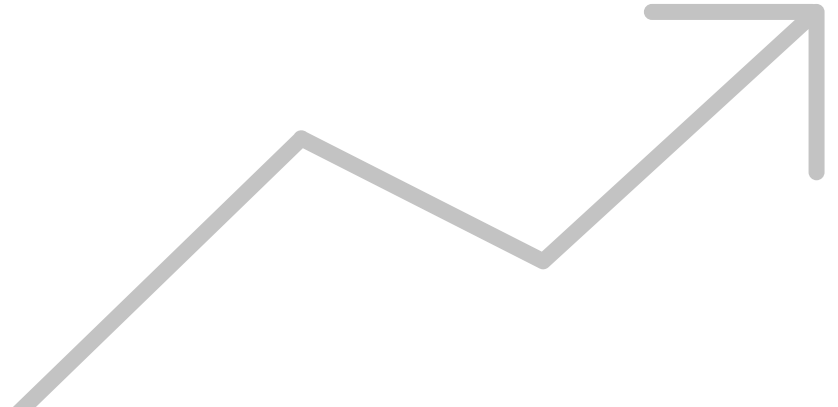
DESTATIS
Statistisches Bundesamt

It's the quality that counts

Florian Dumpert

Federal Statistical Office of Germany

72nd plenary session of the Conference of European Statisticians – Geneva, June 2024



Official Statistics

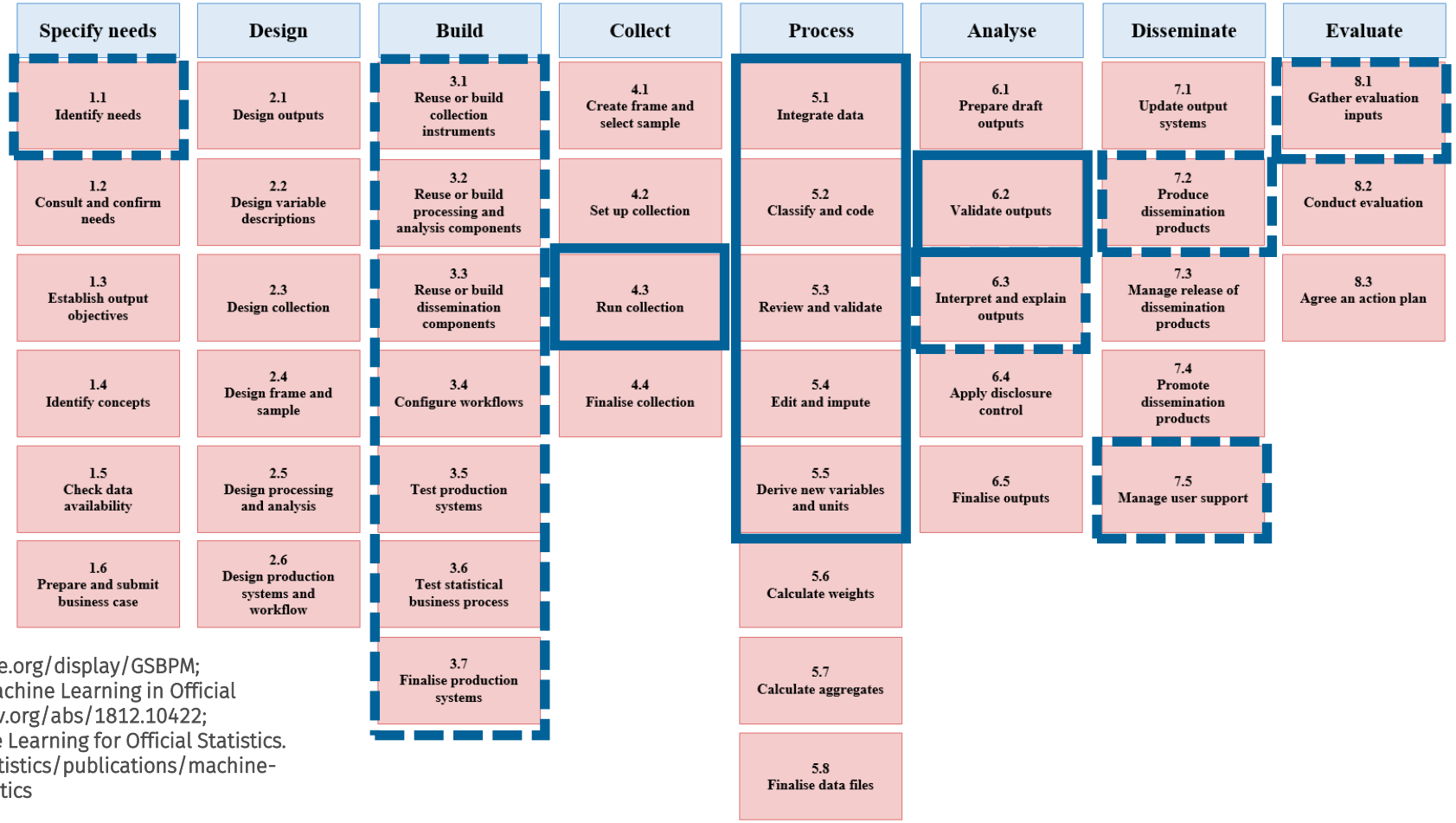
The Generic Statistical Business Process Model (GSBPM)

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

https://statswiki.unece.org/display/GSBPM

Official Statistics

The Generic Statistical Business Process Model (GSBPM)



Core fields for machine learning

Relevant fields for gen. AI

https://statswiki.unece.org/display/GSBPM;
Beck M et al (2018) Machine Learning in Official Statistics. https://arxiv.org/abs/1812.10422;
UNECE (2021) Machine Learning for Official Statistics. https://unece.org/statistics/publications/machine-learning-official-statistics

Applications

- **Mexican National Institute of Statistics and Geography frameworks**
  - Bringing together census data with spatial data (Landsat)
  - Aim: monitoring urban sprawl
  - Comparison of two image segmentation methods

- **Turkish Statistical Institute**
  - Text to code classification for:
    - occupation
    - product
    - education
    - etc.
  - Based on e-invoices, scanner data, web scraping

Coronado A (2024) Comparative analysis of unsupervised and deterministic satellite image segmentation methods for urban monitoring using machine learning. https://unece.org/sites/default/files/2024-05/ECE_CES_2024_24_E.pdf;
Kurban B (2024) Artificial Intelligence (AI) and Generative AI Experiences at Turkish Statistical Institute. https://unece.org/sites/default/files/2024-05/ECE_CES_2024_22_E.pdf

Starting Points

- **Quality frameworks** concerning
  - aspects of the institution
  - aspects of the processes
  - aspects of the products/outputs
- On different levels
- Goal: **fitness for use** of the products

https://www.destatis.de/DE/Methoden/Qualitaet/qualitaetshandbuch.html; https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework; https://unstats.un.org/unsd/unsystem/documents/UNSQAF-2018.pdf

Proposed Quality Dimensions

- Accuracy
- Robustness
- Explainability
- Reproducibility
- Timeliness & Punctuality
- Cost-effectiveness

Predictions	„phenomenon is described correctly“
	„stable results despite small perturbations“
Model	„understand how results are generated“
Infrastructure	„reproduce identical results“
	„deliver up-to-date results punctually“
Processes	„appropriate costs“

Yung W et al (2022) A quality framework for statistical algorithms. Statistical Journal of the IAOS, 38(1), 291–308;
Saidani Y et al (2023) Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik. AStA Wirtschafts- und Sozialstatistisches Archiv, 17(3-4), 253–303

DESTATIS
Statistisches Bundesamt

Thank you!

Florian Dumpert
Federal Statistical Office of Germany

florian.dumpert@destatis.de
+49 611 75-3887

www.destatis.de