Statistical databases in theory and practice
Part III: Statistical information systems (extra material)

Bo Sundgren
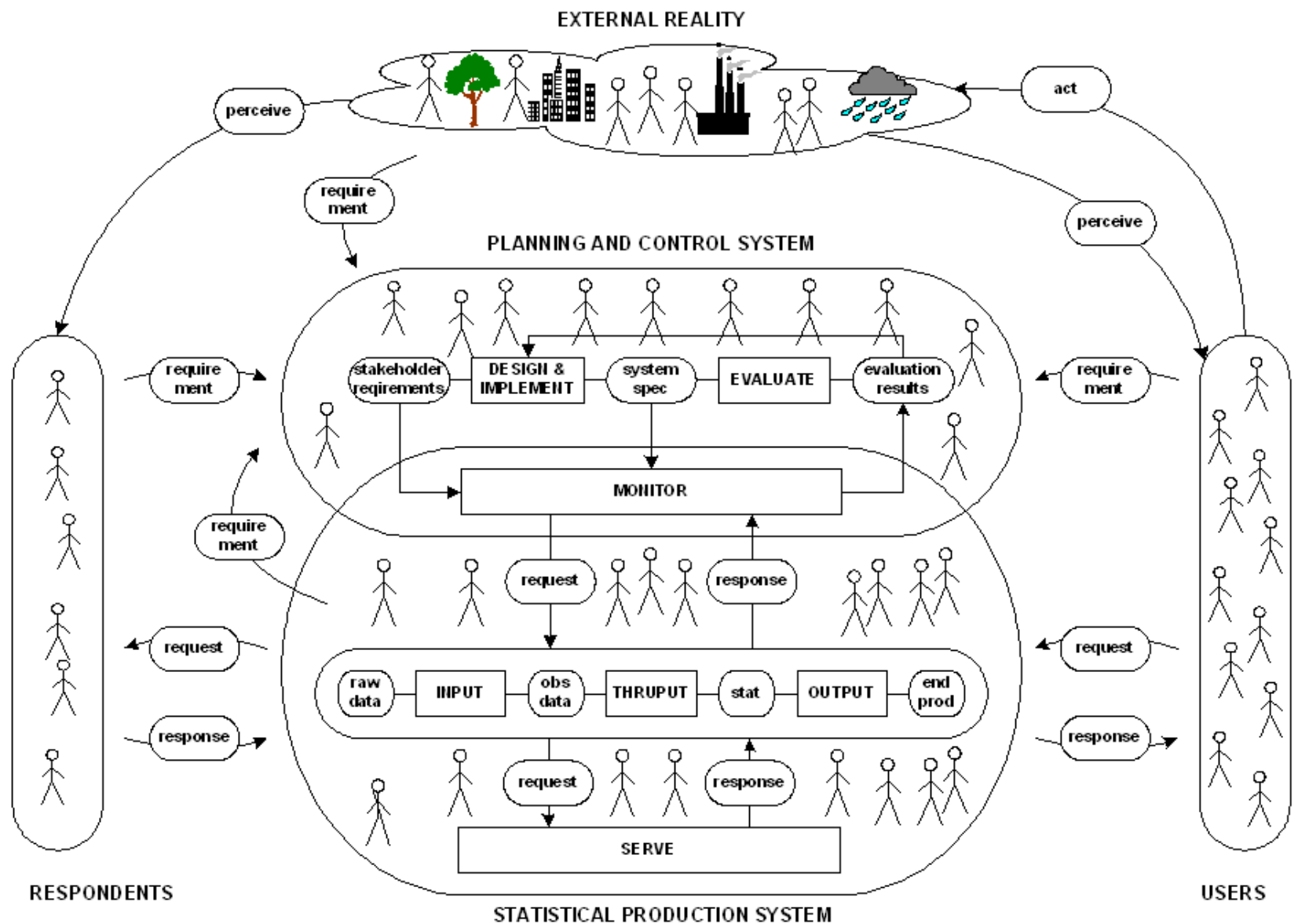
2010

EXTERNAL REALITY

PLANNING AND CONTROL SYSTEM

perceive

require ment

act

perceive

stakeholder reqirements

DESIGN & IMPLEMENT

system spec

EVALUATE

evaluation results

require ment

require ment

MONITOR

require ment

request

response

request

RESPONDENTS

raw data

INPUT

obs data

THRUPUT

stat

OUTPUT

end prod

request

request

response

response

request

response

SERVE

USERS

STATISTICAL PRODUCTION SYSTEM

Figure 2. *A statistical system in its environment. The statistical production process and its customers (users of statistics and respondents and data providers).*
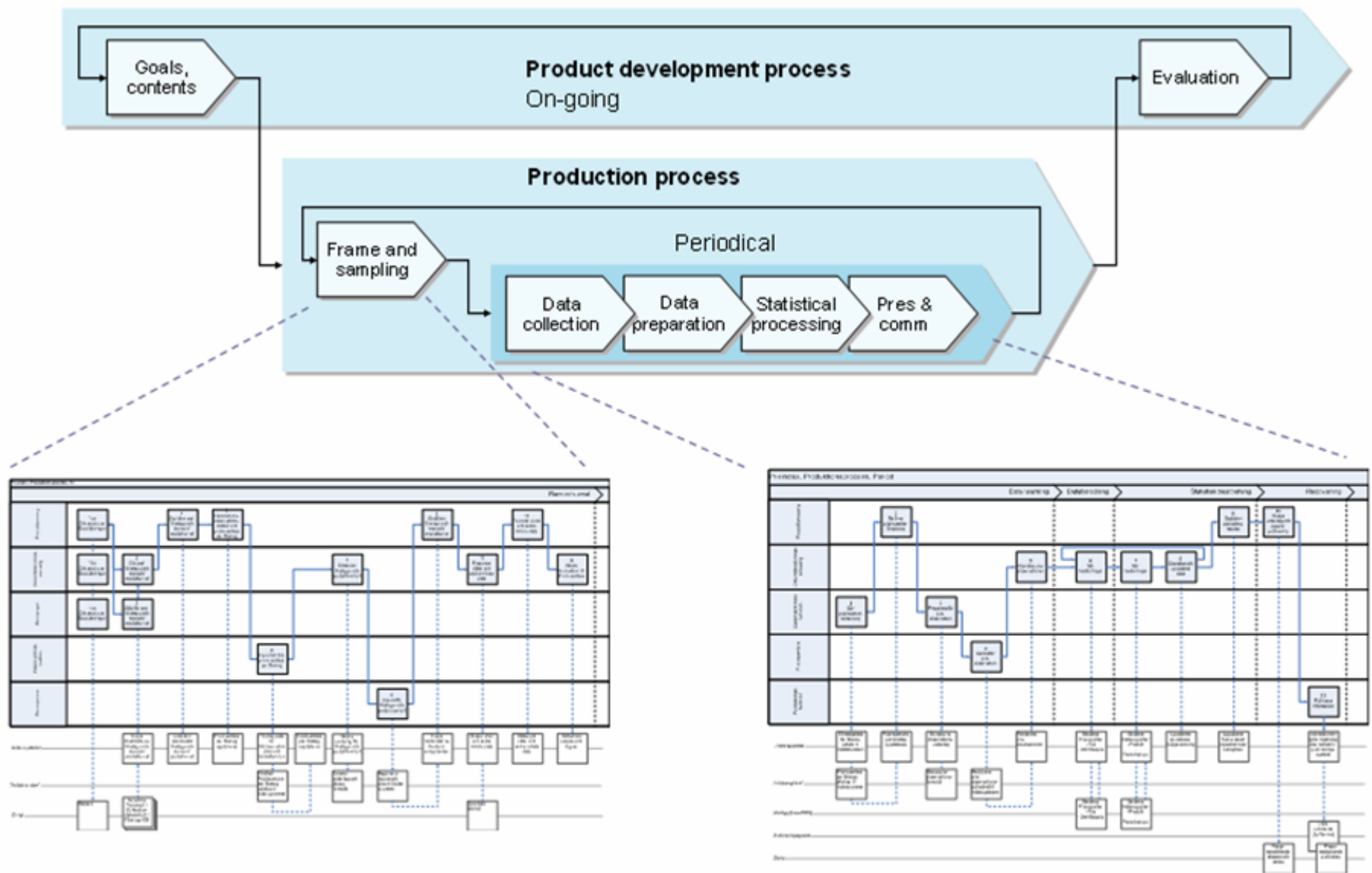
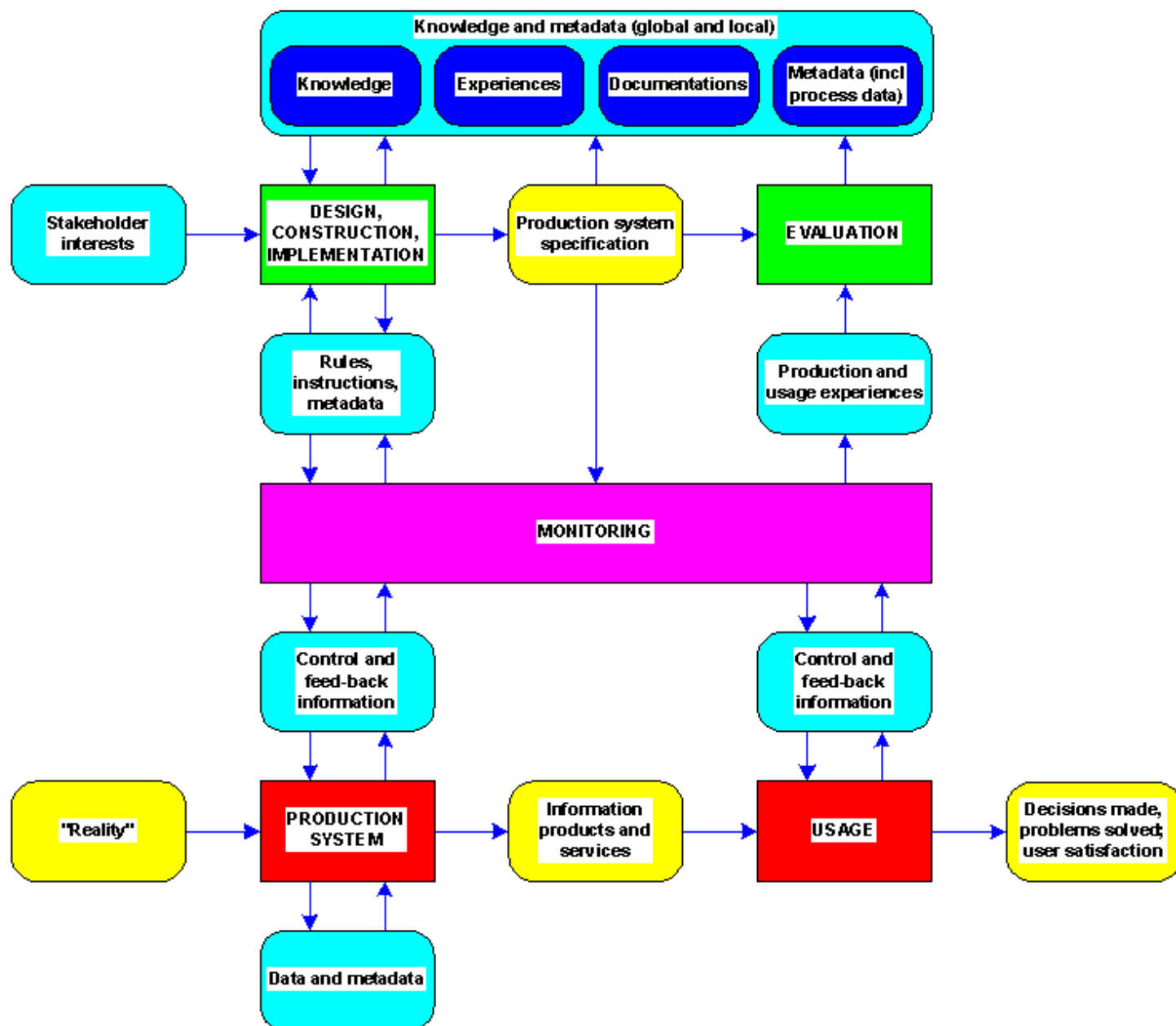**Figure 1.** Statistics production: product development process and production process.

Figure 3. Control and execution of a statistical production system.

The control process flow

Planning → Production system specification + Production system
- Design
    - Purposes, contents, and functionalities (specified in communication with customers)
    - Architecture and infrastructure (adaptation to standard)
        - Organisational, conceptual, methodological, and technical aspects
    - Detailed design of subprocesses and components (adaptation to standards)
- Construction
- Testing
- Implementation

Monitoring → (Possibly) Request for modifications of system and/or instructions
- Systematic collection and use of process data (feed-back) according to standard procedures

Evaluation → (Possibly) Request for replanning of the production system
- Evaluation of feed-back from the production system execution and monitoring
- Evaluation of feed-back from customers (users of statistics and providers of input data)

The execution process flow

Input processes: Data collection and data preparation
Obtain input forms from the design process
Establish an updated frame and draw a sample, if applicable
Obtain a list of respondents and other data sources
Obtain completed input forms from respondents and other data sources
Add coded and derived data to the input forms
Check the completed input forms for missing data and suspected errors
Recontact respondents and other data sources, if necessary
Impute new or modified data, if motivated
Establish and organise final observation data according to standards

Throughput processes: Performing computations and producing statistical estimates
Combine and transform data before aggregations
Compute statistical estimates by means of aggregation processes
Combine and transform data after aggregations
Perform analyses
Establish and organise final statistics and analytical results according to standards

Output processes
Prepare statistical end-products tailored to different needs
Protect statistical confidentiality according to standard procedures
Publish statistical
Make statistical end-products available to customers
Communicate with customers requesting and using statistical end-products

Infrastructure processes supporting the statistical system as a whole
Management of corporate data, metadata, and process data resources
Management of statistical registers (frames)
Management of standard classifications
Management of an archive of questionnaires and questions
Management of libraries of methods and tools

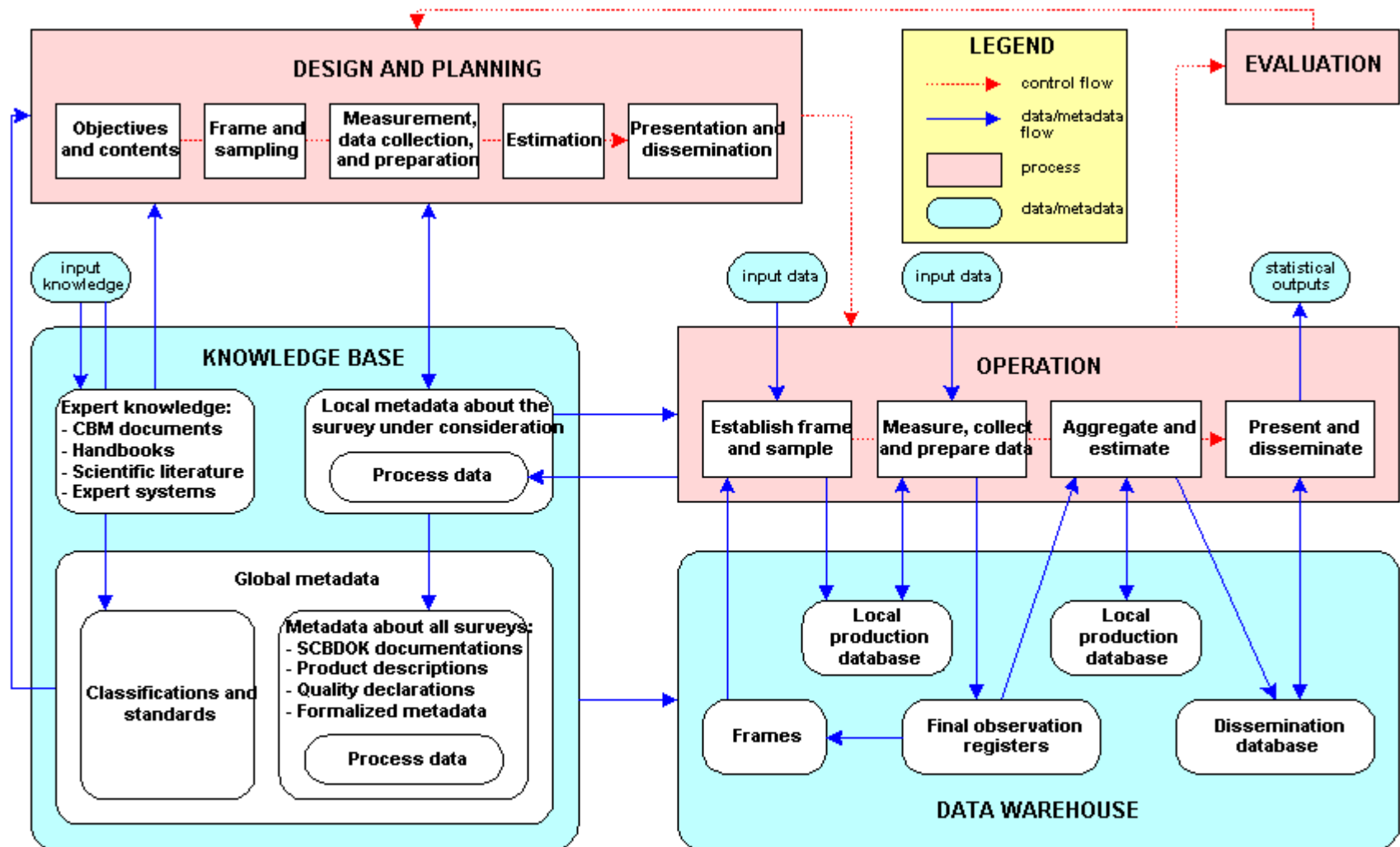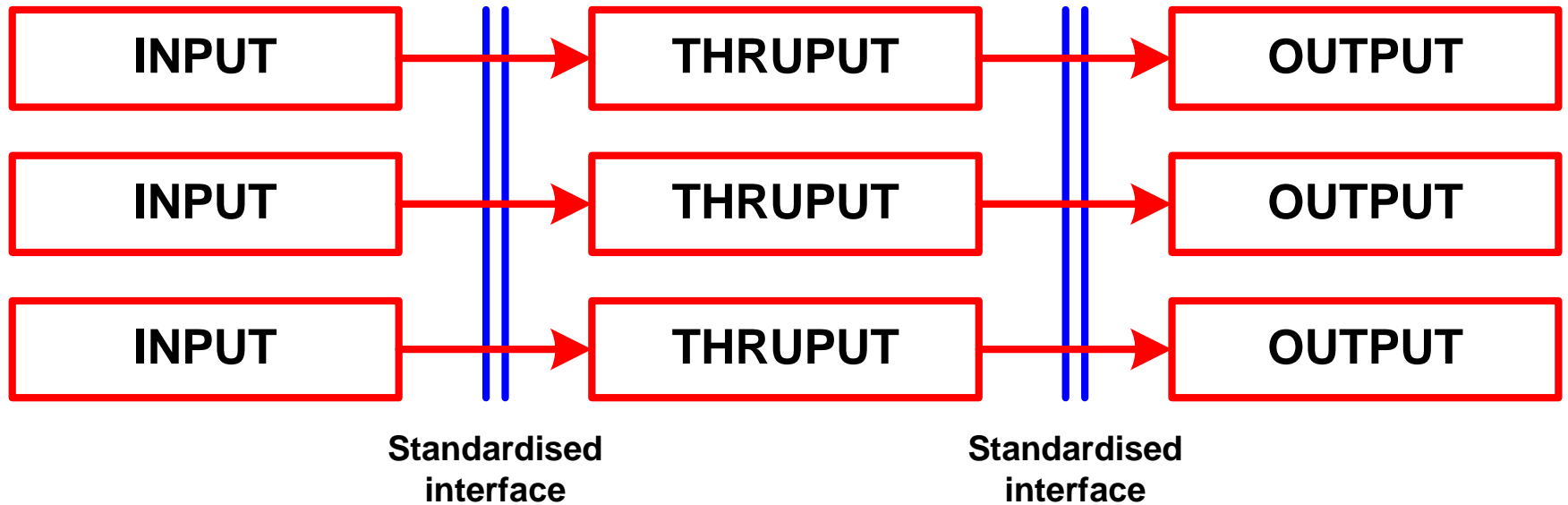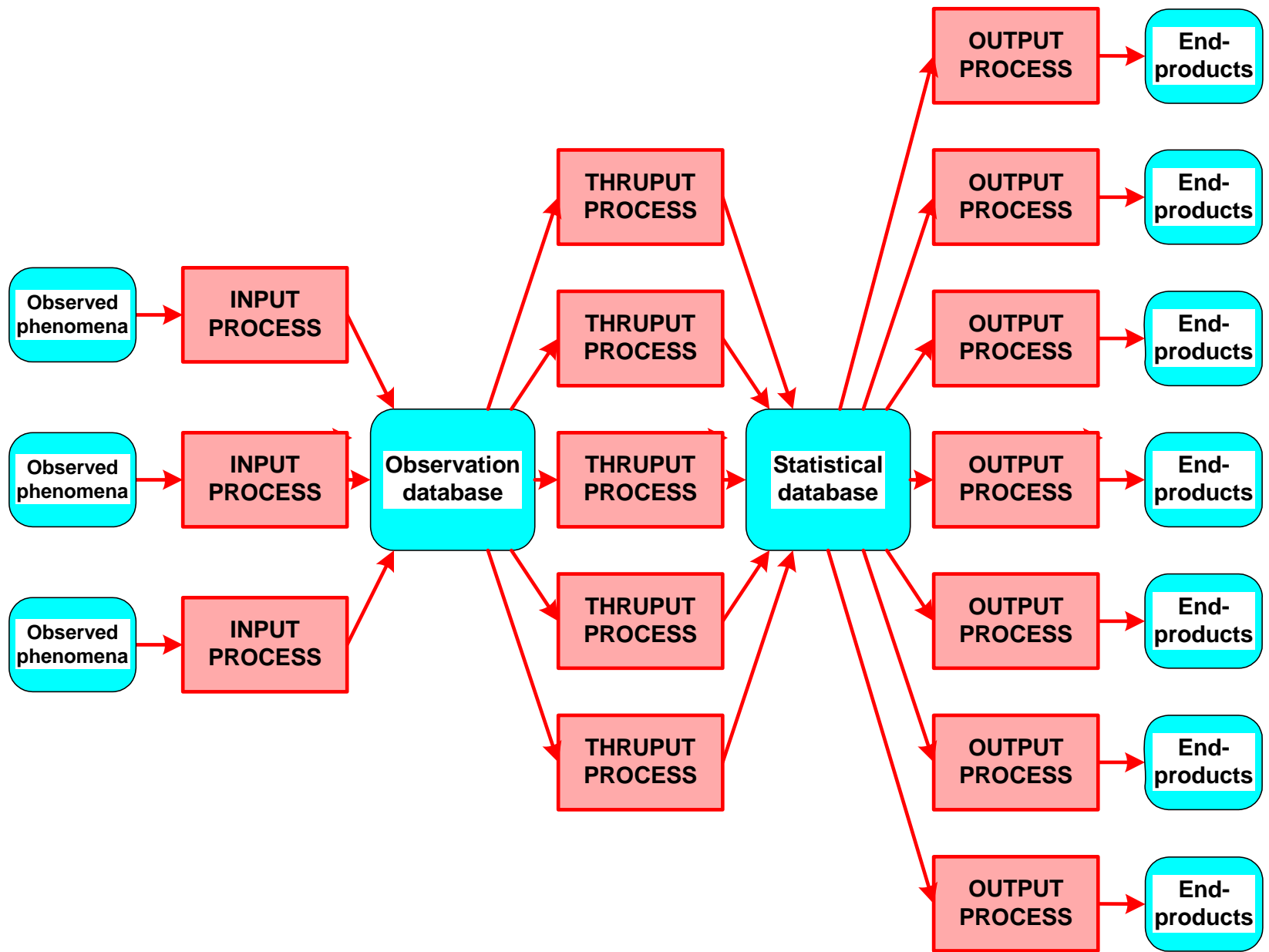**Figure 5.** *Statistics production with integrated data and metadata management.*

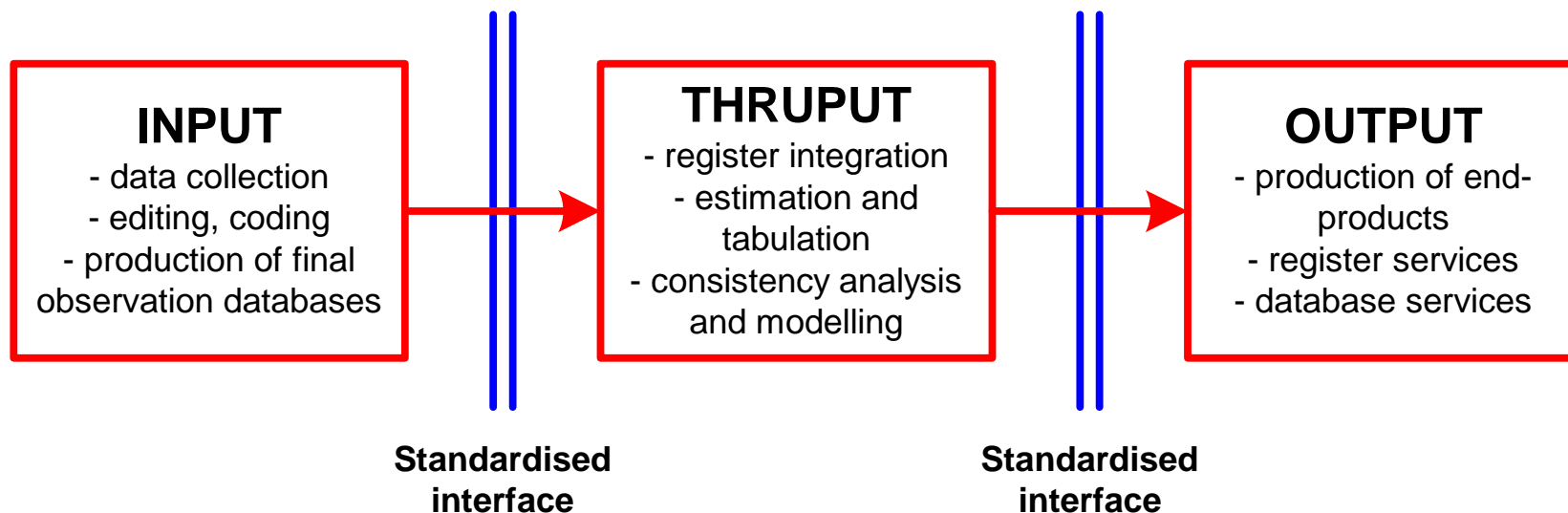# Major components of a statistical (production) system

- data about an external reality and data about other data (metadata)

- processes transforming data and metadata

- resources used for supporting (enabling) and controlling the processes

INPUT	Standardised interface	THRUPUT	Standardised interface	OUTPUT
INPUT		THRUPUT		OUTPUT
INPUT		THRUPUT		OUTPUT

The stovepipe architecture for statistical production systems.

The clearinghouse architecture.

A schematic picture of statistics production based on standardised processes and standardised interfaces.

| SERVICE 1 application | SERVICE 2 application | SERVICE 3 application | ... | SERVICE N application |

SERVICE PLATFORM AND INFRASTRUCTURE
functional components and generalised processes

INFORMATION PLATFORM AND INFRASTRUCTURE
data and metadata components

TECHNOLOGY PLATFORM AND INFRASTRUCTURE
hardware and software

General framework for how a wide range of service applications could be built on top of a well designed hierarchy of standardised platforms. A standardised technology platform supports a standardised information platform, which again support a standardised service platform with standardised functional components and generalised processes.
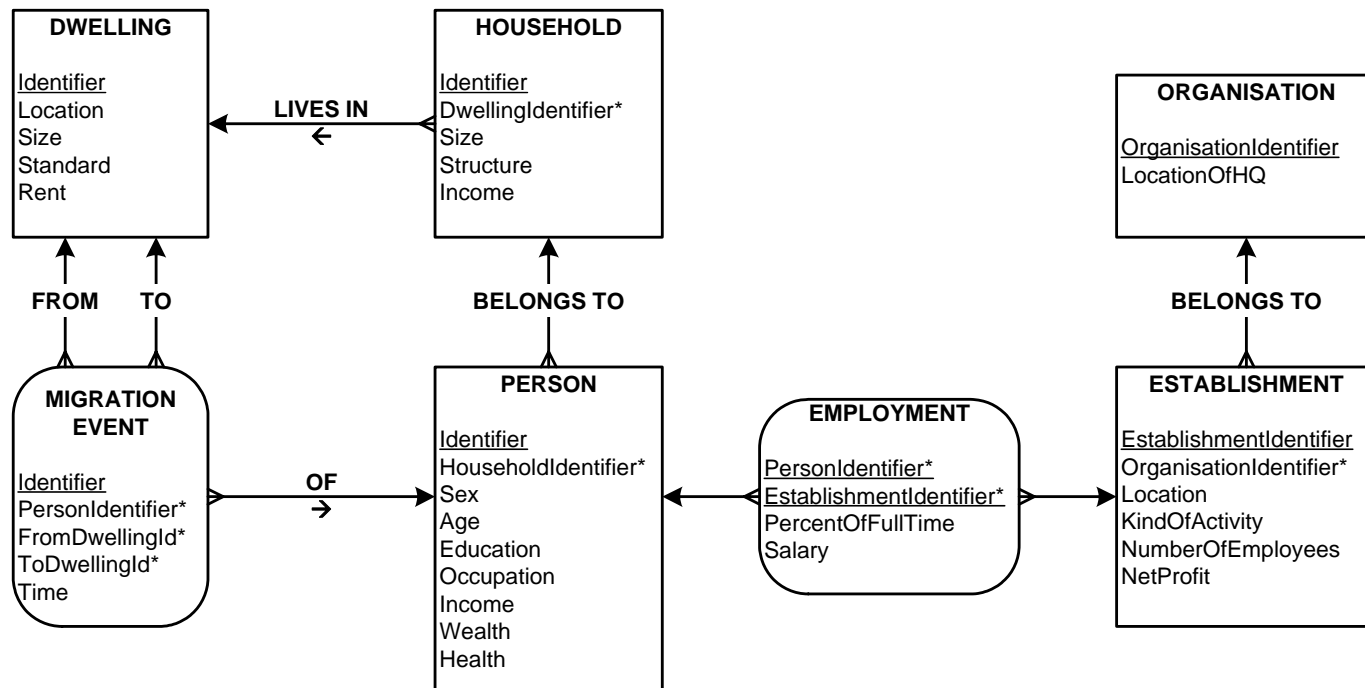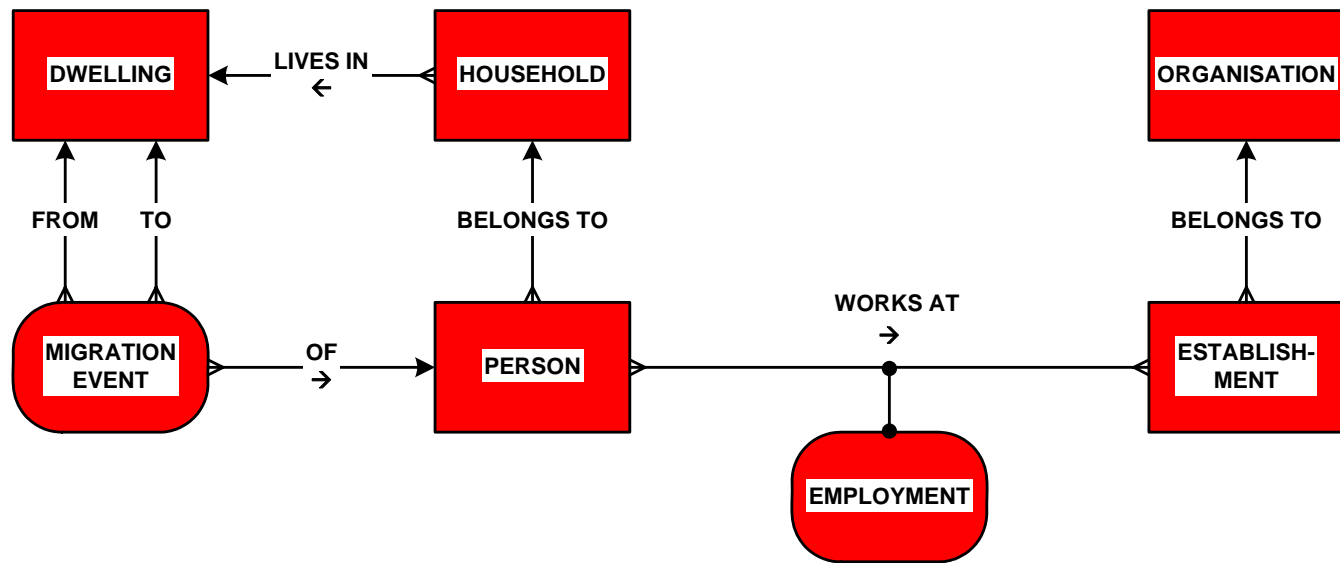
SOA = Service Oriented Architecture.
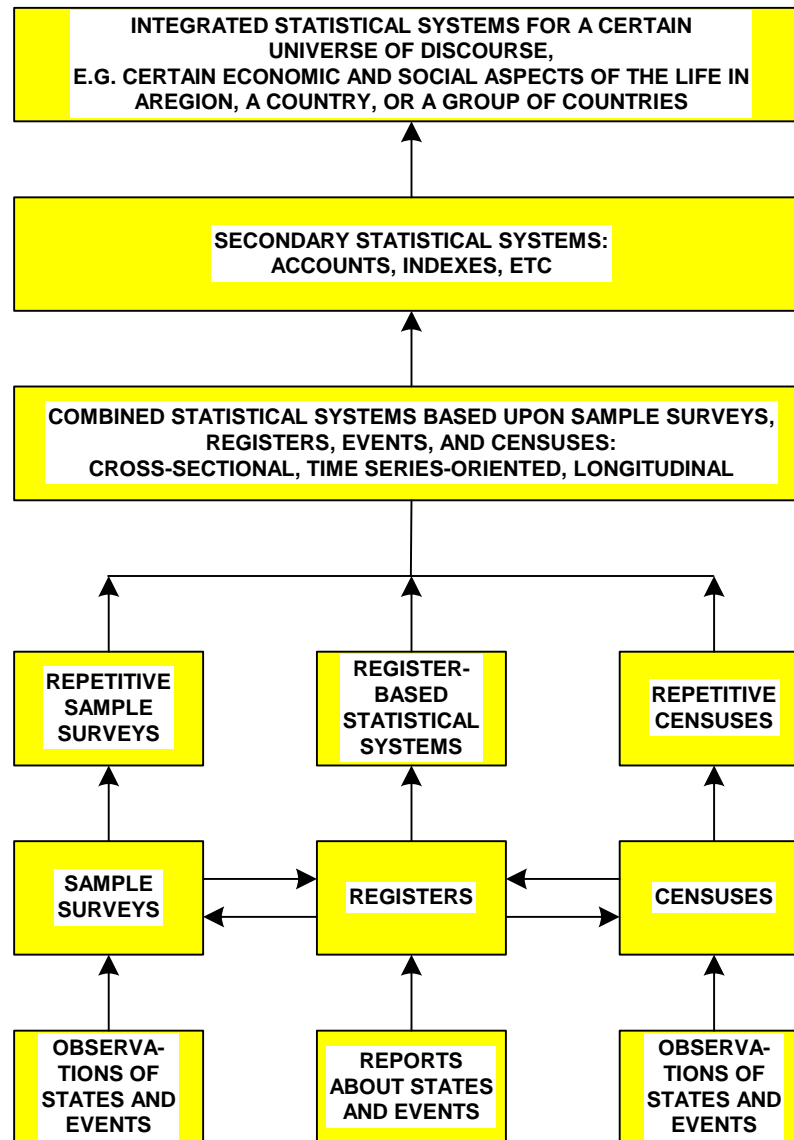
Different types of statistical production systems.

The data capital of a statistical agency

- The history of statistical data warehouses
  - Svein Nordbotten (1960 etc)
- The database concept
  - conceptual models and data models
  - data independence and data/metadata-driven systems
  - database management systems (DBMS)
  - standardisation of data: canonical forms and normalisation
  - statistical databases and statistical database management systems
- Statistical data warehouses
  - enterprise systems and data warehouses in business environments
  - the roles of data warehouses in statistical systems

Conceptual data model:

DWELLING

LIVES IN ←

HOUSEHOLD

ORGANISATION

FROM   TO

BELONGS TO

BELONGS TO

MIGRATION EVENT

OF →

PERSON

WORKS AT →

ESTABLISH-MENT

EMPLOYMENT

Relational data model:

DWELLING

Identifier
Location
Size
Standard
Rent

LIVES IN ←

HOUSEHOLD

Identifier
DwellingIdentifier*
Size
Structure
Income

ORGANISATION

OrganisationIdentifier
LocationOfHQ

FROM   TO

BELONGS TO

BELONGS TO

MIGRATION EVENT

Identifier
PersonIdentifier*
FromDwellingId*
ToDwellingId*
Time

OF →

PERSON

Identifier
HouseholdIdentifier*
Sex
Age
Education
Occupation
Income
Wealth
Health

EMPLOYMENT

PersonIdentifier*
EstablishmentIdentifier*
PercentOfFullTime
Salary

ESTABLISHMENT

EstablishmentIdentifier
OrganisationIdentifier*
Location
KindOfActivity
NumberOfEmployees
NetProfit

Conceptual data model and relational data model in normalised form.
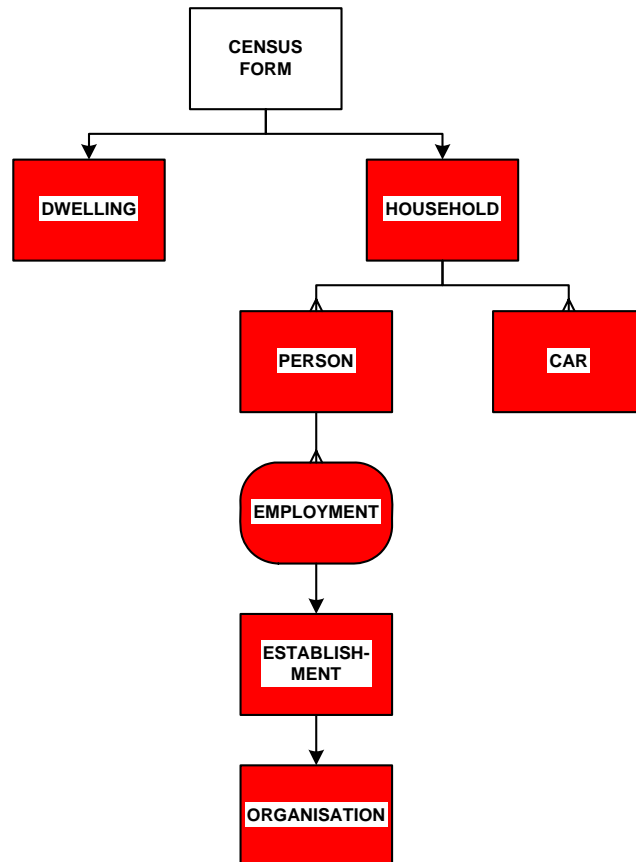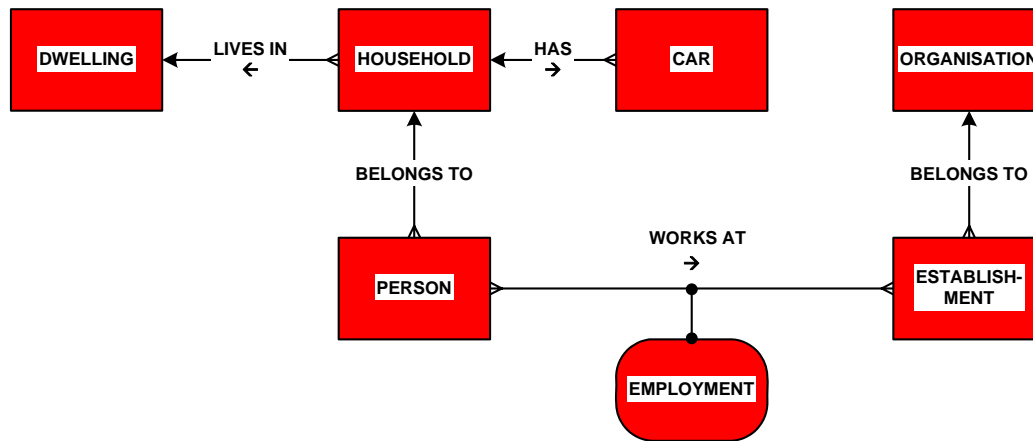
Mapping between conceptual model and input form.

Production process for statistics product	① Transfer of data and metadata	Sweden's statistical databases

Rules, guidelines, good advice

Templates, tools, routines → Production of electronic documents (Word) A4 format ②

③ Conversion to PDF format and production and dissemination of pre-subscribed copies ④

③ Conversion to HTML format

④ Printed SM or press release

④ Electronic archive (PDF, DOC, XLS, PX, etc.)

Web pages

⑤ Print On Demand → Printed SM or press release

⑤

⑤ User-defined print-out

Workflow for database-driven electronic publishing of statistics.
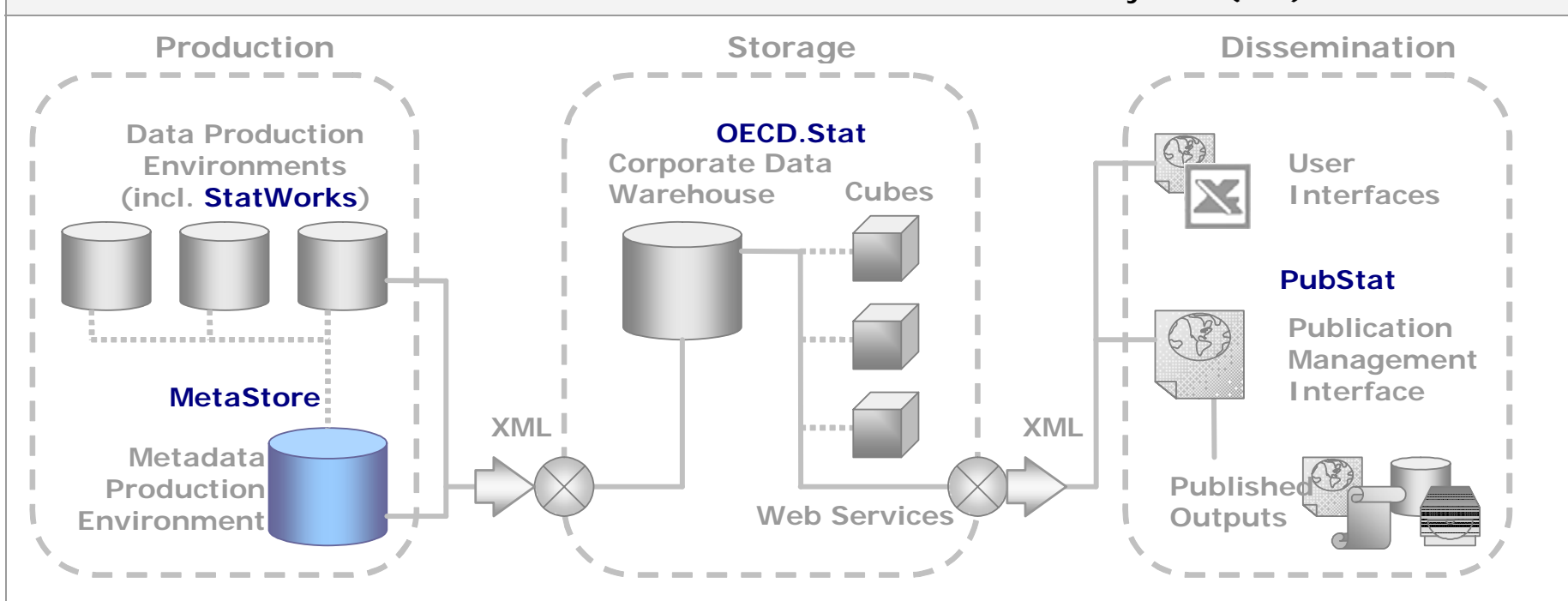
Database-driven statistics production and publishing at the OECD.