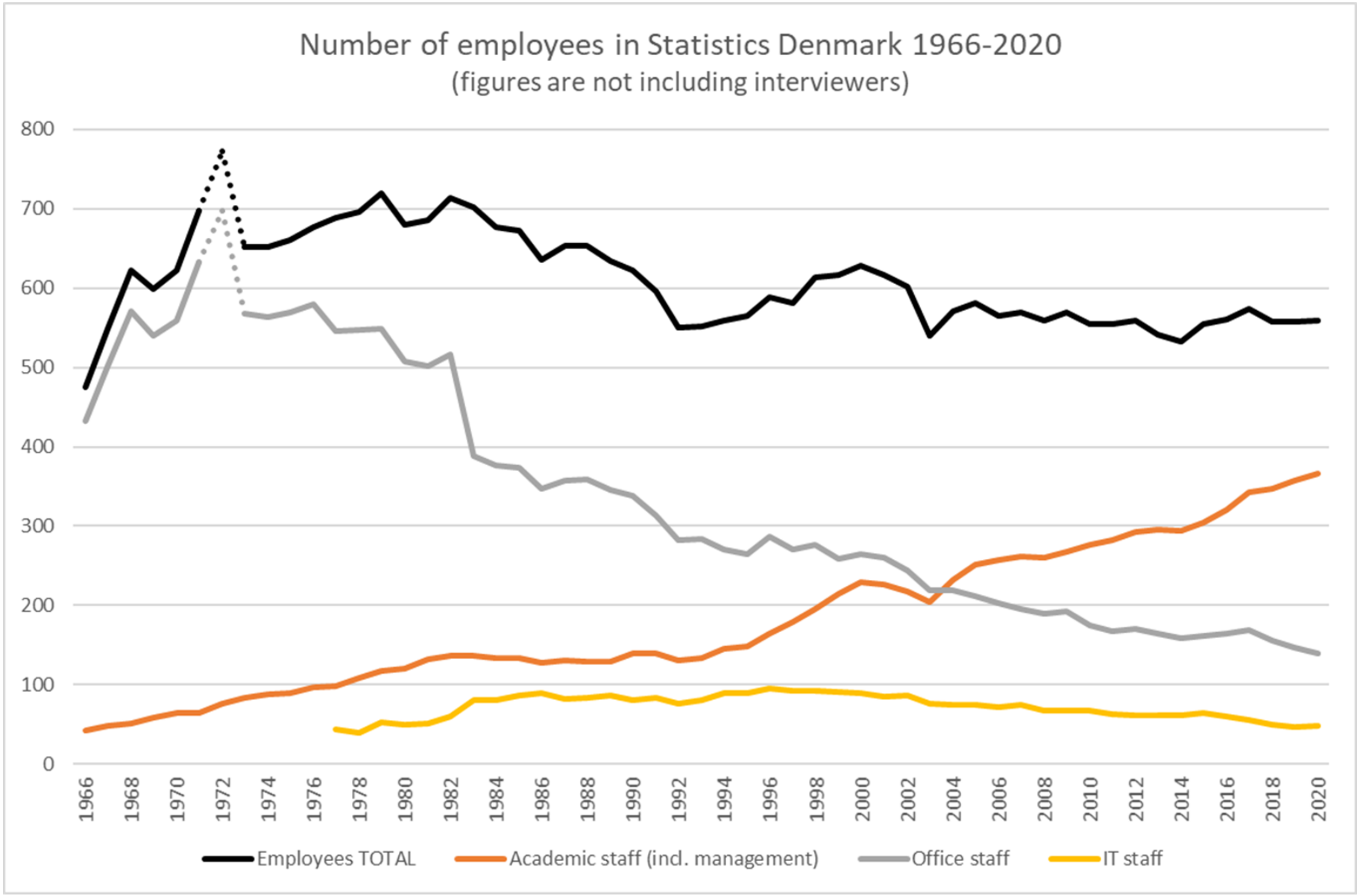
- **Understanding the importance of good data**
  - Data providers need to understand the flow of data and the importance and usability of high quality data
  - Ideally the data must be initially entered correctly into the system
    - Correct data from the beginning means less work of correcting and manipulating data, less misunderstandings about the data, and a high degree of comparability between the original source of the administrative data and the final statistical product
  - It can however be a demanding task to raise awareness about the importance of high quality registrations
  - Data entry institutions are often far away from the capital both in geographical and mental terms

- **Defining the data needed**
  - Be aware that access to administrative data is not just a matter of getting access to everything in the data provider's system
  - Data should be carefully selected and described in order for the NSI to understand what data they are dealing with
    - The work of describing, selecting and documenting administrative data from an external data provider can easily be underestimated
  - Statistical time series based on administrative data can provide new and unknown problems
    With administrative data, the NSI has no direct control over the data collection
    - Periodically, administrative data may be affected by regulatory changes and changes in registration practices. This requires new statistical methods and processes - and not least good cooperation between institutions

Challenge #3:
Reorganizing the statistics production and the NSI

- **Regional offices role in a new set-up**
  - Survey-based statistical production often implies decentralized NSI's with a number of regional offices. The role of regional offices is reduced when introducing a system-wide change towards use of administrative data
    - The NSI thus needs to rethink its organization and maybe relocate staff

- **New staff skills needed and how to train**
  - Using administrative data for statistical purposes requires new skills in the staff and thus training. Human resource efforts need to be initiated to define the skills needed and how to ensure existing and new staff has such skills
    - These skills will often require IT competences that are highly difficult to recruit and retain


Number of employees in Statistics Denmark 1966-2020
(figures are not including interviewers)
800
700
600
500
400
300
200
100
0
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
Employees TOTAL
Academic staff (incl. management)
Office staff
IT staff

Challenge #4:
IT challenges

- **Data security**
  - Integrating administrative data from a large number of public authorities increases the risk of misuse by hackers (or even internal staff)
  - Measures to address the security risks that the digitation will imply must be set up
- **Back-up systems**
- **Reliable power and internet supply**
  - Poor power and internet supply is not uncommon, and can lead to a disaster for the organization if the statistical production relies on electronic data transmissions and digital communication
- **Standardized software across the organization for processing data**
  - Having only one or few software solutions will in the long run make internal training and staff rotations easier
- **IT hardware**
  - A general challenges for many NSIs is server capacity and modern computers to handle large amounts of data. Many countries have low server capacity and need to enlarge their capacity if they are to handle large amounts of administrative data
- **IT staff and funding**
  - IT is often understaffed and underfunded. Establishing a statistical production based on wider use of administrative data requires understanding the importance of qualified IT staff, up-to-date and sufficient hardware as well as strategically chosen and systemized software
- **Cooperation between statistical units and IT department**
  - A strong cooperation between the statistical units and the central IT-department will be needed