Register-based statistics production

Administrative data used for statistical purposes

Bo Sundgren

2010

Part 1

Basic definition of a register
(register in a strict and narrow sense)

- A register is an authorized, up-to-date list of all objects belonging to a certain population
- The objects are uniquely identified by an authorized identifier, such as person number for persons, orgaisation number for enterprises and other organisations, etc
- In addition to the identifier, a register may contain additional basic and up-to-date information about the objects, such as name (not necessarily unique) and location and other contact information, e.g. address and telephone number

Cf. Wallgren&Wallgren, page 4. But many other definitions are used throughout the book, often vaguely or implicitly.

Extended register

(register in a broader sense)

- An authorized, up-to-date list of all objects belonging to a certain population
- The objects are uniquely identified by an authorized identifier
- In addition to the identifier, a register may contain additional basic and up-to-date information about the objects, such as name (not necessarily unique) and location and other contact information, e.g. address and telephone number
- Furthermore, an extended register may contain links to other registers and data sources, as well as additional information from those other sources

Administrative registers and statistical registers

- An *administrative register* is a register used for administrative purposes, e.g. by a government agency or an enterprise
- A *statistical register* is a register used for statistical purposes, e.g. by a statistical agency or an enterprise
- A statistical register may be created from one or more administrative registers, sometimes in combination with information from other sources, such as other administrative sources, e.g. administrative databases, or other statistical sources, e.g. surveys and statistical databases
- A basic function of a statistical register is to serve as a (sampling) frame for (sample) surveys
- Extended statistical registers are also useful by themselves as sources for statistics production

Sources of data used by a statistical agency

- censuses
- surveys
  - sample surveys and total surveys
  - repeated surveys and ad hoc surveys
- administrative systems
  - administrative registers
  - other administrative data collections, e.g. databases
- archives
  - data already collected and stored by the statistical agency or by somebody else

About 97-99% of the data used by Statistics Sweden come from administrative sources. The cost of a value collected from an administrative source is about 1% of the cost of collecting it by a survey. Cf the virtual census of the Netherlands: 3 M€ vs 300 M€.

Wallgren&Wallgren, page 2:

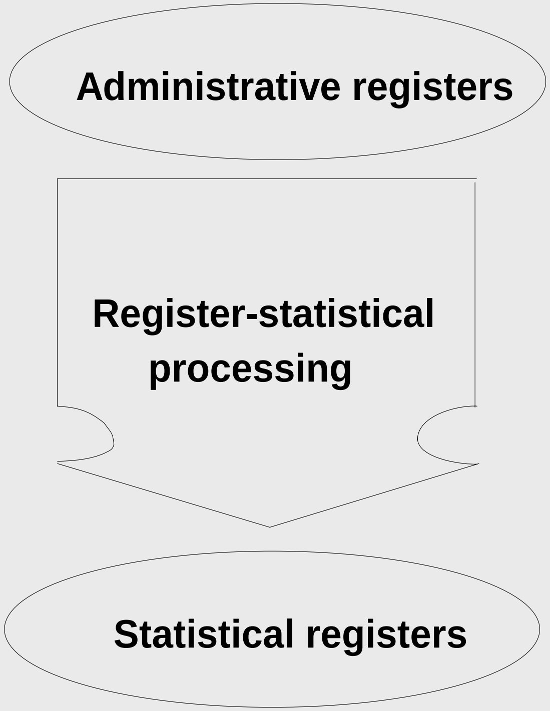
Chart 1.1 Four principles on how to use administrative data

1. A statistical office should have access to administrative registers kept by public authorities. This right should be supported by law as the protection of privacy.
2. These administrative registers should be transformed into statistical registers. Many sources should be used and compared during this transformation.
3. All statistical registers should be included in a coordinated register system. This system will ensure that all data can be integrated and used effectively.
4. Consistency regarding populations and variables are necessary for the coherence of estimates from different register-based surveys.

2. These administrative registers should be transformed into statistical registers. Many sources should be used and compared during this transformation.

Wallgren&Wallgren, page 5:

Chart 1.3 From an administrative register to a statistical register



Administrative registers

Register-statistical processing

Statistical registers

Register-statistical processing:

The administrative registers are processed so that objects and variables meet statistical needs:

- Editing of data
- Coding of variables
- Handling of missing objects and missing values
- Matching and selections
- Processing of time references
- Creating derived objects
- Creating derived variables

The statistical register is used to produce statistics

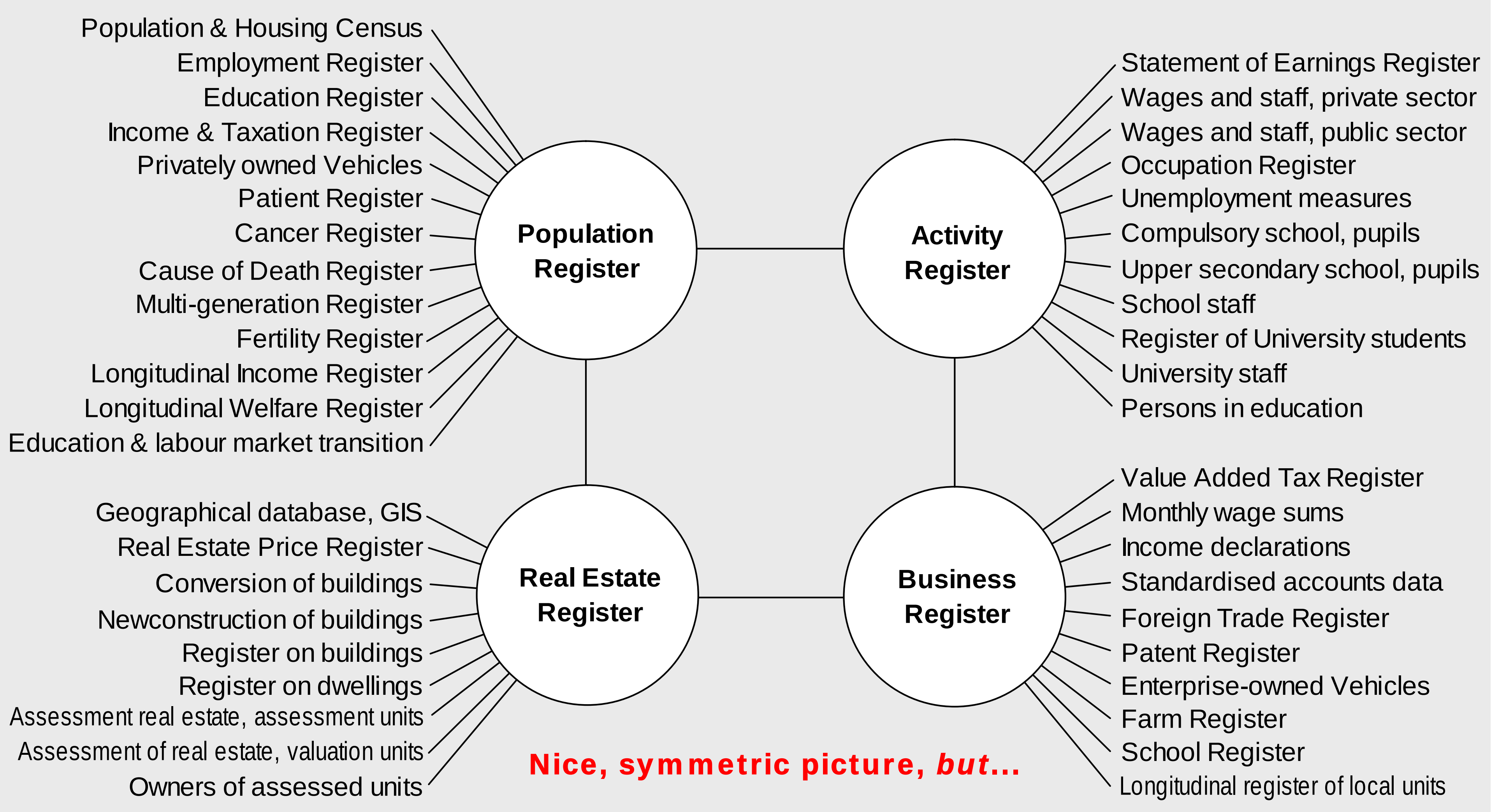
Quality assurance:

- Contacts with data suppliers
- Checking received data
- Missing values: reasons and extent
- Causes and extent of mismatches
- Evaluate quality of objects and variables
- Register maintenance surveys
- Inconsistencies are investigated and reported

3. All statistical registers should be included in a coordinated register system. This system will ensure that all data can be integrated and used effectively.

Wallgren&Wallgren Page 30: **(Note: A very broad and fuzzy register concept is used here.)**

Chart 2.10 A system of statistical registers – registers by object type and subject field



Population Register

Population & Housing Census
Employment Register
Education Register
Income & Taxation Register
Privately owned Vehicles
Patient Register
Cancer Register
Cause of Death Register
Multi-generation Register
Fertility Register
Longitudinal Income Register
Longitudinal Welfare Register
Education & labour market transition

Activity Register

Statement of Earnings Register
Wages and staff, private sector
Wages and staff, public sector
Occupation Register
Unemployment measures
Compulsory school, pupils
Upper secondary school, pupils
School staff
Register of University students
University staff
Persons in education

Real Estate Register

Geographical database, GIS
Real Estate Price Register
Conversion of buildings
Newconstruction of buildings
Register on buildings
Register on dwellings
Assessment real estate, assessment units
Assessment of real estate, valuation units
Owners of assessed units

Business Register

Value Added Tax Register
Monthly wage sums
Income declarations
Standardised accounts data
Foreign Trade Register
Patent Register
Enterprise-owned Vehicles
Farm Register
School Register
Longitudinal register of local units

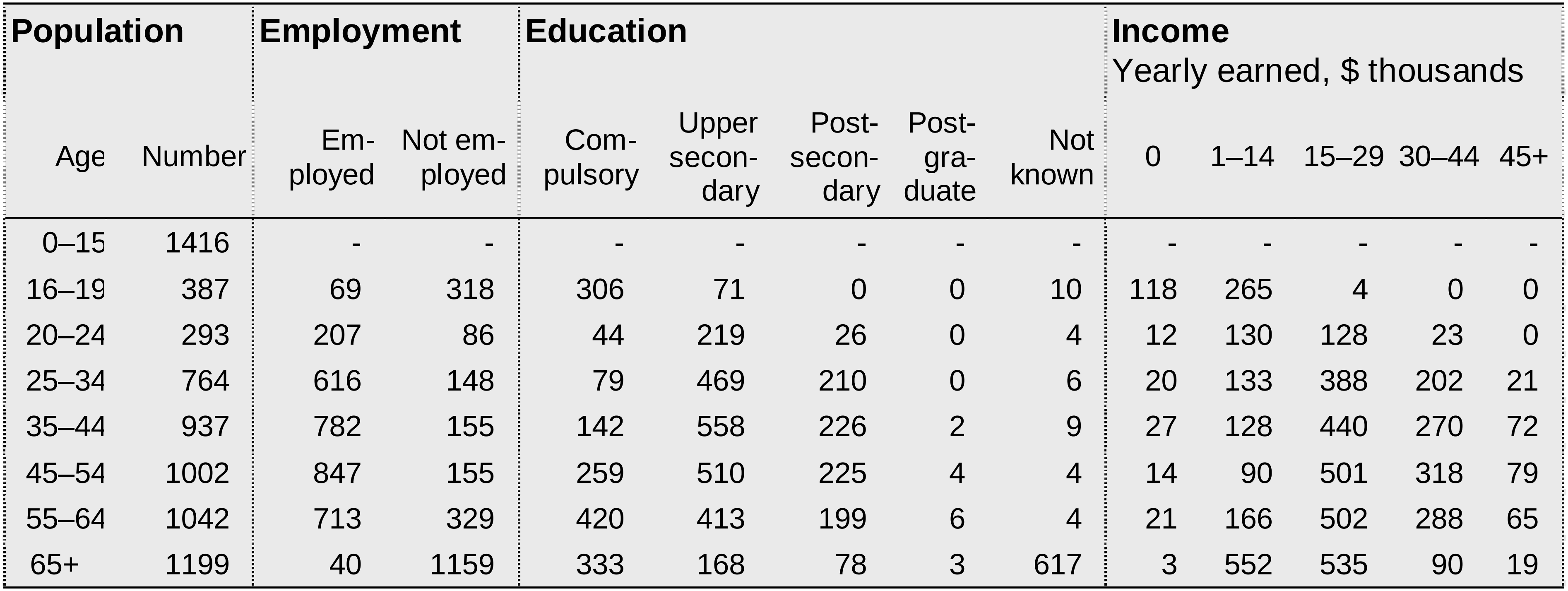
Nice, symmetric picture, *but*…

Note: This model is not a model of Statistics Sweden’s system, the model shows general possibilities.

4. Consistency regarding populations and variables are necessary for the coherence of estimates from different register-based surveys.

Wallgren&Wallgren, Page 32:

Chart 2.13 Register-based statistics for one small municipality in Sweden 2003



Population		Employment		Education					Income Yearly earned, $ thousands				
Age	Number	Employed	Not employed	Compulsory	Upper secondary	Post-secondary	Post-graduate	Not known	0	1–14	15–29	30–44	45+
0–15	1416	-	-	-	-	-	-	-	-	-	-	-	-
16–19	387	69	318	306	71	0	0	10	118	265	4	0	0
20–24	293	207	86	44	219	26	0	4	12	130	128	23	0
25–34	764	616	148	79	469	210	0	6	20	133	388	202	21
35–44	937	782	155	142	558	226	2	9	27	128	440	270	72
45–54	1002	847	155	259	510	225	4	4	14	90	501	318	79
55–64	1042	713	329	420	413	199	6	4	21	166	502	288	65
65+	1199	40	1159	333	168	78	3	617	3	552	535	90	19

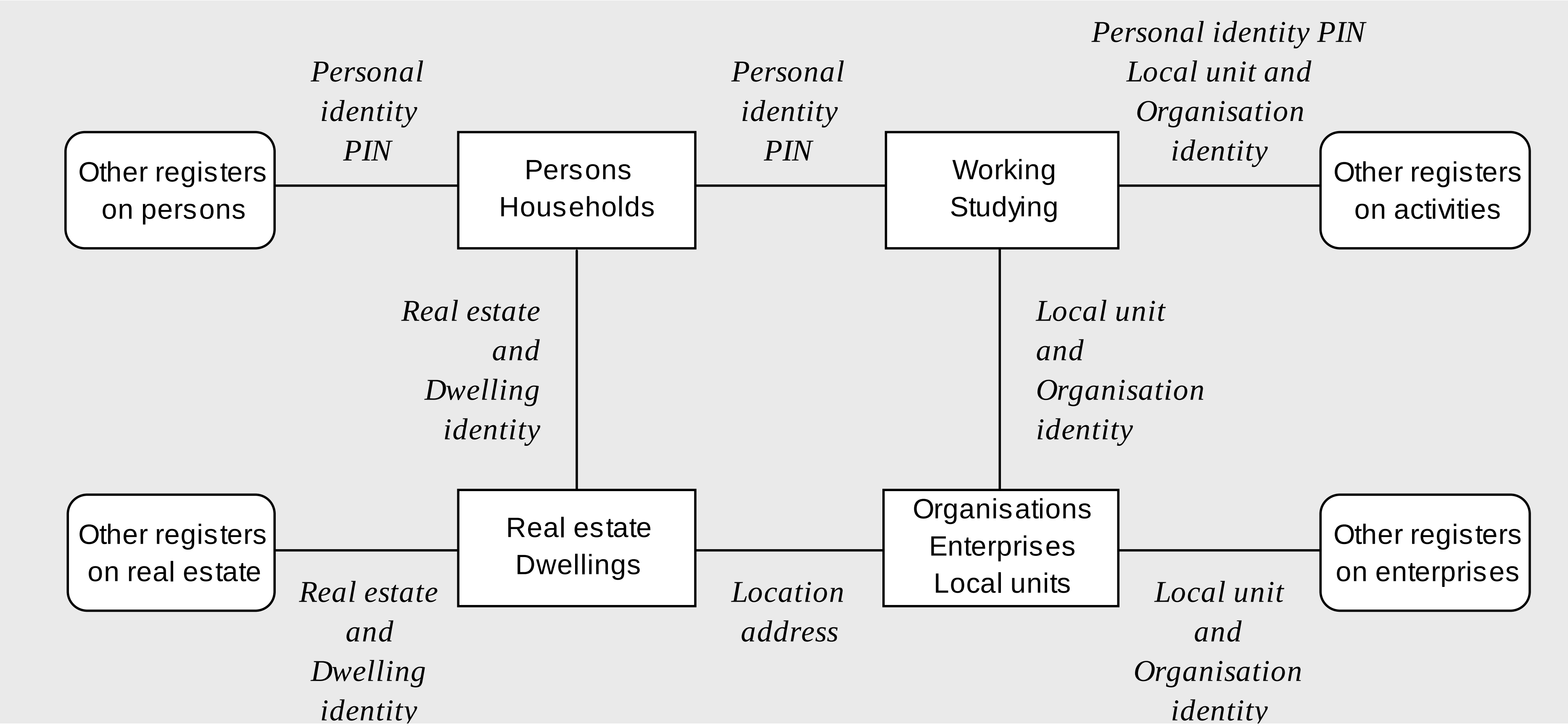
”Register system” or ”Register-based statistical system”

- A national statistical system could be defined as a system of statistical data about socio-economic conditions and developments in a country
- A system always consists of related (coordinated, integrated) subsystems, parts, or components
- The related data collections in a statistical system could come from different sources: censuses, surveys, registers, (other) administrative sources
- In a register-based statistical system, registers play an important role as a *”backbone”* and coordination tool within the statistical system

Chapter 2

Wallgren&Wallgren, Page 22:

Chart 2.5 A conceptual model of a register system of statistics on society



Personal identity PIN
Personal identity PIN
Personal identity PIN Local unit and Organisation identity
Other registers on persons
Persons Households
Working Studying
Other registers on activities
Real estate and Dwelling identity
Local unit and Organisation identity
Other registers on real estate
Real estate Dwellings
Organisations Enterprises Local units
Other registers on enterprises
Real estate and Dwelling identity
Location address
Local unit and Organisation identity

Wallgren&Wallgren, Page 23:

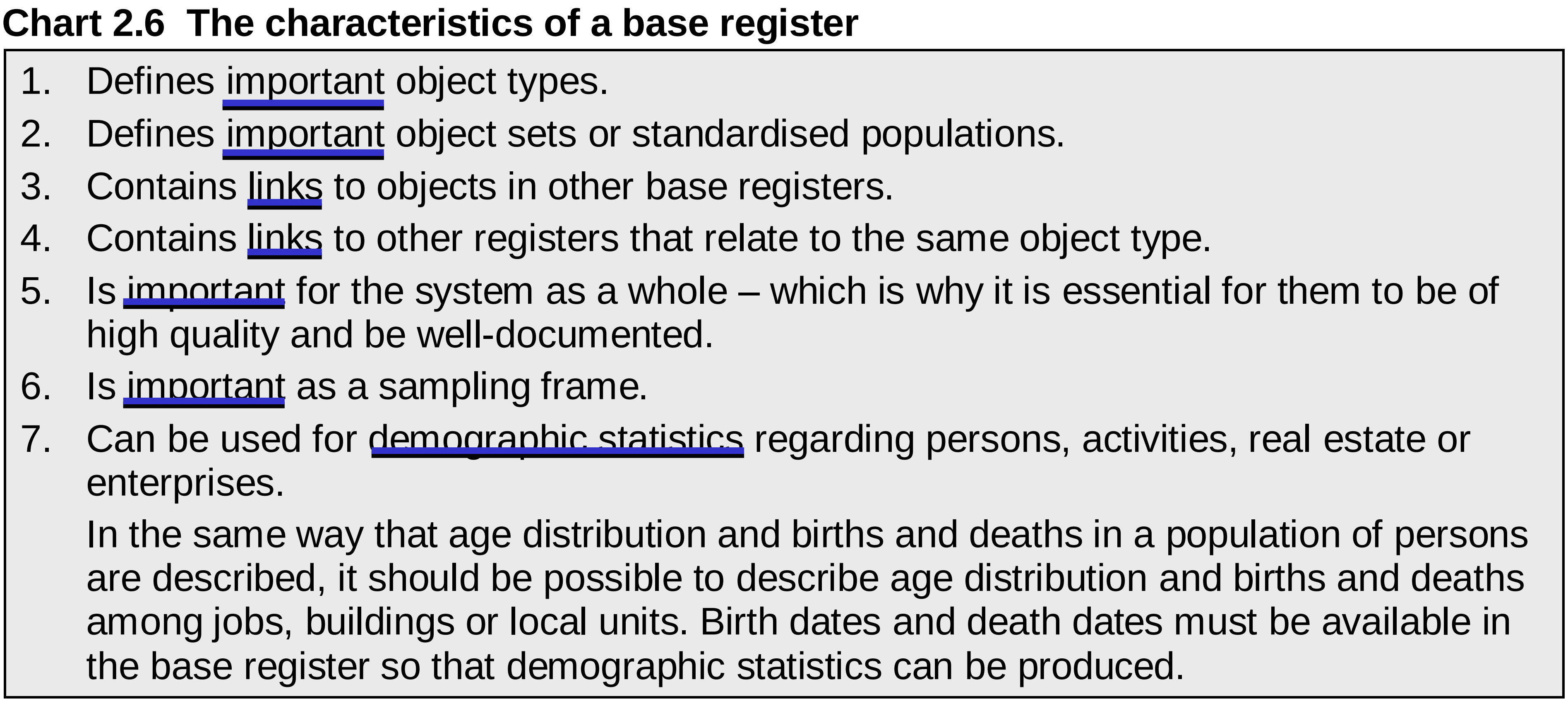


Chart 2.6 The characteristics of a base register

1. Defines important object types.
2. Defines important object sets or standardised populations.
3. Contains links to objects in other base registers.
4. Contains links to other registers that relate to the same object type.
5. Is important for the system as a whole – which is why it is essential for them to be of high quality and be well-documented.
6. Is important as a sampling frame.
7. Can be used for demographic statistics regarding persons, activities, real estate or enterprises.

   In the same way that age distribution and births and deaths in a population of persons are described, it should be possible to describe age distribution and births and deaths among jobs, buildings or local units. Birth dates and death dates must be available in the base register so that demographic statistics can be produced.

Backbone of the Swedish register system

Population register – exists and works well; person, household?

Activity register – exists; note: relational objects

Business register – exists; object type: (part of) organisation

Real estate register – exists; dwellings underway

Chart 2.8 The relation between registers on persons, activities and enterprises

Population Register – Persons	
Person	**Wage sum**
PIN1	450 000
PIN2	210 000
PIN3	270 000

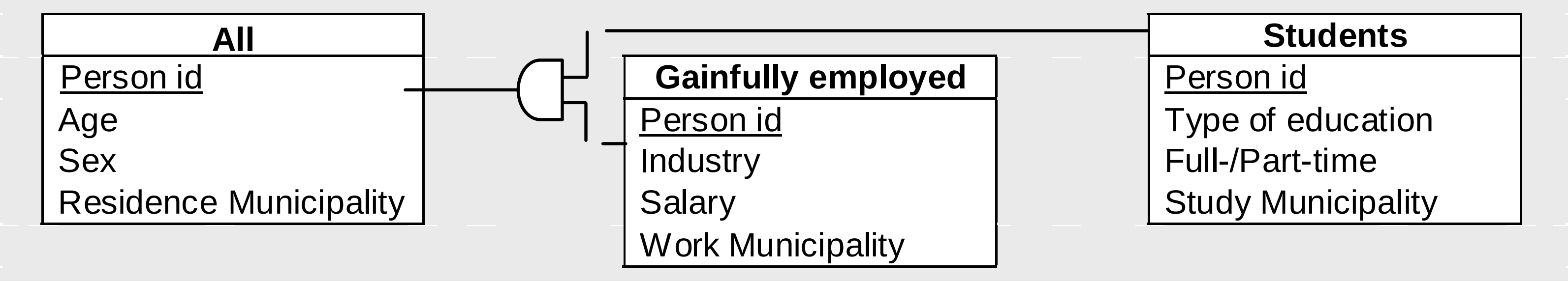
Activity Register – Jobs			
Job	**Person**	**Local unit**	**Wage sum**
J1	PIN1	LU1	220 000
J2	PIN3	LU1	180 000
J3	PIN1	LU2	230 000
J4	PIN2	LU2	210 000
J5	PIN3	LU2	90 000

Business Register – Local units	
Local unit	**Wage sum**
LU1	400 000
LU2	530 000

The Activity Register contains the bi-variate distribution and the Business and Population Registers contain marginal distributions

Conceptual models and data models (cf previous course)

Chart 3.5 A database on individuals with three database tables
A. Conceptual database model



All
Person id
Age
Sex
Residence Municipality

Gainfully employed
Person id
Industry
Salary
Work Municipality

Students
Person id
Type of education
Full-/Part-time
Study Municipality

Chart 3.5 A database on individuals with three database tables
B. Example of content in the database

All

Person id	Age	Sex	ResMun.
PIN1	20	F	0586
PIN2	23	M	0586
PIN3	31	M	0586
PIN4	32	F	0586
PIN5	33	M	0586
PIN6	40	F	0586
PIN7	59	F	0586
PIN8	65	M	0586
PIN9	71	F	0586

Gainfully employed

Person id	Industry	Salary	WorkMun.
PIN2	G	52 000	0586
PIN3	G	287 000	0580
PIN4	A	193 000	0586
PIN6	D	291 000	0586
PIN7	D	314 000	0580

Students

Person id	Educ.Type	Full/Part-time	StudMun.
PIN1	AdultEduc	100	0586
PIN2	Univ	100	0580
PIN5	Univ	100	0580

Example

Chart 3.6 Two data matrices for different statistical purposes

A. Data matrix: Employment register

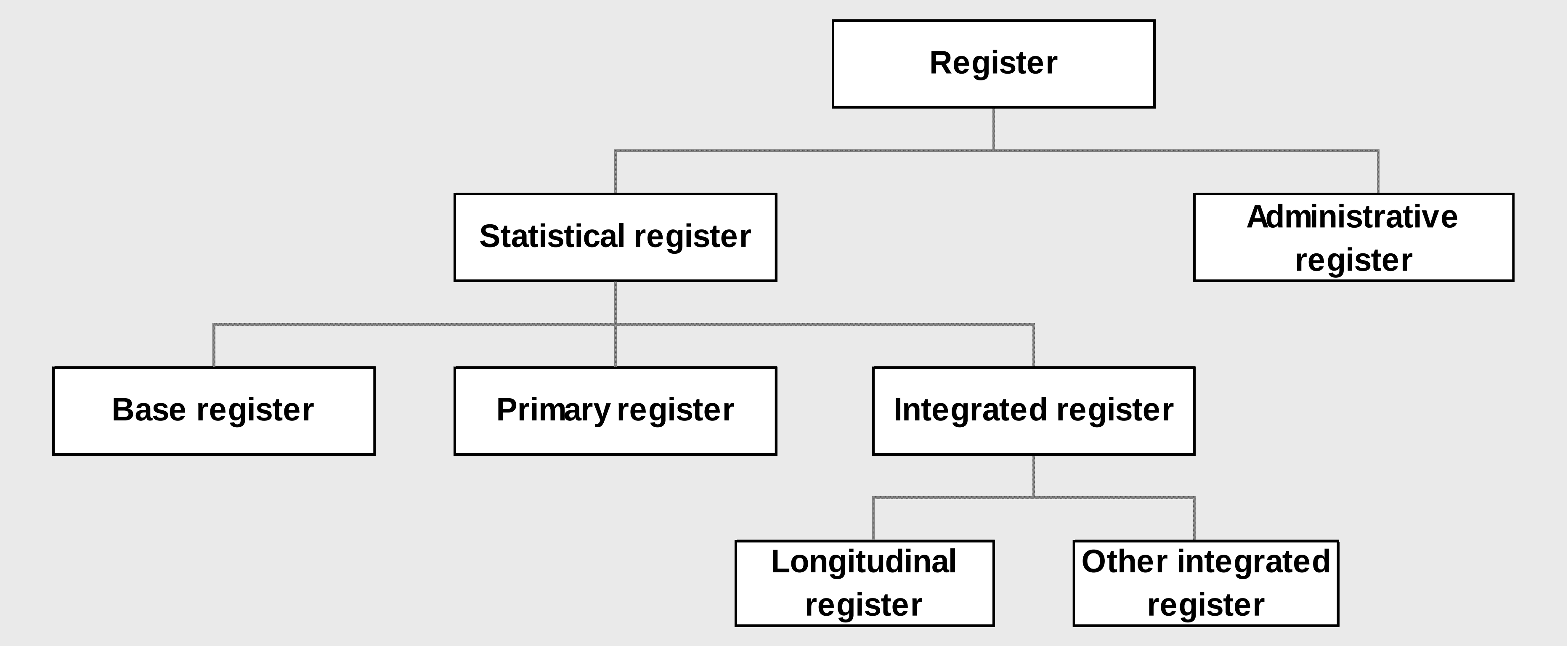
Person	Age	Sex	Employed	Industry	Salary
PIN1	20	F	No	null	0
PIN2	23	M	Yes	G	52 000
PIN3	31	M	Yes	G	287 000
PIN4	32	F	Yes	A	193 000
PIN5	33	M	No	null	0
PIN6	40	F	Yes	D	291 000
PIN7	59	F	Yes	D	314 000
PIN8	65	M	No	null	0
PIN9	71	F	No	null	0

B. Data matrix: Commuting register

Person	ResMun.	WorkMun.	StudMun.	Commuting
PIN1	0586	null	0586	0
PIN2	0586	0586	0580	1
PIN3	0586	0580	null	1
PIN4	0586	0586	null	0
PIN5	0586	null	0580	1
PIN6	0586	0586	null	0
PIN7	0586	0580	null	1
PIN8	0586	null	null	0
PIN9	0586	null	null	0

Exercise: Reconstruct the conceptual models behind the two data matrices!

Chart 3.8 Different kinds of registers



- Base register: important object type, important population, links to other objects or to same objects in other register, important as a frame, important for the system as a whole, can be used for demographic statistics
- Primary register: directly based on at least one administrative register
- Integrated register: created from other statistical registers only
- Longitudinal register: integrated register, where it is possible to follow objects over time

Registers and time

- Current stock register: updated with all information on currently active/live objects; used as frame for (sample) surveys
- Register referring to a specific point in time, e.g. turn of the year, updated *after* the point in time when reports about ”all” events have arrived; used for register-based surveys
- Calendar year register: containing all objects that have existed at any point during a specific year; used for register-based surveys
- Events register: all events during a specific period
- Historical register: all events, used for longitudinal surveys
- Longitudinal register: all events during a time period

3.2.4 Registers and time page 47-48

1. The *current stock register* …is used as frame population for sample surveys or censuses.
2. The *register referring to a specific point in time*, … Is used for register-based surveys.
3. The *calendar year register* .. It is used as register populations for register-based surveys.

Chart 3.9 Calendar year register for 2002

Object identity	Existed 1/1	Added	Ceased to exist	Existed 31/12	Other variables
Idnr 1	Yes	-	20020517	No	…
Idnr 2	Yes	-	-	Yes	…

4. The *events register* for a specific period, It is used in register-based surveys.

Chart 3.10 Events register for 2002 regarding change of address

Object identity	Address 1/1	Date of change of address	New address
Idnr 1	Address 11	20020517	Address 21

5. The *historical register* It is used for longitudinal surveys.

Chart 3.11 Historical register regarding change of address

Object identity	From address	Date of change of address	To address
Idnr 1	Born	19670517	Address 1
Idnr 1	Address 1	19810606	Address 2

6. A *longitudinal register* for a period of time

Chart 3.12 Longitudinal register for 2000–2002

Object identity	Existed 1/1/2000	Added	Ceased to exist	Income 2000	Income 2001	Income 2002
Idnr 1	Yes	-	20010517	183 450	97 600	-
Idnr 2	Yes	-	-	273 500	281 360	258 340

Derived objects
Example 1: Household

- http://stats.oecd.org/glossary/detail.asp?ID=1255
- "A household is a small group of persons who share the same living accommodation, who pool some, or all, of their income and wealth and who consume certain types of goods and services collectively, mainly housing and food." (SNA 4.132 [4.20]).

Derived objects
Example 2: Part of organisation

- “An organisation is a unique framework of authority within which a person or persons act, or are designated to act, towards some purpose.” (OECD)
- “An enterprise is an institutional unit in its capacity as a producer of goods and services; an enterprise may be a corporation, a quasi- corporation, a non-profit institution, or an unincorporated enterprise.” (SNA)
- “An institutional unit is an economic entity that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transactions with other entities.” (SNA)
- Eurostat: The Council Regulation ((EEC), No. 696/93 of 15 March 1993) on statistical units for the observation and analysis of the production system in the Community lays down a list of eight (types of) statistical units:

— the Enterprise;
— the Institutional Unit;
— the Enterprise Group;
— the Kind-of-activity Unit (KAU);
— the Unit of Homogeneous Production (UHP);
— the Local Unit;
— the Local Kind-of-Activity Unit (local KAU);
— the Local Unit of Homogeneous Production (local UHP).

Different kinds of variables

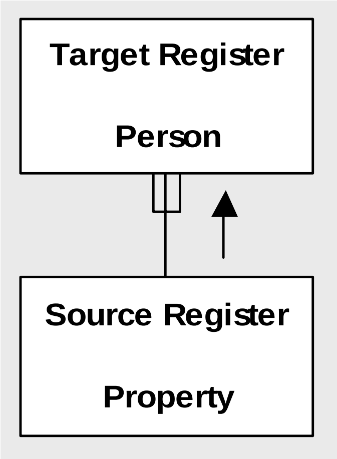
- Stocks and flows
  - a stock variable is measured *at one specific time*, and represents a quantity existing at that point in time, which may have been accumulated over time
  - a flow variable is measured over an interval of time and would be measured *per unit of time*
- Quantities and qualities
  - a quantitative variable has values that can be summarised in a meaningful way
  - a qualitative variable has values (codes) that cannot be summarised in a meaningful way, although the codes may sometimes be numerical (sex = 1, 2)
- Open values (free text) vs predefined codes
- Single-valued vs multi-valued variables

Derived variables

1. Variables derived by grouping values and dividing into class intervals

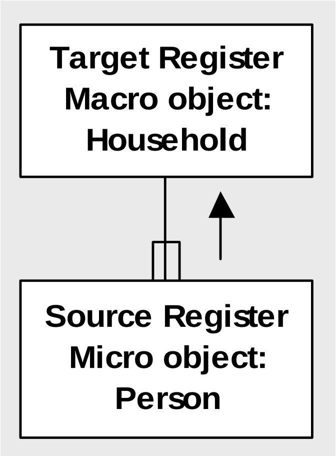
2. Variables derived by arithmetic operations using variables in the data matrix

3. Variables derived by adjoining



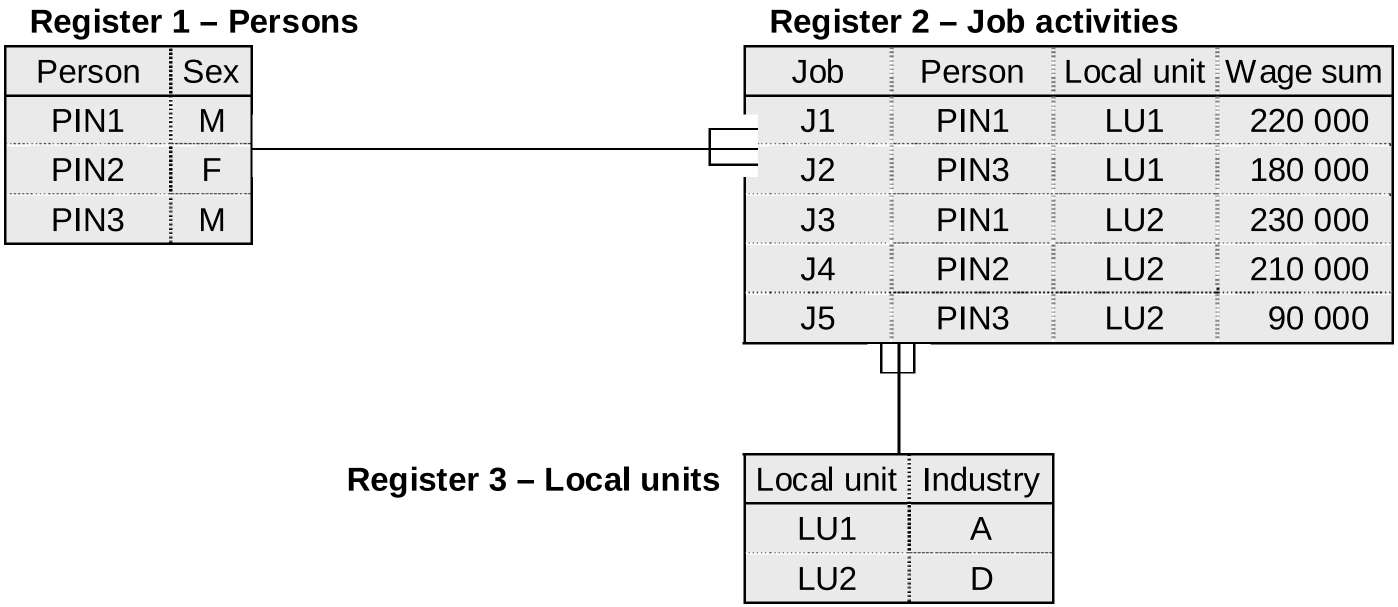
This involves creating a derived variable in a register using variables from another register. The objects in the first register can be linked to objects in the second register in a *one-to-one* relationship or a *one-to-many* relationship. This means that every object in the source register can be linked to one or many objects in the target register. Using this relationship, variables in the source register can be adjoined to the objects in the target register.

4. Variables derived by aggregation



This involves creating a derived variable in a register using variables from another register. The objects in the source register can be linked to the objects in the target register using a *many-to-one* relationship. One or many objects in the source register can be linked to one object in the target register. It is possible to aggregate values, in a way that is relevant for the survey, for the *micro objects* in the source register that is linked to the respective *macro object* in the target register.

Chart 3.14A The relations between persons, activities and local units



Register 1 – Persons

Person	Sex
PIN1	M
PIN2	F
PIN3	M

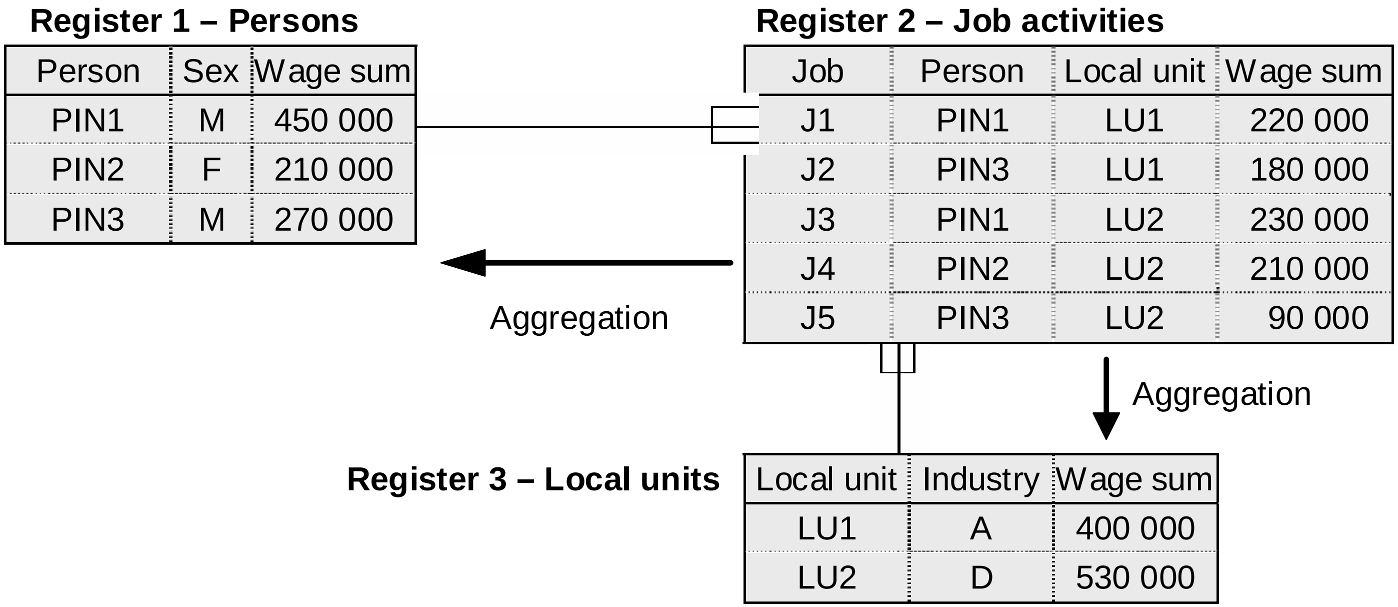
Register 2 – Job activities

Job	Person	Local unit	Wage sum
J1	PIN1	LU1	220 000
J2	PIN3	LU1	180 000
J3	PIN1	LU2	230 000
J4	PIN2	LU2	210 000
J5	PIN3	LU2	90 000

Register 3 – Local units

Local unit	Industry
LU1	A
LU2	D

Chart 3.14B Wage sums for persons and local units created by aggregation



Register 1 – Persons

Person	Sex	Wage sum
PIN1	M	450 000
PIN2	F	210 000
PIN3	M	270 000

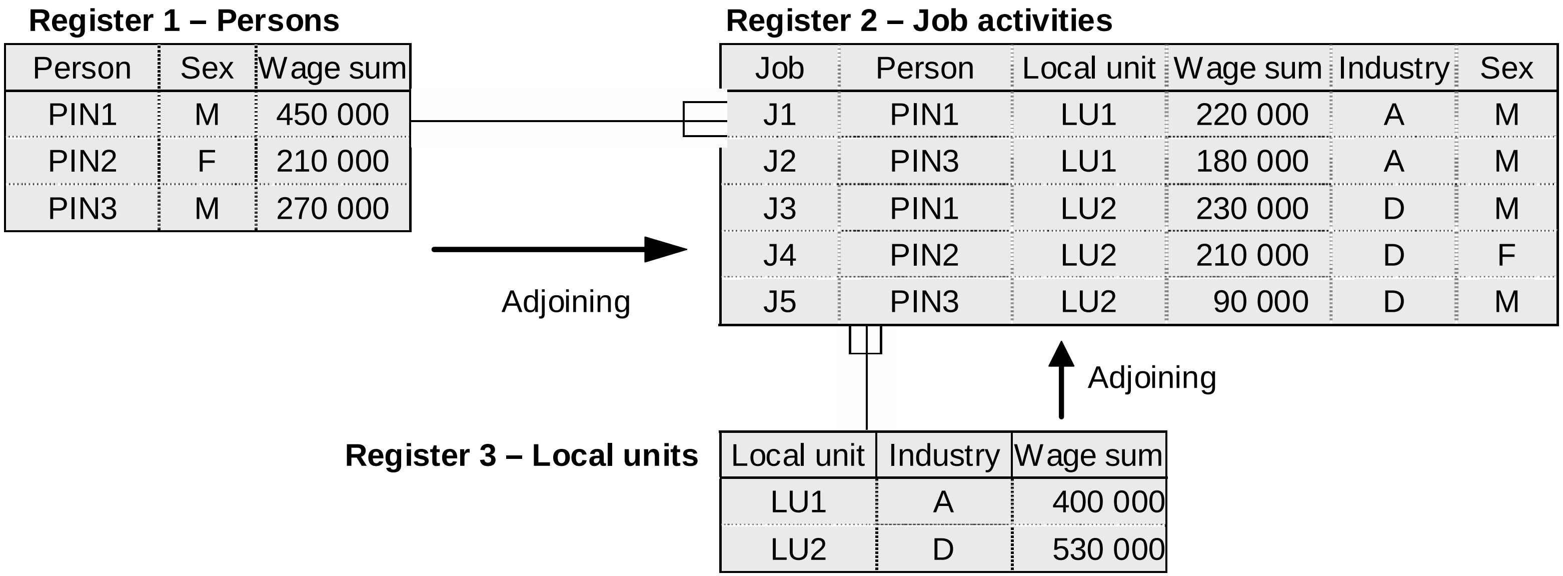
Register 2 – Job activities

Job	Person	Local unit	Wage sum
J1	PIN1	LU1	220 000
J2	PIN3	LU1	180 000
J3	PIN1	LU2	230 000
J4	PIN2	LU2	210 000
J5	PIN3	LU2	90 000

Register 3 – Local units

Local unit	Industry	Wage sum
LU1	A	400 000
LU2	D	530 000

Chart 3.14C Industry and sex as derived variables for jobs created by adjoining



Register 1 – Persons

Person	Sex	Wage sum
PIN1	M	450 000
PIN2	F	210 000
PIN3	M	270 000

Register 2 – Job activities

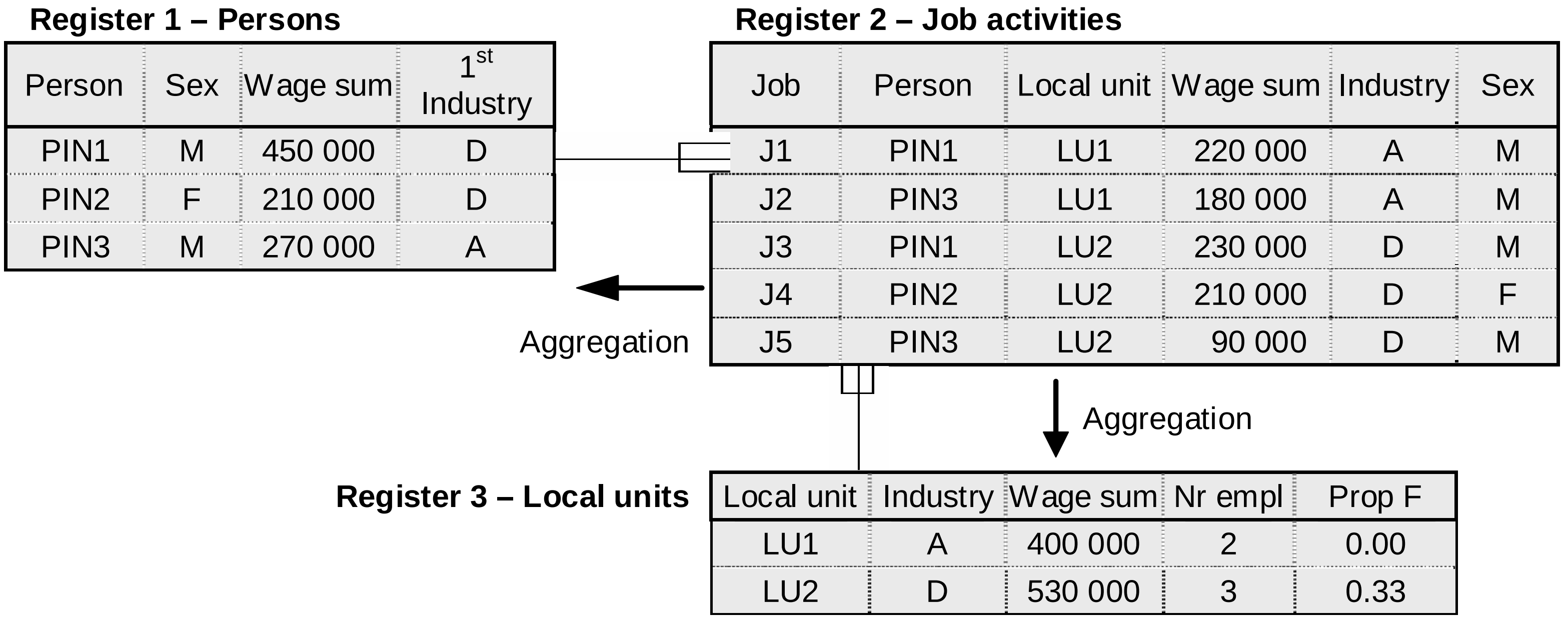
Job	Person	Local unit	Wage sum	Industry	Sex
J1	PIN1	LU1	220 000	A	M
J2	PIN3	LU1	180 000	A	M
J3	PIN1	LU2	230 000	D	M
J4	PIN2	LU2	210 000	D	F
J5	PIN3	LU2	90 000	D	M

Register 3 – Local units

Local unit	Industry	Wage sum
LU1	A	400 000
LU2	D	530 000

Chart 3.14D

Industry, number of employees and proportion of females as derived variables – by aggregation



Register 1 – Persons

Person	Sex	Wage sum	1st Industry
PIN1	M	450 000	D
PIN2	F	210 000	D
PIN3	M	270 000	A

Register 2 – Job activities

Job	Person	Local unit	Wage sum	Industry	Sex
J1	PIN1	LU1	220 000	A	M
J2	PIN3	LU1	180 000	A	M
J3	PIN1	LU2	230 000	D	M
J4	PIN2	LU2	210 000	D	F
J5	PIN3	LU2	90 000	D	M

Register 3 – Local units

Local unit	Industry	Wage sum	Nr empl	Prop F
LU1	A	400 000	2	0.00
LU2	D	530 000	3	0.33

Variables with different origins

- Local primary variable: coming directly from an administrative register
- Locally derived variable: derived from other variables in the same register
- Imported primary variable: coming from another statistical register with identical objects
- Imported derived variable: derived variable from another statistical register with identical objects

Variables with different functions

- identifying variable (primary key)
- reference variable (foreign key, link)
- communication/location variable
- time variable: reference time, reporting time, ...
- statistical variable: classification/spanning variable, summation/response variable, ...
- metadata/technical variable: source of value, computed/corrected value, weight, comment, ...

Variables used for matching/linking

- Linking identical objects in different registers or database tables, e.g. linking different variables about the same person in different registers, using the person number
- Linking objects in the same register, or in different registers, which are related to each other in a certain way, e.g. linking a person to his/her spouse (via person number), or his/her dwelling (via dwelling identifier)

Chart 3.17 A register's primary role in the system

Type of register	**Types of variables**	**Role and responsibility**
Base register	*Local primary variables:* Identifying variables Communication variables Reference variables Time references	Receive administrative data Create object sets Define objects Create some basic spanning variables Produce demographic statistics
Primary register	*Local primary variables:* Identifying variables Statistical variables	Receive administrative data Create the actual statistical variables
Integrated register	*Imported variables:* Identifying variables Statistical variables Locally derived variables, adjoined and aggregated variables	Create new information without data collection Compile information from different fields Compile information from different time periods

Sample surveys and registers

- How can sample surveys benefit from registers?
- How can registers benefit from sample surveys?
- Combining register-based surveys and sample surveys
- Comparing sample surveys and register-based surveys

How can sample surveys benefit from registers?

- When selecting the sample, an appropriate register is used as a *sampling frame*, and register variables are used to stratify the population: *stratification variables*
- Measurements can be made easier by eliminating the need for questions on data that already exists in the registers → *reduced response burden*
- During the estimation phase, register variables can be used as auxiliary information to increase precision and compensate for non-response → *improve quality (precision) and/or decrease costs (sample size)*

How can registers benefit from sample surveys?

- Quality checks and quality improvements
  - overcoverage and undercoverage
  - non-response, missing values
  - bias (e.g. because of administrative purposes of registers)
  - other imperfections in the register as a frame

Combining register-based surveys and sample surveys

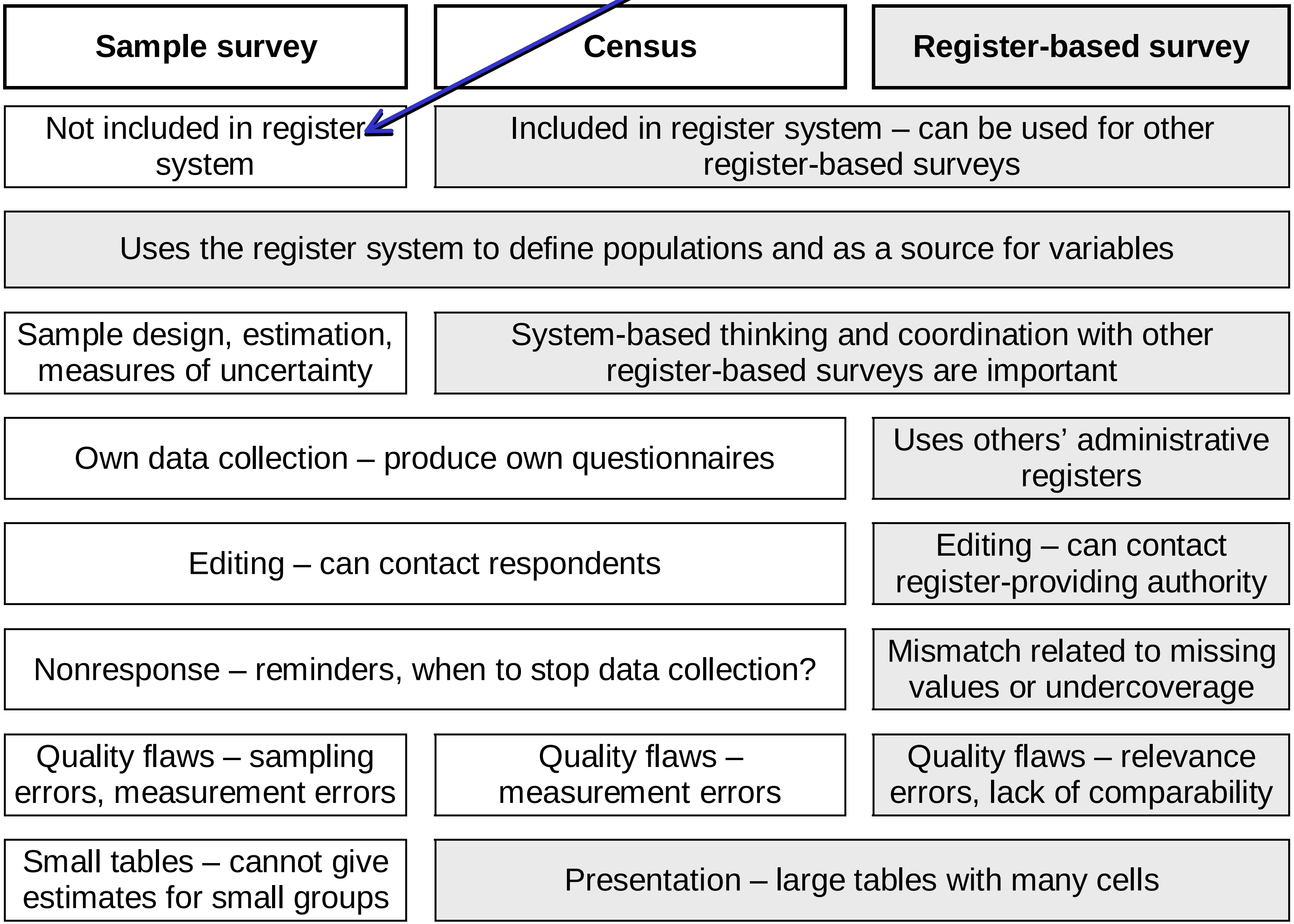
- Defining a precise target population
- Sample surveys can give indications on register quality
- Register maintenance surveys could be used to improve register quality
- An administrative register can be complemented with special data collection
- Sample surveys can be used when creating derived variables in registers
- Small area estimation
- Virtual censuses

Comparing register-based surveys and sample surveys

- When designing a sample survey, the designer is, in principle, free to define populations, parameters, observation objects, and observation variables (e.g. questions in a questionnaire) first, starting from user needs and priorities
- When designing a register-based survey, the designer has to start from available data in available registers and other available data sets
- Different conditions for the editing process, e.g. when data are missing or suspicious
- Different conditions for changing measurement processes and measurement instruments
- Different possibilities and problems in the presentation process, e.g. precision and confidentiality problems in small groups

Note: What is not included in a register system may be included in a register-based statistical system!

Chart 4.2 Similarities and differences between the different types of survey



Sample survey	Census	Register-based survey
Not included in register system	Included in register system – can be used for other register-based surveys	
Uses the register system to define populations and as a source for variables		
Sample design, estimation, measures of uncertainty	System-based thinking and coordination with other register-based surveys are important	
Own data collection – produce own questionnaires		Uses others' administrative registers
Editing – can contact respondents		Editing – can contact register-providing authority
Nonresponse – reminders, when to stop data collection?		Mismatch related to missing values or undercoverage
Quality flaws – sampling errors, measurement errors	Quality flaws – measurement errors	Quality flaws – relevance errors, lack of comparability
Small tables – cannot give estimates for small groups	Presentation – large tables with many cells	

The objects and population(s) of a register

- Object type and object instance
- Object set, set of object instances:
  - population: ”all” objects of a type at a time
  - domain, subpopulation: a subset of a population, often defined by a crossclassification of variables, classification variables
- Population of interest: the ”ideal” population, sometimes rather vaguely defined
- Target population: the population decided in the design process to be aimed at in practice, precisely defined, in an operational way
- Register population: the set of objects actually obtained/registered

Example: Population register of Sweden

- Population of interest: People permanently living in Sweden on December 31

Relevance error

- Target population: People registered by the appropriate agency as living in Sweden on December 31

Coverage error

- Register population: People actually registered as living in Sweden on December 31, according to information obtained up to January 31 next year

Creating a statistical register

- Determine the objectives: Which statistical needs are to be fulfilled by the register? Examples:
  - cross-sectional status data
  - events and changes
  - time series data (macro-level)
  - longitudinal studies
- Define the desirable contents of the register in terms of objects and variables
- The inventory phase: Which sources are available, administrative and statistical?
- Editing and integration of the sources:
  - match, check, edit, redefine, and reconcile objects, and synchronise them as regards times of reference
  - combine, check, edit, redefine, and synchronise variables

Defining populations of registers

- General methodology
  - Define the target population
  - Select the intended object set from the base register, giving the register population
  - Match against registers containing interesting registers
  - When receiving hits: import the variable values to the register which is created
  - When receiving mismatches: show missing values (item nonresponse
- Standardised populations created for general usage:
  - end of year version: suitable for annual stock statistics, such as the population on December 31
  - calendar year version: suitable for annual flow statistics, such as the population's income during a specific year
  - monthly/quarterly version: suitable for monthly/quarterly statistics

Important requirements on base registers

- A base register should contain time references, i.e. all events that affect the register’s objects should be dated
  - dates of events (birth/deaths, moves, category changes...)
  - dates of registration/update
- A base register should have good coverage (neither overcoverage, nor undercoverage)
- Linkage variables should be of high quality
- Classification/spanning variables should be of high quality, otherwise there will be coverage errors in subpopulations (domains of interest)

Register matching

- When unique, officially authorized identities exist (like for persons and organisations in Sweden), and are used in registers involved, register matching (also called record linkage) is relatively easy
- Nevertheless, errors may occur, because of
  - errors in identities (not so common)
  - errors in references reflecting relations to other objects
  - coverage errors in the registers involved
- When unique, officially authorized identities do not exist (like in many countries), or are not used, more complex and error-prone matching has to take place
- Statistical matching is something else, where the purpose is to find *similar* objects for analytical purposes (or imputation)

Chart 5.7 Frame populations and annual registers

A. Frame population
formed in Nov year 1
for years 1 and year 2

Enterprise id	Industry
Idnr 1	DE
Idnr 2	DB
Idnr 3	**DA**
Idnr 4	DC
-	-
-	-

B. Calendar year register
formed in autumn year 2
regarding year 1

Enterprise id	Industry
-	-
Idnr 2	DB
Idnr 3	**DB**
Idnr 4	DC
Idnr 5	DG
-	-

C. Calendar year register
formed in autumn year 3
regarding year 2

Enterprise id	Industry
-	-
-	-
Idnr 3	**DB**
Idnr 4	DC
Idnr 5	DG
Idnr 6	DC

Chart 5.8 Population definitions in different kinds of surveys

	Advantages	**Disadvantages**
Survey statistics, own data collection	Can be up-to-date	Significant problems with over- and undercoverage and errors in spanning variables if changes are reported late
Register-based statistics	Good coverage, more correct spanning variables	In certain cases, a long delay between the event to the statistics becoming available

A register population, created in the correct manner, has always better quality than the corresponding frame population, as it is based on more and better information.

Creating register variables and their values

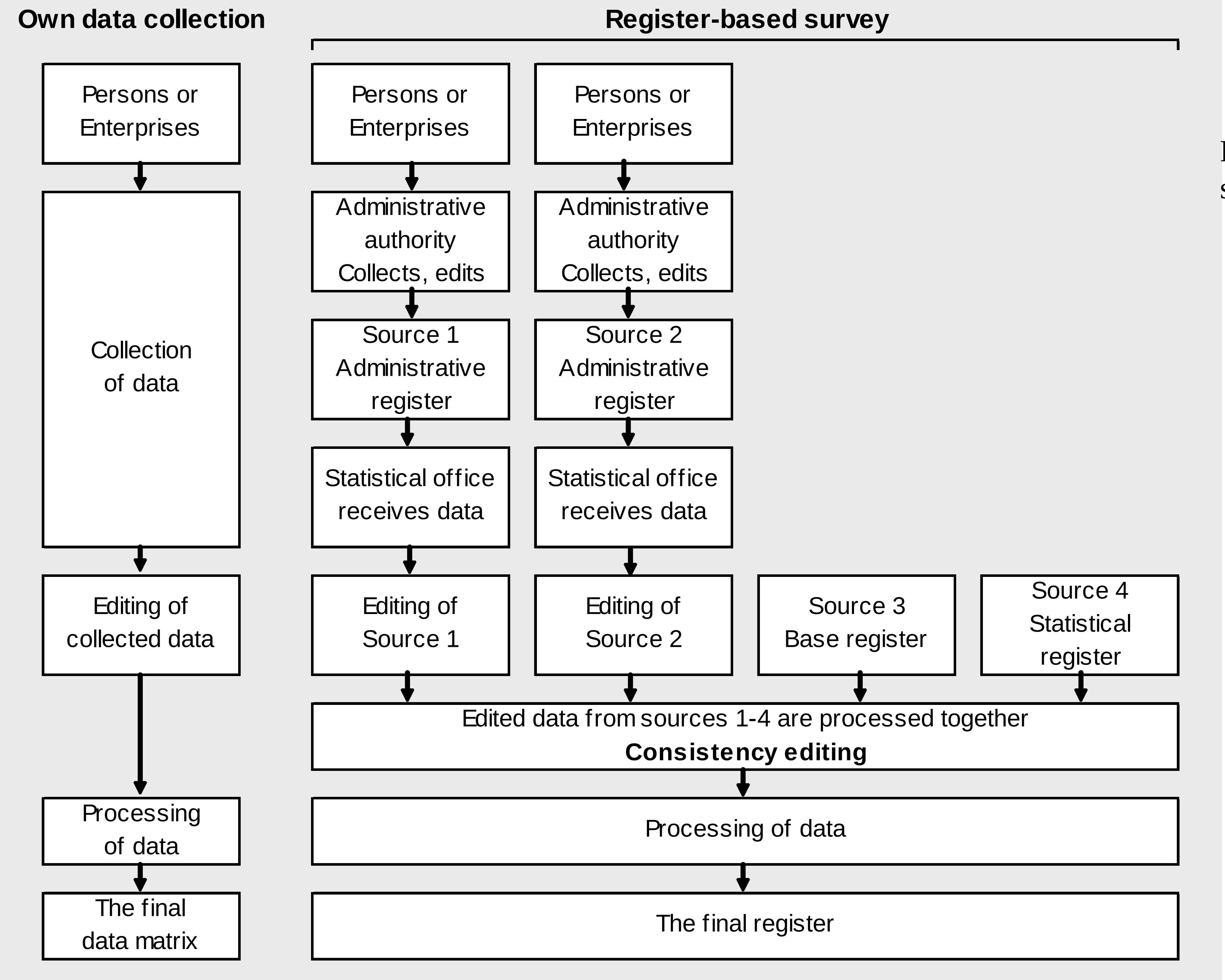
- When creating a statistical register, both objects and variables may come from different sources and need to be carefully checked and reconciled before they are accepted
- The checking and editing that has taken place in the source register, will have been done for other purposes, e.g. administrative purposes
- Derivation of variables (discussed before) and imputation of (missing or suspicious) values of variables are related but different phenomena: a derived variable is created for all objects in a register, whereas an imputed variable value is only formed for the objects in a register where values are missing (or deemed erroneous)

Editing processes in a register system

- Create a data matrix and combine all records that belong to the same object
- Check the register population
- Check that the data regarding a specific identity from different sources really refer to the same object
- Check that the data delivery from administrative sources are complete, both regarding objects and variables; differentiate between missing data and true zero values
- Check variable values for ”obvious” errors
- Make sure that the editing process is documented

For more explanations and illustrative examples: see Wallgren & Wallgren, Chapter 6.

Chart 6.5 Editing in surveys with their own data collection and register-based surveys


Own data collection
Register-based survey
Persons or Enterprises
Collection of data
Editing of collected data
Processing of data
The final data matrix
Persons or Enterprises
Administrative authority Collects, edits
Source 1 Administrative register
Statistical office receives data
Editing of Source 1
Persons or Enterprises
Administrative authority Collects, edits
Source 2 Administrative register
Statistical office receives data
Editing of Source 2
Source 3 Base register
Source 4 Statistical register
Edited data from sources 1-4 are processed together
Consistency editing
Processing of data
The final register

Page 101
sid 99

Important aspects of data editing

- In many cases, a small number of huge errors destroy data – as a rule it is easy to find and correct these errors
- Use selective editing to find the most important errors first
- Capture knowledge and experiences from domain experts and use this information in documentation and software – neural networks an interesting possibility
- Automatic editing and imputation – pros and cons

Part 1: Extra material

Bo Sundgren

2010

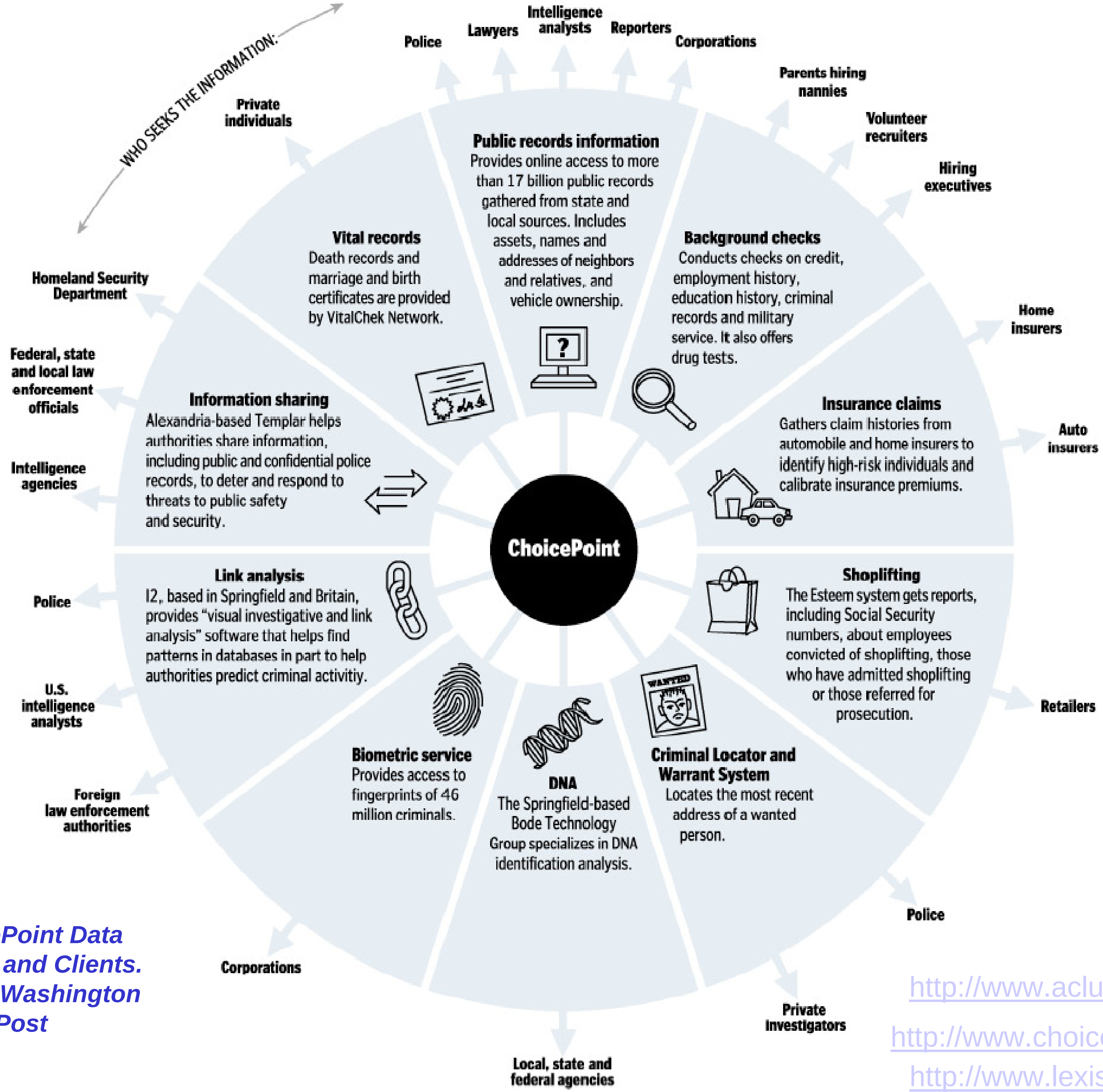
Registers...

- *… US version: ”war against terrorism”*

http://www.aclu.org/pizza/

http://www.choicepoint.com/

http://www.lexisnexis.com/



WHO SEEKS THE INFORMATION:

Police

Lawyers

Intelligence analysts

Reporters

Corporations

Parents hiring nannies

Volunteer recruiters

Hiring executives

Home insurers

Auto insurers

Retailers

Police

Private investigators

Local, state and federal agencies

Corporations

Foreign law enforcement authorities

U.S. intelligence analysts

Police

Intelligence agencies

Federal, state and local law enforcement officials

Homeland Security Department

Private individuals

ChoicePoint

Public records information
Provides online access to more than 17 billion public records gathered from state and local sources. Includes assets, names and addresses of neighbors and relatives, and vehicle ownership.

Background checks
Conducts checks on credit, employment history, education history, criminal records and military service. It also offers drug tests.

Insurance claims
Gathers claim histories from automobile and home insurers to identify high-risk individuals and calibrate insurance premiums.

Shoplifting
The Esteem system gets reports, including Social Security numbers, about employees convicted of shoplifting, those who have admitted shoplifting or those referred for prosecution.

Criminal Locator and Warrant System
Locates the most recent address of a wanted person.

DNA
The Springfield-based Bode Technology Group specializes in DNA identification analysis.

Biometric service
Provides access to fingerprints of 46 million criminals.

Link analysis
I2, based in Springfield and Britain, provides "visual investigative and link analysis" software that helps find patterns in databases in part to help authorities predict criminal activitiy.

Information sharing
Alexandria-based Templar helps authorities share information, including public and confidential police records, to deter and respond to threats to public safety and security.

Vital records
Death records and marriage and birth certificates are provided by VitalChek Network.

ChoicePoint Data Sources and Clients. Source: Washington Post

http://www.aclu.org/pizza/

http://www.choicepoint.com/

http://www.lexisnexis.com/