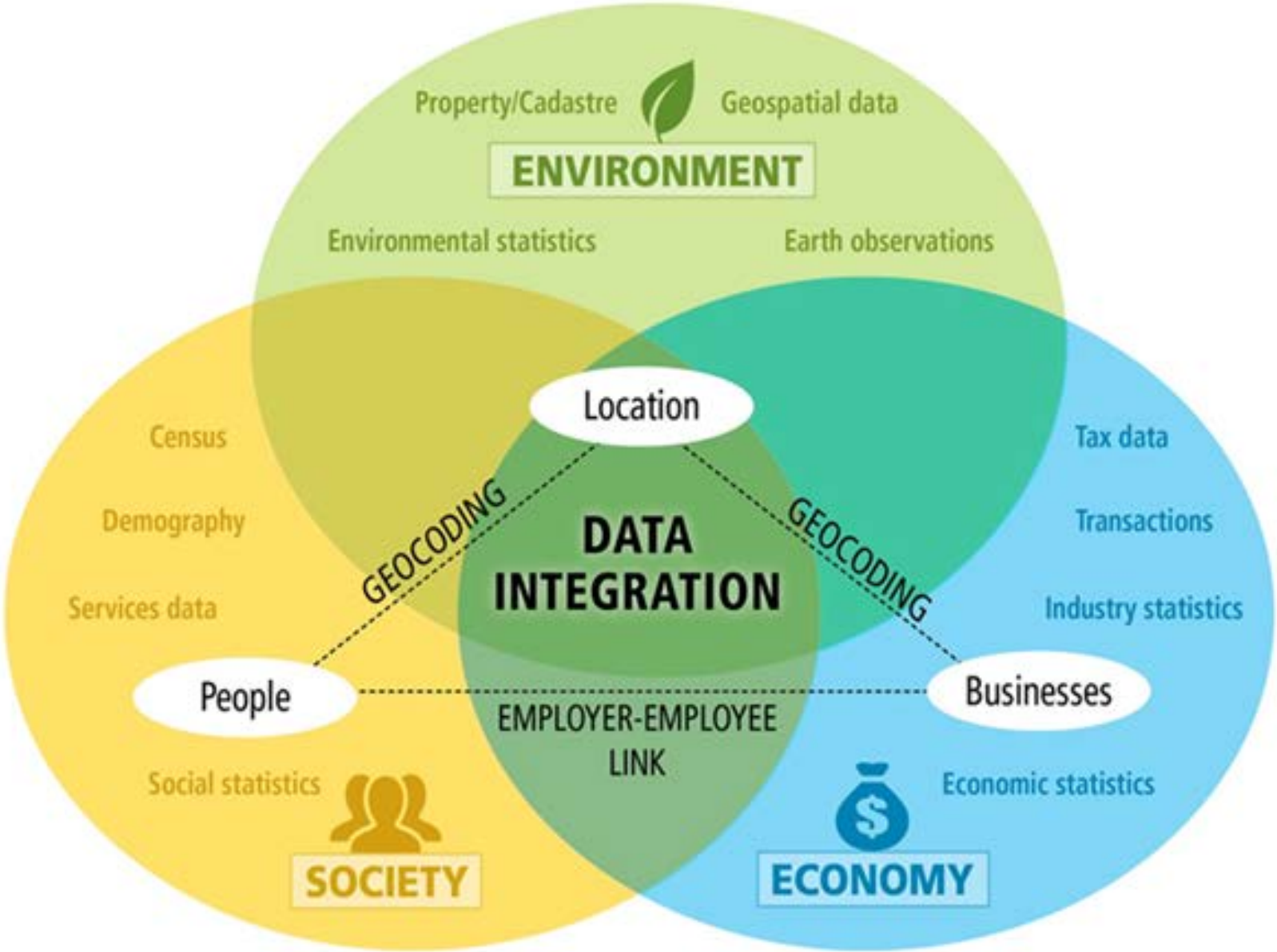
GEOSTAT 2

A Point-Based Foundation for Statistics in the EU

Marie Haldorson, Statistics Sweden





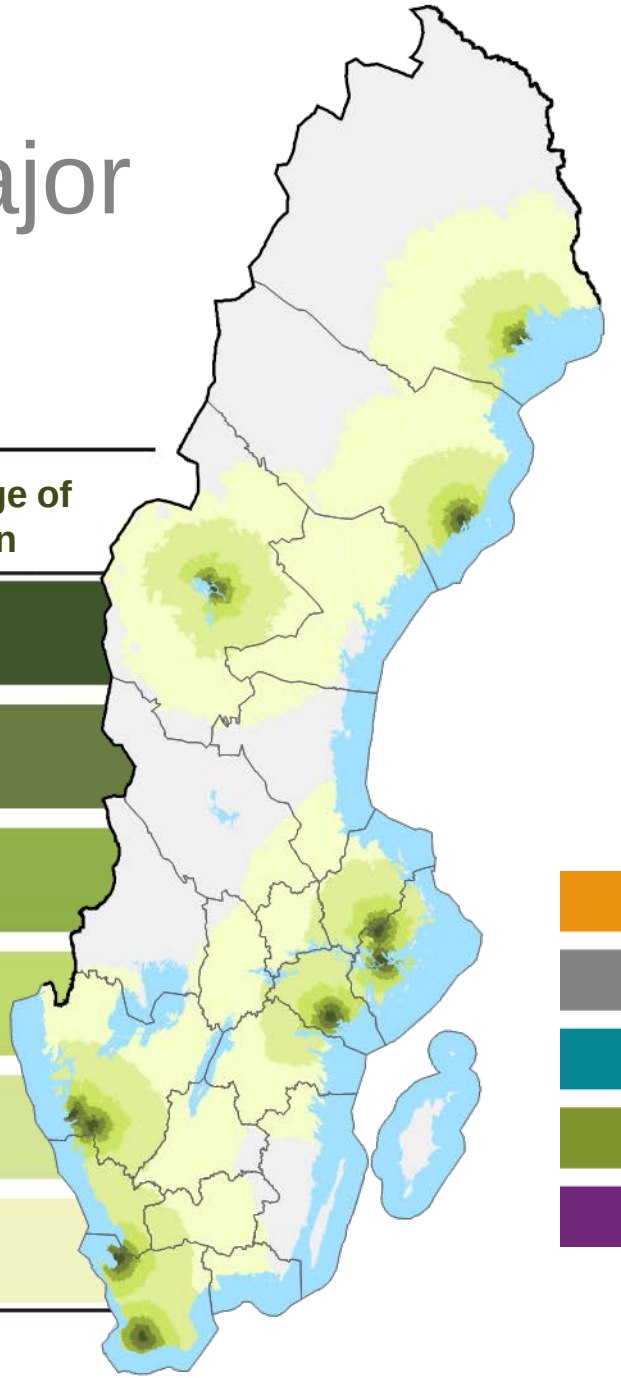


Source: Statistics Australia and Statistics New Zealand

Policy making, impact assessment:

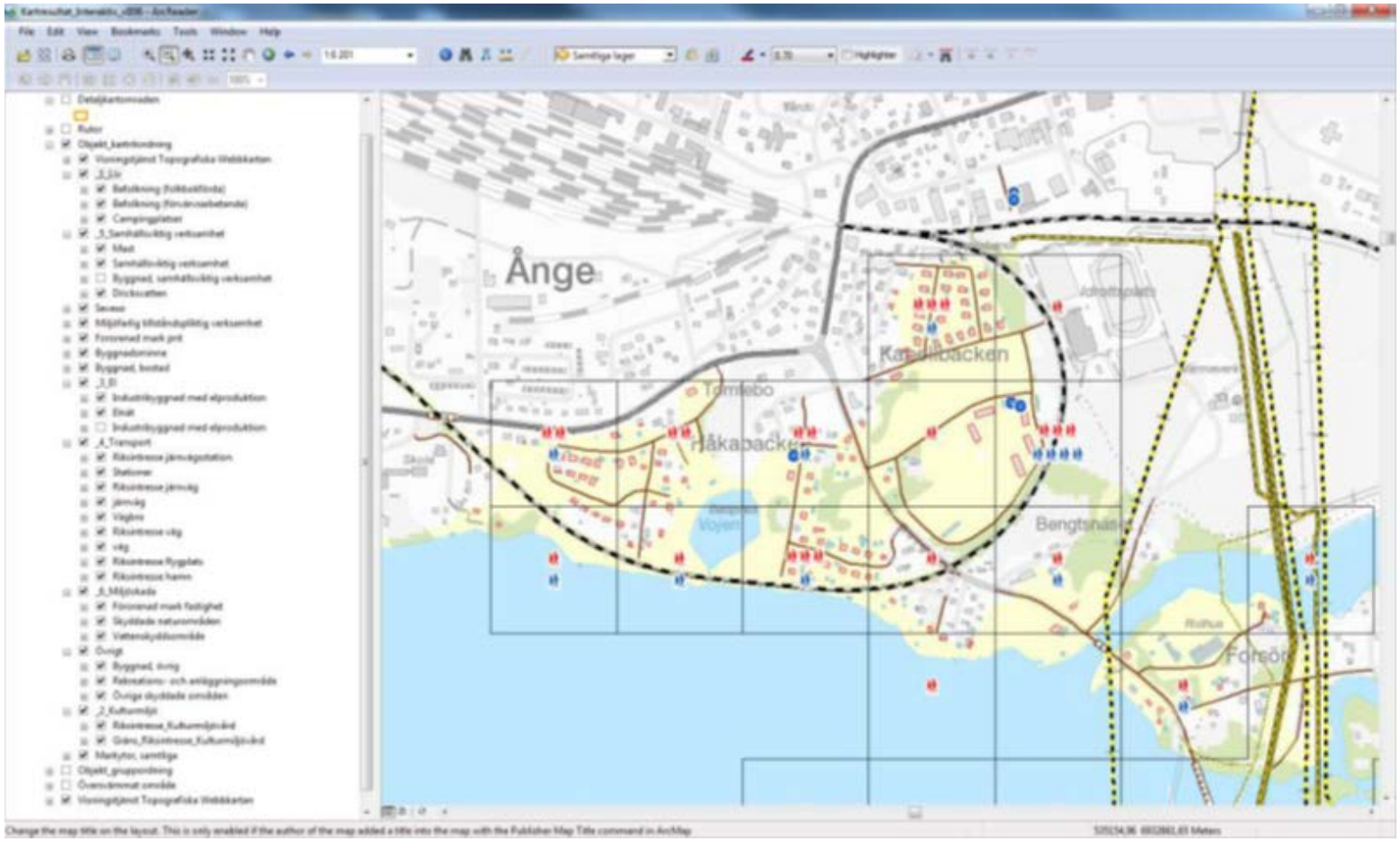
Catchment area for major airports

Catchment area	Number of inhabitants	Percentage of population
10 km	751 000	8 %
20 km	2 300 000	24 %
30 km	3 364 000	36 %
50 km	4 618 000	49 %
100 km	6 181 000	66 %
200 km	8 443 000	90 %



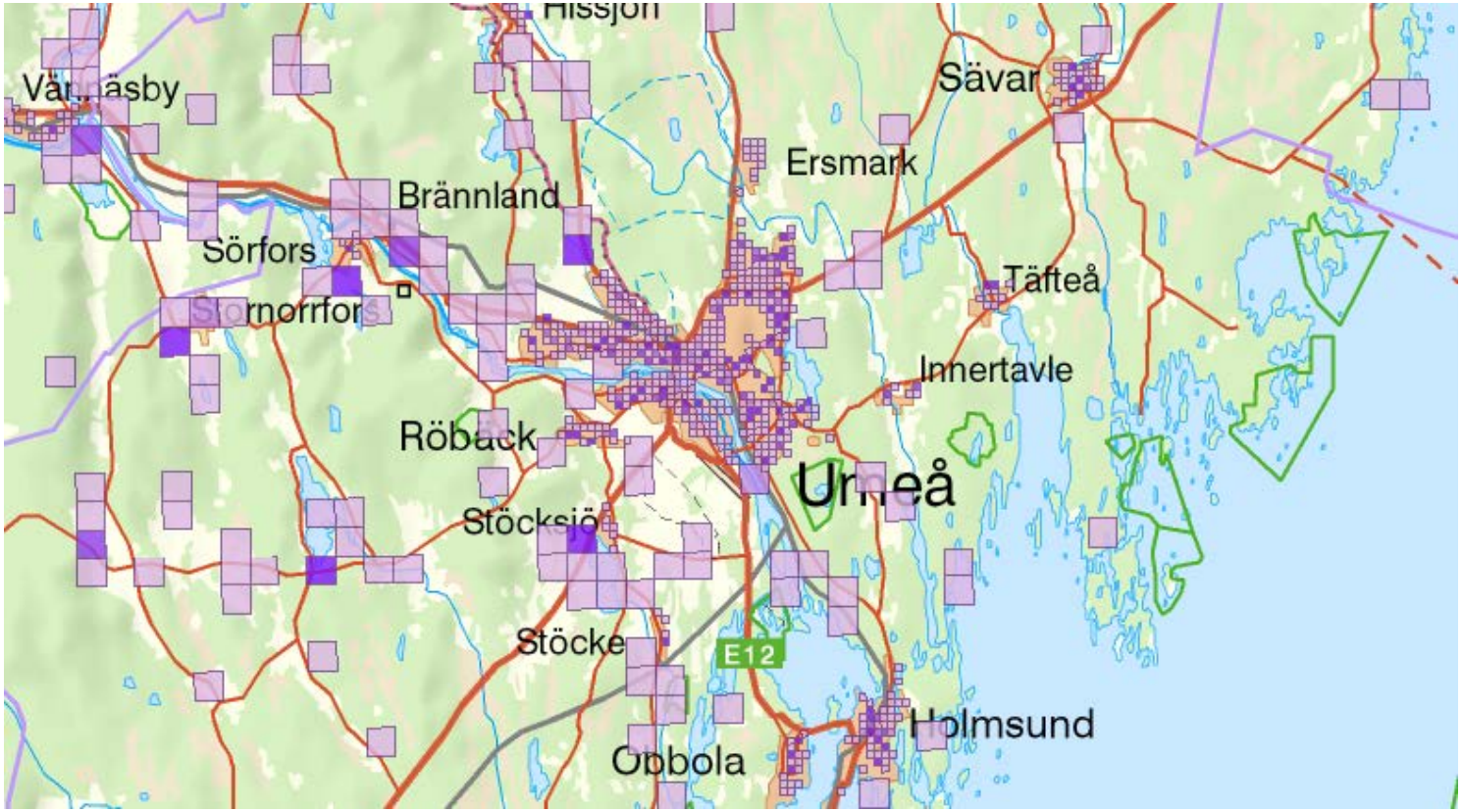
Preparing for extreme events:

Flooding as a consequence of dam failure



Planning, find best location:

Where do people work?



The GEOSTAT Projects

- GEOSTAT 1: Creating a population grid for Europe (2010-2014)
- GEOSTAT 2: A point-based foundation for statistics (2015-2016)
- GEOSTAT 3: The ESS Statistical Geospatial Framework (2017-2018)

GEOSTAT 2

- Propose a model for a point-based geocoding infrastructure for statistics based on address, buildings and/or dwelling registers.
- A priority of the proposed setup has been the ESS vision of a fully geocoded population census 2021.
- The model should be considered suitable for statistics in the widest possible sense.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Project Team

Marie Haldorson, Jerker Moström & Karin Hedeklint (Sweden)

Erik Engelien & Ola Nordbeck (Norway)

Rina Tammisto (Finland)

Vincent Loonis (France)

Ingrid Kaminger (Austria)

Amelia Wardzińska-Sharif (Poland)

Ana Santos (Portugal)

Jerker Moström and Rina Tammisto has kindly shared their presentations, this presentation include a number of their slides!

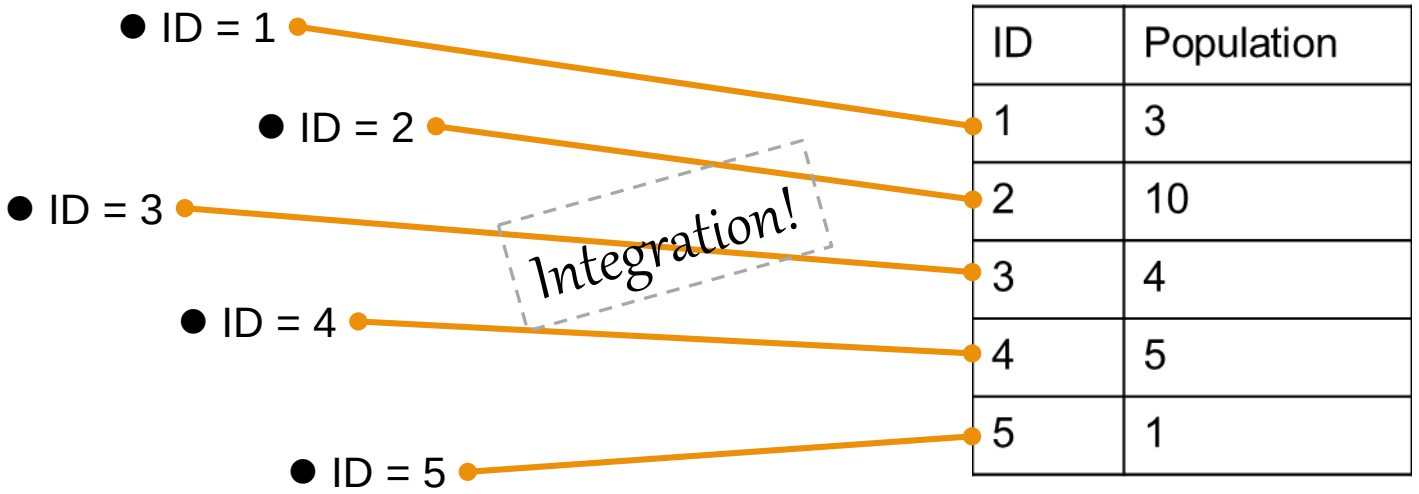
EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

The Concept

Location data

Unit record data

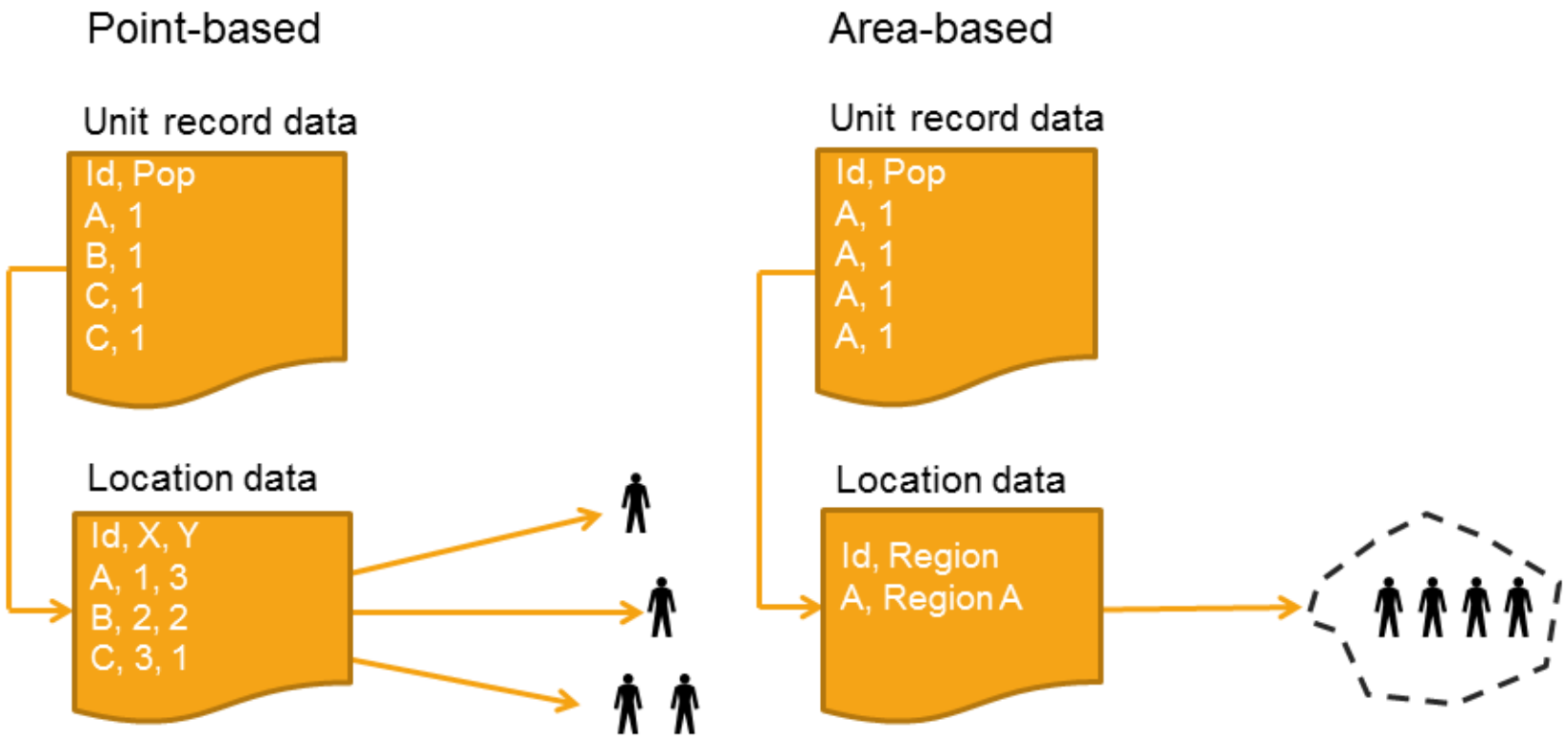


ID	Population
1	3
2	10
3	4
4	5
5	1

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

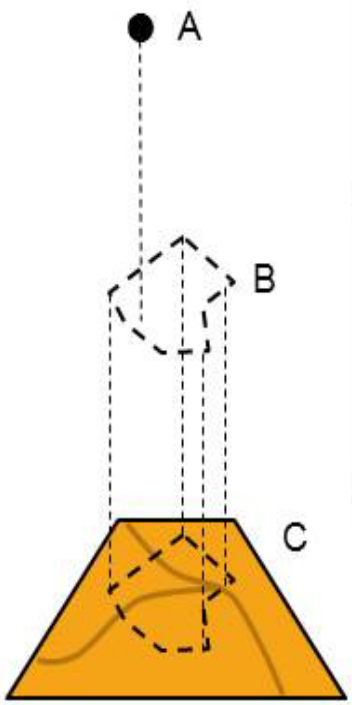
The Concept



EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Geospatial infrastructure:
The proposed ”Data Model”



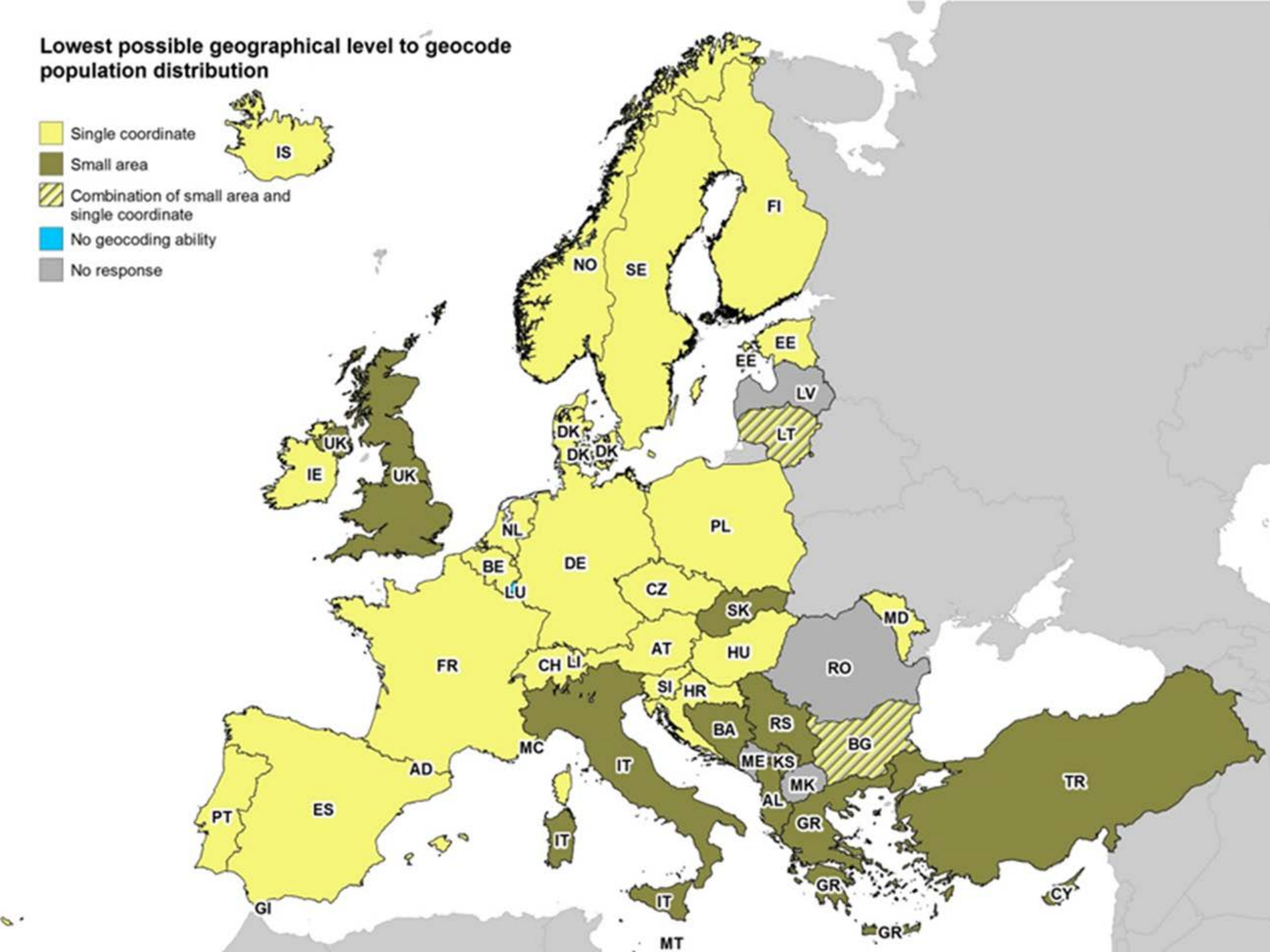
Tier 1- Address location, enumeration districts, statistical regions and grids

Tier 2- Buildings/dwellings, cadastral parcels, transport network, land parcel information

Geospatial Infrastructure data

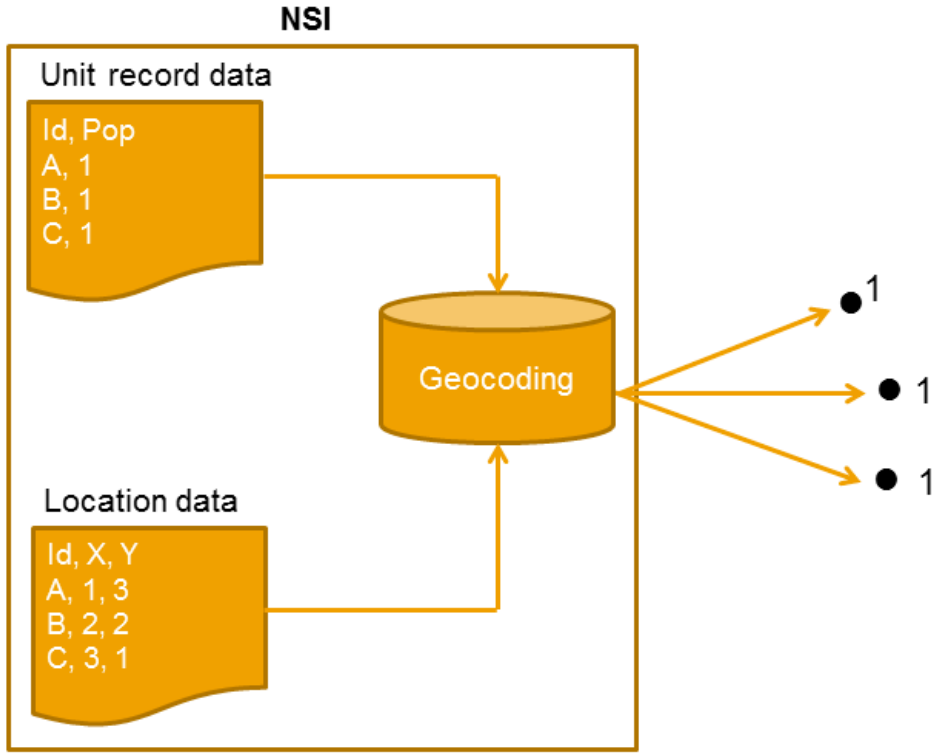
Tier 3 – Authoritative topographic data, ortho photo and satellite imagery, land cover data, DEMs

Geospatial data for statistical content


Lowest possible geographical level to geocode population distribution
Single coordinate
Small area
Combination of small area and single coordinate
No geocoding ability
No response
IS
NO
SE
FI
EE
EE
LV
LT
DK
DK
DK
UK
IE
UK
NL
BE
LU
DE
PL
CZ
SK
MD
AT
HU
RO
FR
CH
LI
SI
HR
BA
RS
BG
MC
IT
ME
KS
MK
AL
GR
TR
AD
PT
ES
IT
GR
CY
GI
IT
GR
MT

Different Approaches in EU Countries: "In-house"

- Both location data and statistical data are collected and managed completely within the NSI

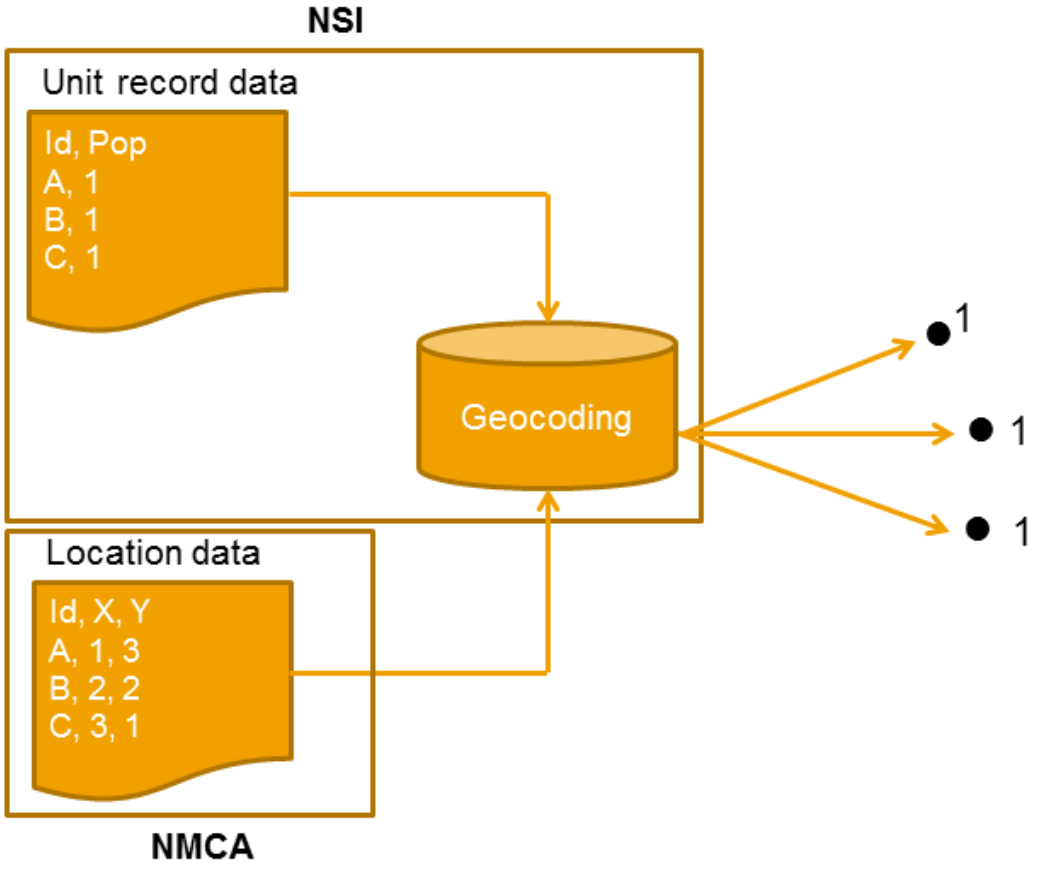


EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Different Approaches in EU Countries: "Hybrid"

- Location data is collected and managed outside the NSI and statistical data within

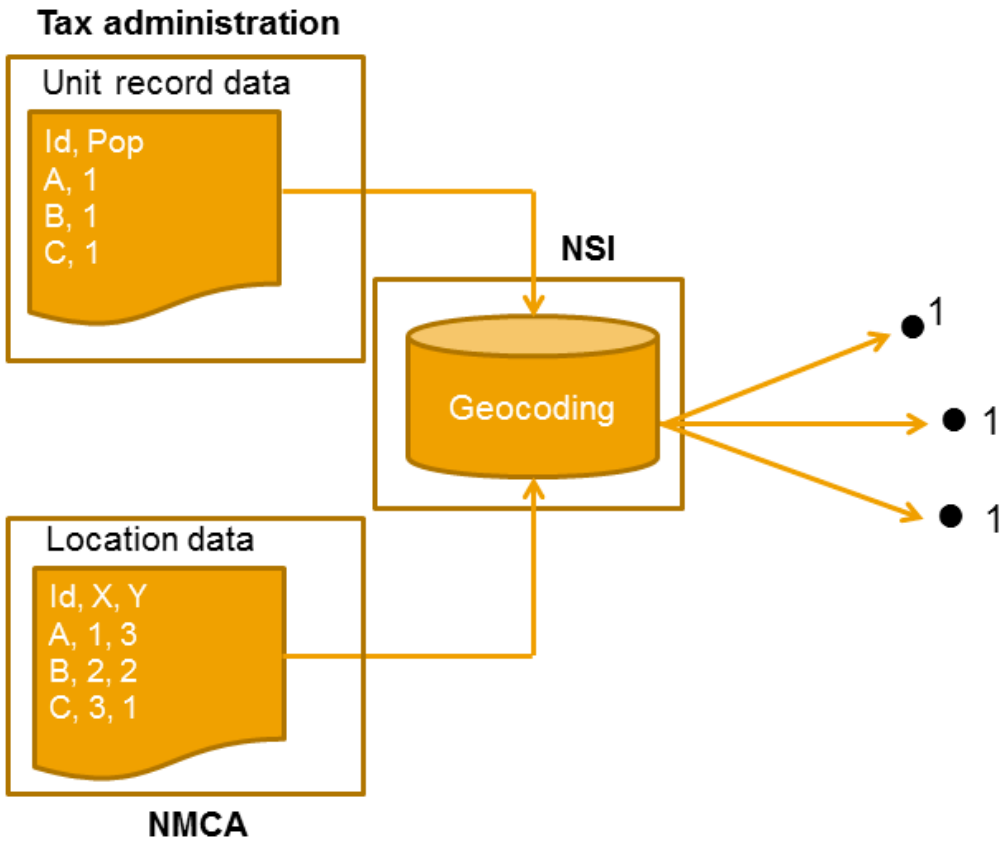


EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Different Approaches in EU Countries: "Data Broker"

- Both location data and statistical data are collected and managed outside the NSI



EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Processes

Including Geospatial Processes in the Generic Statistical Business Process Model

GSBPM


Specify needs
Design
Build
Collect
Process
Analyse
Disseminate
Evaluate
Metadata Management
Quality Management

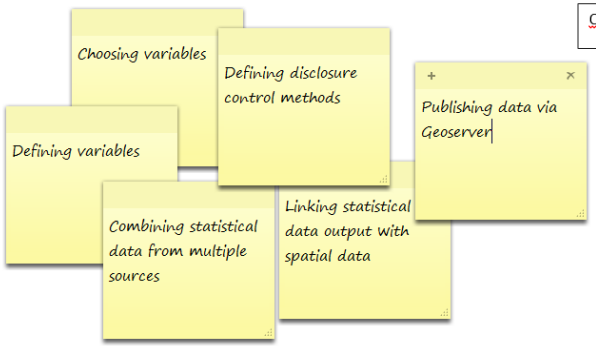
GSBPM + Geospatial

- GSBPM lack references to geospatial data and processes
- It is possible to include geospatial elements in the generic process descriptions
- Useful for identifying where geospatial expertise is needed
- Common language and understanding -> quality

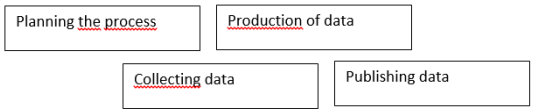
Story Mapping Method

1. Writing down free cognitive associations (tasks) of the process
2. Classification of notes and identification of phases + completion by phases still missing
3. Identification of GSBPM phases + completion by phases still missing
4. Creation of the model + recognising actors

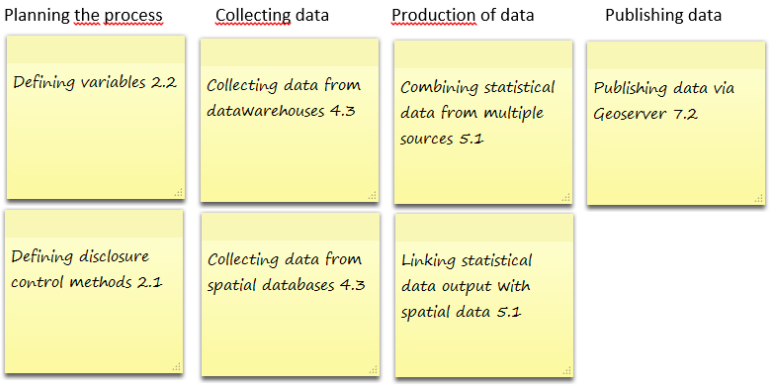
TASKS: CLASSIFICATION



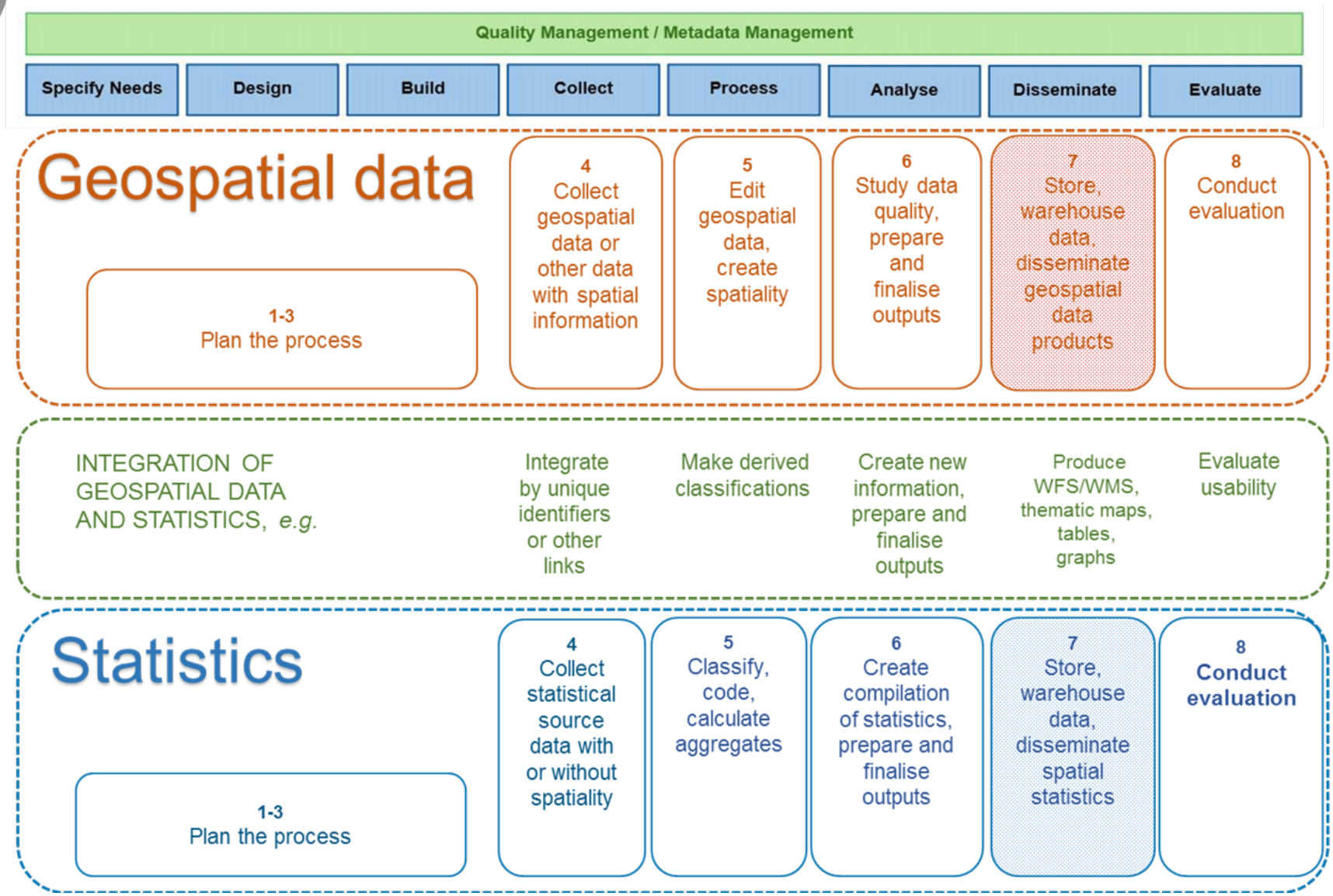
IDENTIFICATION OF PHASES



IDENTIFICATION OF TASKS IN GSBPM

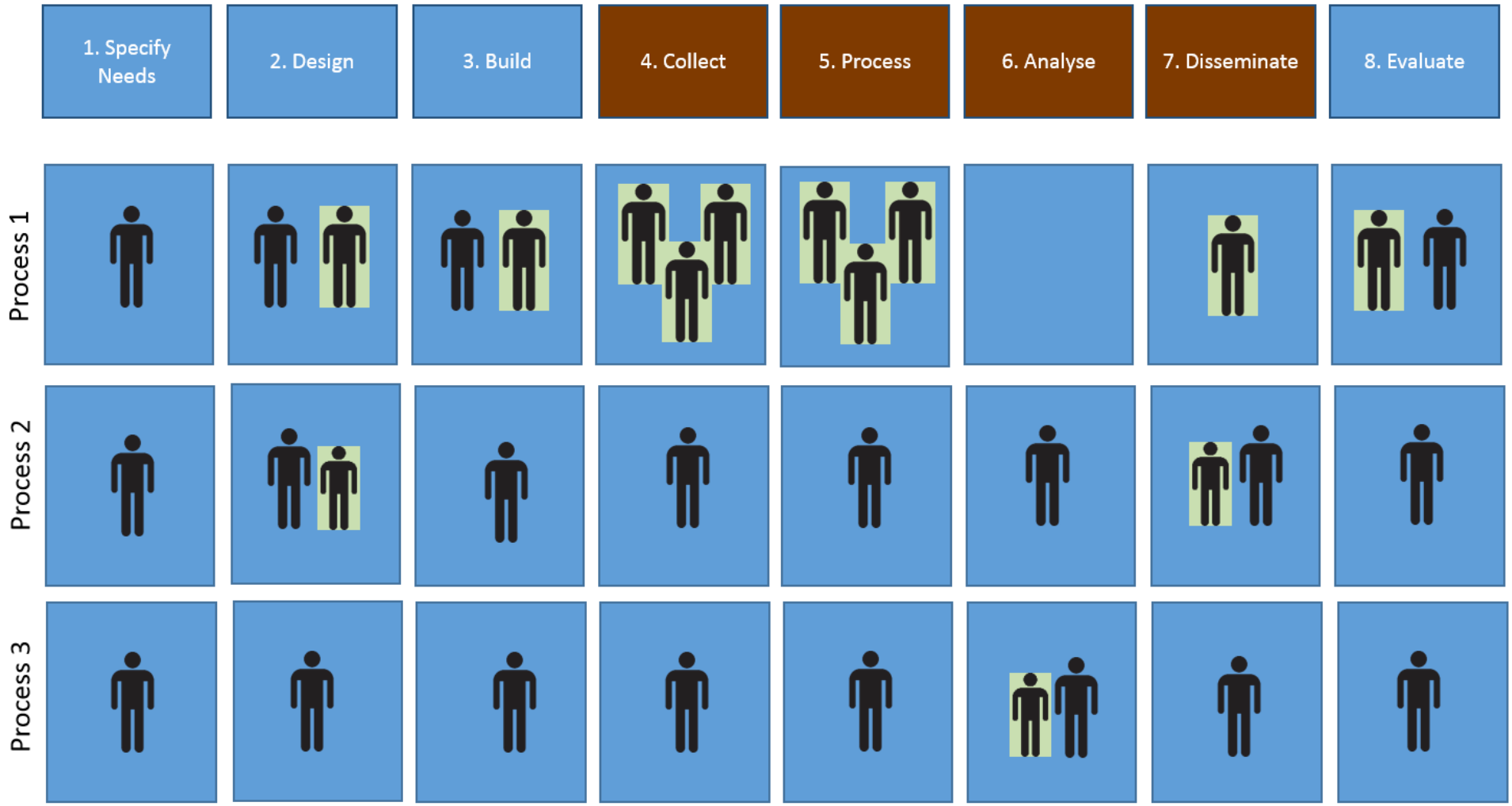


Defining processes: GSBPM



Source: GEOSTAT 2

Need for geospatial data and expertise



Lessons learned in Finland

1. Modelling not a goal in itself, important to **understand the processes**
2. Describe the process as it is, after selecting the level of granularity
3. Apply the model, use it as a tool

Statistics Finland

National experiences

- The GSBPM and geospatial specialists were found, the discussions were fruitful, their support was on the right level to promote the work
- The right persons were found, discussions were partly fruitful but practical results were missing. Applying the GSBPM seemed to be too big a workload
- The GSBPM experts were found but discussions produced no proper benefits, no support for applying the GSBPM
- The project member could not find any support for the work

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

The ”Cookbook”

Getting started

Getting started: Key Tasks

1. Find out what the users need
2. Promote geospatial statistics and the potential of geospatial information
3. Recognise geospatial data sources
4. Assess data processing capacity
5. Specify geospatial statistics output
6. Create a flexible production set-up

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

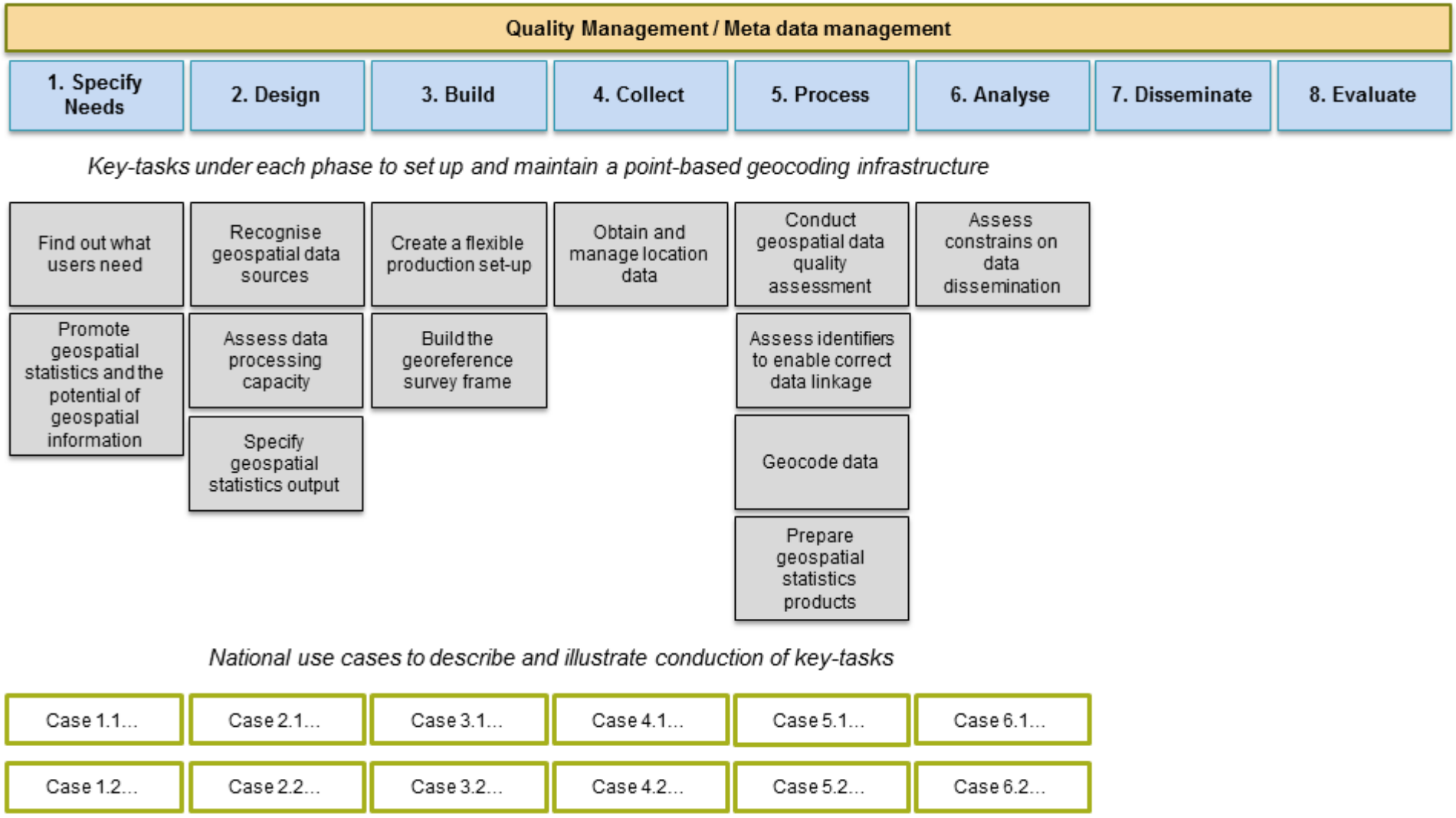
Getting started: Key Tasks cont.

7. Build the geocoded survey frame
8. Obtain and manage geospatial data
9. Conduct geospatial data quality assessment
10. Assess identifiers to enable correct data linkage
11. Geocode data
12. Prepare geospatial statistics products
13. Assess data dissemination constraints

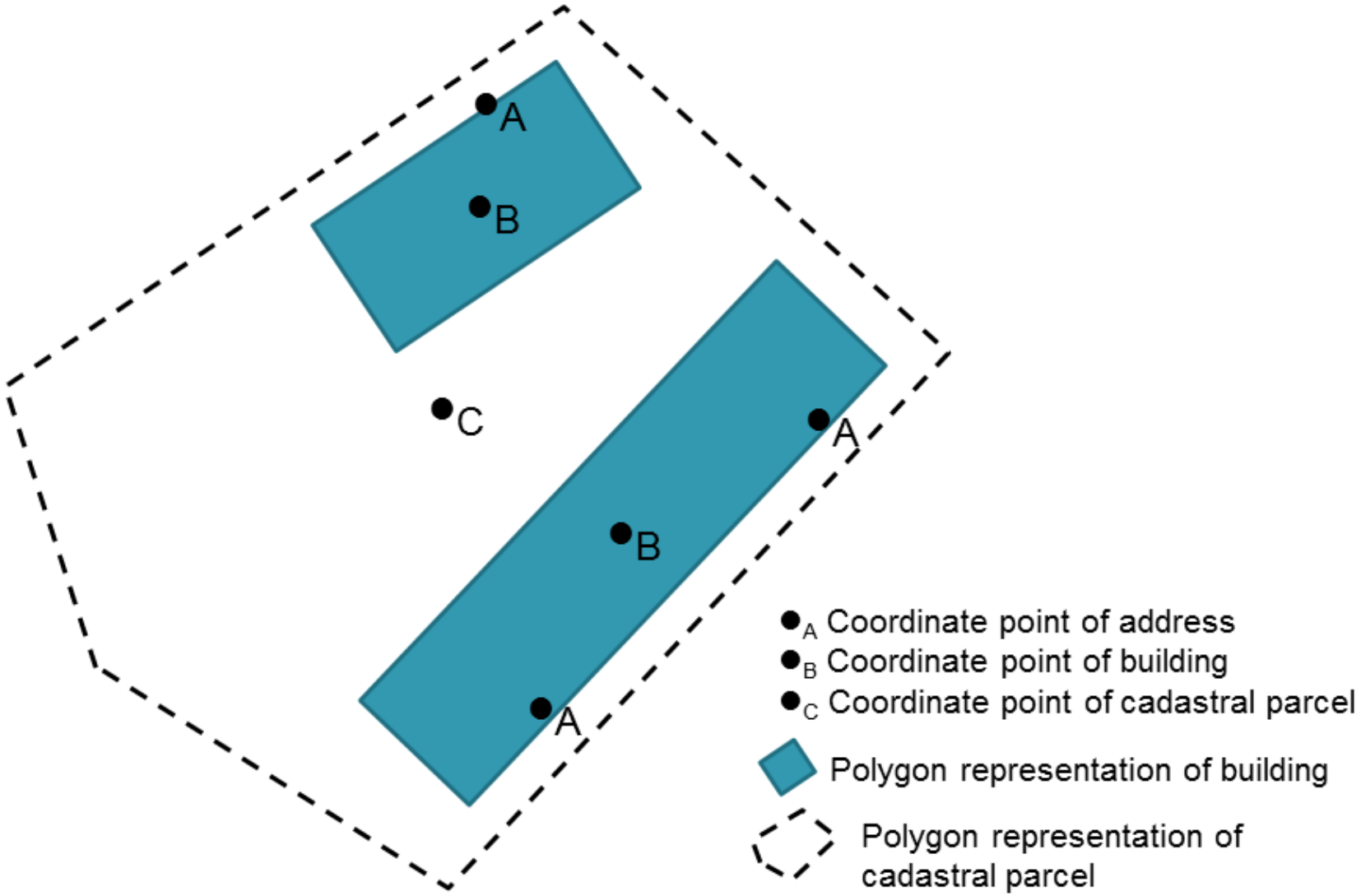
EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Key Tasks aligned with GSBPM



Example from ”Design”:
Recognise geospatial data sources



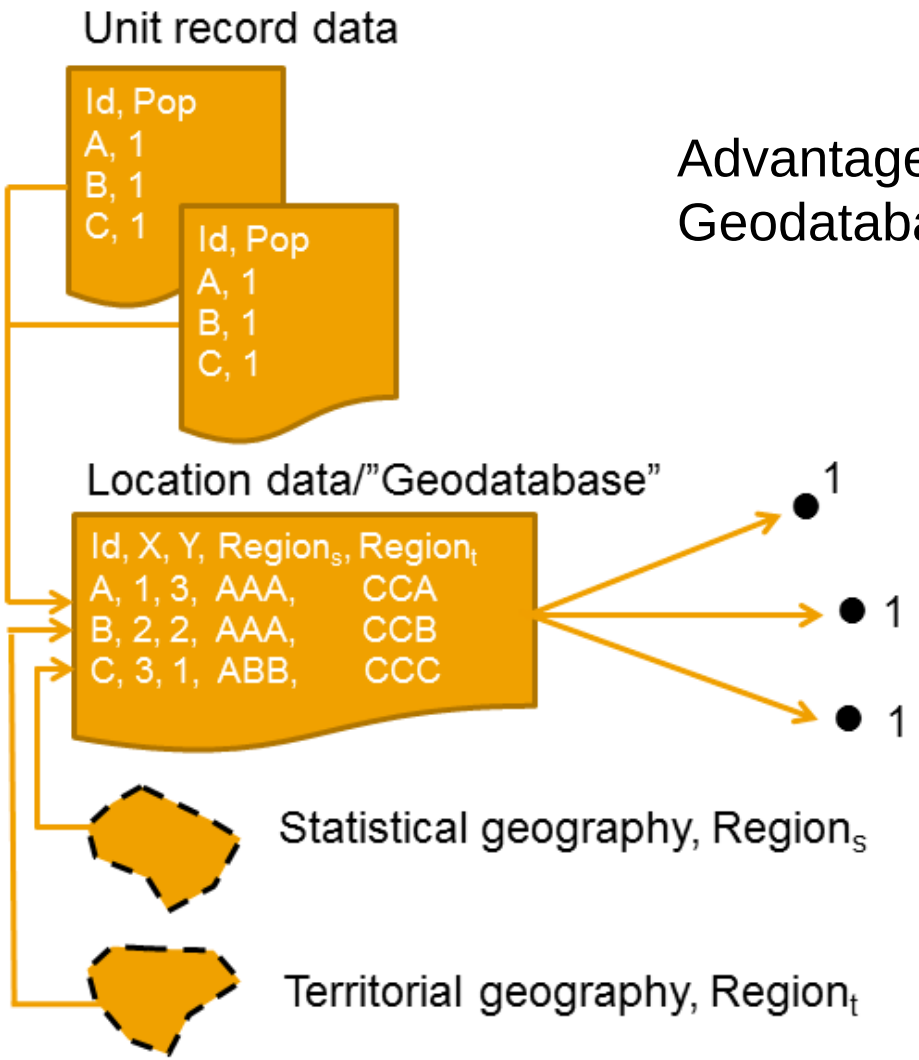
Guidance: Which data source to choose?

- Completeness and coverage
- Existence of sustainable maintenance
- Existence of data specifications and metadata
- Consistent use of geocodes or identifiers
- Suitability of location data
- Spatial accuracy
- Coordinate reference systems

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Example from ”Build”:
Create a flexible production set-up

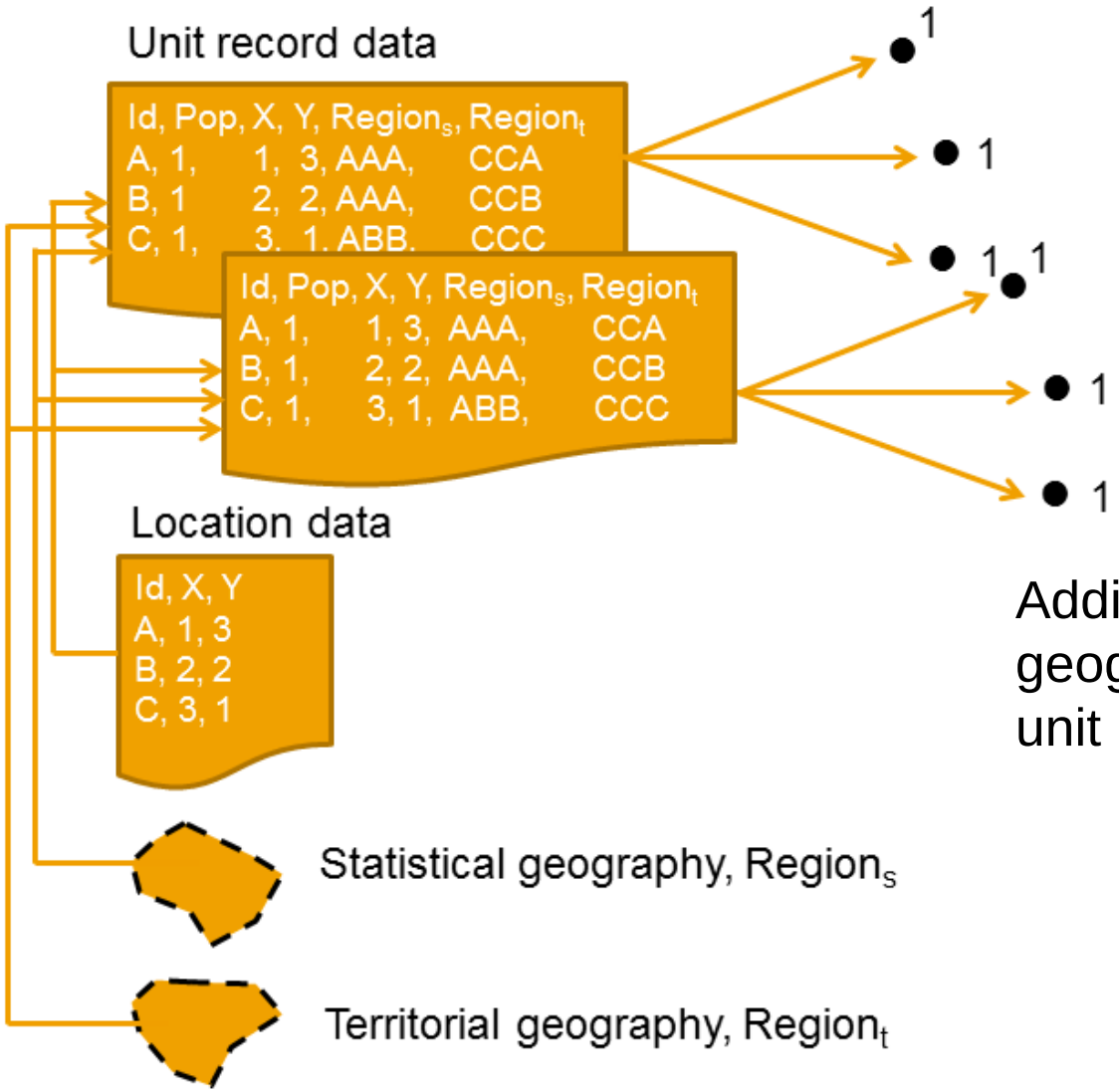


Advantages of a Geodatabase

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Example from ”Build”:
Create a flexible production set-up



Adding geography to unit record data

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Conclusions and Recommendations

General Conclusions

- “Location” needs to be tightly and fully integrated into the statistical production process.
- A point-based foundation for statistics enables NSIs and the ESS to increase relevance and timeliness of statistics but also to improve efficiency and flexibility of production.
- A point-based geocoding infrastructure for statistics is the corner stone of such foundation.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Recommendations (1)

- Make sure to obtain a good understanding of the expectations and requirements of the users, both externally and internally.
- Introduction of a point-based foundation for statistics will and should ideally impact several domains of the statistical production chain (survey design, collection methods, processing techniques, disclosure etc.)
- Engage the organisation broadly to find out how and where a point-based infrastructure for statistics can improve business processes.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Recommendations (2)

- Make a thorough assessment of location data sources that can be potentially used to set up a point-based geocoding infrastructure.
- Fundamental and authoritative geospatial data from the National Spatial Data Infrastructure should be the first-hand option!
- Make sure that the provider has a consistent scheme of maintenance to keep the dataset updated and that data is properly described in line with metadata standards or any other official data specifications.

Recommendations (3)

- Build formal working relationships with external producers of geospatial information (e.g. NMCAs) as to safeguard long-term provision of data.
- Cooperation between institutions should ideally rely on (formal) agreements or legislation, but the agreements themselves are no guarantee for a good cooperation! (UN GGIM: Europe)
- Cooperation with data custodians should entail feed-back routines for reporting and correction of errors found in geospatial information.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Recommendations (4)

- Develop a strategy to manage the geospatial data streams in the best possible way.
- Address human resources needed as well as the technical infrastructure allowing for efficient processing of geospatial data (software, data storage, computing capabilities etc).
- The goal should be a flexible technical environment and to avoid duplication of data and smooth processes for maintenance.

Recommendations (5)

- Develop routines to deal with the temporal aspects of data in order to obtain the best possible temporal cohesion between location data and statistical and administrative data to be geocoded.
- Develop routines for a uniform approach to geocoding, most notably workarounds to handle erroneous data (address validation tools, homogenisation of address information or interpolation of address location points).

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

GEOSTAT 2 – Final Report

- Main report – explaining concepts, providing recommendations
- Annex 1. Use cases (linked to recommendation in the main report)
- Annex 2. Results from national exercises on assessing the GSBPM to structure management of geospatial data in production + recommendations for revision of the GSBPM
- Annex 3. Ten-point-guide to use the GSBPM
- Annex 4. Detailed proposals for improving the GSBPM

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

http://www.efgs.info/geostat/

eurostat

GEOSTAT 3

GEOSTAT 3 Objectives

1. Develop the European adaption of the SGF (ESS-SGF) and provide recommendations on how to implement it in EU member states within and outside NSIs;
2. Test the usefulness of the ESS-SGF and its implementation on selected indicators for the 2030 Agenda;
3. Support EFGS, disseminate results, benefit from GGIM activities etc.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

Project Team

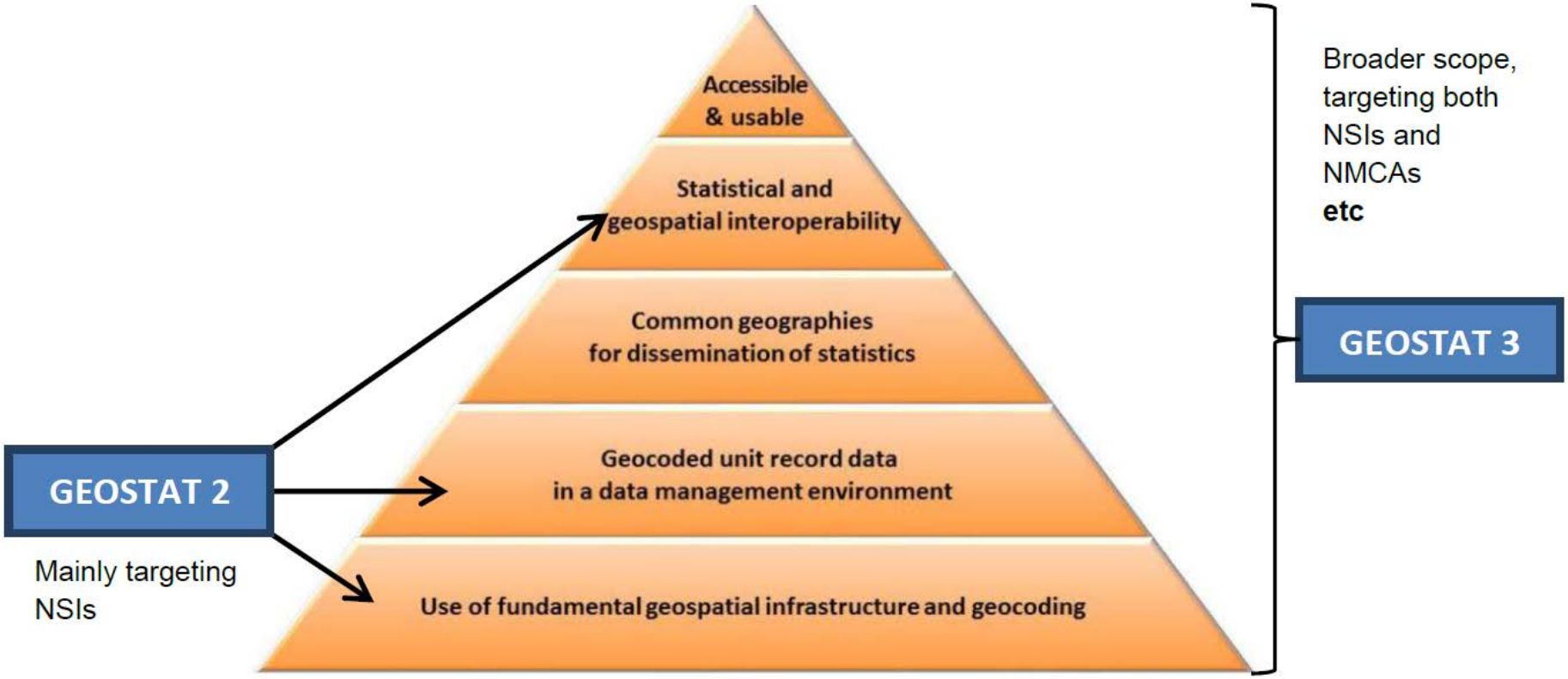
- Coordinator: Sweden
- Participating NSIs:
  - Austria
  - Estonia
  - Finland
  - Netherlands
  - Norway
  - Poland
  - Portugal
- Sub-contractors:
  - Kartverket, Norway
  - BKG, Germany
  - Marina Backer (EFGS Secretary)
- Sponsor: Eurostat

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat

WP-1:

Develop a European Statistical Geospatial Framework

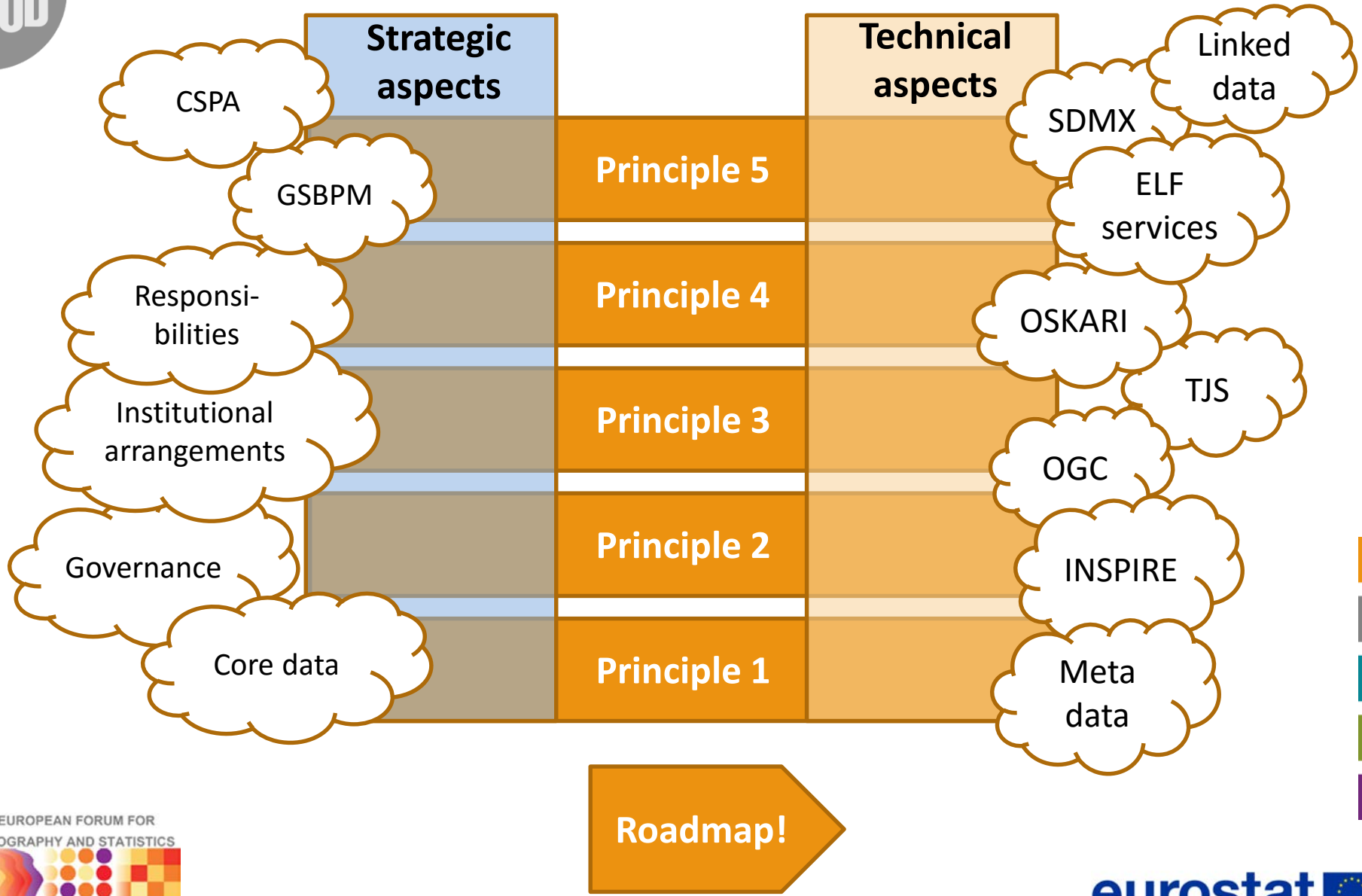


Deliverables of WP1

- A detailed description of all the principles and elements of the ESS-SGF and the differences with the global SGF.
- Strategic and technical recommendations for the implementation of the ESS-SGF in NSIs but also, where needed and/or possible, in other public authorities concerned.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

eurostat


Strategic aspects
Technical aspects
CSPA
GSBPM
Responsi-bilities
Institutional arrangements
Governance
Core data
Principle 5
Principle 4
Principle 3
Principle 2
Principle 1
Linked data
SDMX
ELF services
OSKARI
TJS
OGC
INSPIRE
Meta data
Roadmap!

WP-2 Testing the ESS-SGF

- Applying the ESS-SGF to the indicator framework of the 2030 Agenda
- Applying the ESS-SGF to Census indicators (1 km grid)
- Implementing technical considerations from WP 1.
- Choice of indicators to be tested!

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS

GGIM UNITED NATIONS INITIATIVE ON GLOBAL GEOSPATIAL INFORMATION MANAGEMENT

eurostat


More information:
http://www.efgs.info/geostat/