# Practical approach in developing the integration of statistics and geospatial information through value chains and data quality

Rina Tammisto[1], Godfrey Lowndes[2,] Mervi Haakana[3]

[123]*Statistics Finland, Finland*

## Abstract

One challenge for the smooth integration and interoperability of statistical and geospatial information lies in ensuring a seamless flow of data between administrative sectors. Achieving interoperability necessitates cooperation, both between and inside of organisations, and a shared understanding of specific development measures and their benefits and agreed-upon implementation.

In the joint GSGF in Finland EU Grant project (2023 to 2025), Statistics Finland, the Land Survey of Finland, and the Finnish Environment Institute aim to lay ground for the future development of integration of statistics and geospatial information in Finland. The project seeks to establish a national adaptation of the GSGF model (GSGF Finland), aligning it with the Global Statistical Geospatial Framework (GSGF) and its European counterpart (GSGF Europe).

Adapting the frameworks to the national level requires more practical tools; as part of our approach, the project is piloting value chains and utilising the quality framework. This work started by conducting a current state analysis, focusing on a case study on geocoding process for Statistics Finland's Business Register. The evaluation encompassed the statistical production process from data collection to dissemination of statistical products. The evaluation was grounded in the concept of data value chains, tracing information flow of geospatial data from one actor to another and across different production phases. Notably, the innovation emerged from recognizing the significance of quality as the value chain unfolds.

In this pilot, the national data quality framework served as the quality model. Its eleven perspectives, ie. correctness, accuracy, consistency, currency, completeness, traceability, understandability, compliance, portability, user rights and punctuality, were used in analysing the quality changes in the piloted process.

This paper shed more light on the evaluation of the geospatial-statistical production process from a quality perspective, the adapted value chain concept behind it, and initial thoughts on how the results can be utilised as a basis for development roadmap.

**Keywords:** data value chains, geospatial information, quality framework, GSGF Europe, interoperability

# 1    Introduction

In the GSGF in Finland EU Grant project (2023 to 2025), information value chains were examined at the request of the national project's steering body. Through the case study of the Business Register, the flow of information and the changes in its value through the information process were analysed. As the work progressed, it was observed that the information introduced at a certain stage of the process may not necessarily meet the needs of the next stage and may require correction or modification. These value degradations can be observed both within an organisation's own process and more likely when information flows from one organisation to another. An innovative idea emerged during the project: Utilising the national Data Quality Framework for value chain analysis, particularly focusing on value degradations. By adopting a value chain perspective, we can understand, describe, and measure the quality of information at critical points in its flow. The 11 quality criteria of the Data Quality Framework provide a standardised structure for assessing information quality at these critical points. This situational understanding can be shared with other stakeholders or other process stage participants, and successful collaboration can lead to specific changes that eliminate value reductions or reduce flow impediments. As a result, information flow becomes more efficient, streamlining processes where information flows directly.

# 2    Value Chains

The value chain perspective allows us to describe and understand the increase in value of information within the context of information flow. Information is produced from a specific viewpoint and provided to users as predetermined or otherwise defined. The original information producer acts according to what has been deemed best for producing information, and as required by legal obligations. However, within the information ecosystem, actions are more complex, and not all user needs are necessarily known to individual information producers—or there may be no means to pay attention to them. Nevertheless, information flows despite these complexities, although not as efficiently as it could.

By applying the value chain perspective, we can examine the flow of data needed for and produced by statistical processes. In this case, the viewpoint for examination is that of the entity responsible for the statistical process. Other actors in the broader data process value chain include the original data producers and data processors prior to the start of the statistical process, as well as data users and processors of statistical products after the statistical process.

## 3 Data Quality Framework and Quality Criteria

Efficient utilisation of information requires investment not only in describing the content of data sets but also in assessing the quality of data usage restrictions. To address this need, a data quality framework has been created. The framework consists of quality criteria that describe data quality and metrics that provide concrete values for these criteria.

Data quality descriptions provide information about data sets to both the data set owner and external reviewers. Producing these descriptions also encourages the use of consistent terminology, thereby promoting data interoperability. Consistent descriptions allow for comparing the quality of different data sets and selecting the most suitable data for specific purposes.

## 4 Value Chain and Quality Framework in Information Processes

In the ideal state, the data used for statistical production is a seamless part of a broader information ecosystem. Where necessary data is readily available for use, its content aligns with information needs, and the quality of the data is high, requiring no corrective actions.

However, in practice, data doesn't always flow as efficiently as desired. Data may be difficult to access due to technical or rights-related reasons. Separate efforts may be needed to obtain the data, or agreements may need to be made regarding details for repeated data retrieval. It's also possible that the structure of the data doesn't match the required format for utilisation, leading the receiving party to initiate their own data processing to align the structure appropriately. The most challenging situation arises when the available data contains outright errors that become apparent when combining it with other data in the statistical process.

It's important to note that not all of the aforementioned issues necessarily result from negligence. Data producers may genuinely believe that the quality of the data they provide is good. The quality of information, however, is determined by its use, and it's possible that the information producer may not be aware of all the intended uses or specific needs of information users. Often, the information producer may not receive feedback about corrective actions taken by recipients in their processes related to the information.

Through value chain descriptions, it is possible to create a shared understanding of information flow and gain insights, especially at critical points in the chain where information transitions from one actor to another or from one process stage to another. Describing value chains provides a broad view of information flow, allowing the original information producer to understand the subsequent life of the information in the hands of other actors. Even small technical changes can support and enhance the entire chain of information utilization. Seeing

the entire chain likely involves traversing multiple administrative domains, highlighting that operating in separate silos restricts the optimal flow of information within the information value chain.

A common language for discussing the effectiveness or ineffectiveness of value chains can be found in the information quality framework. The 11 perspectives on quality within the framework provide structure for discussions about the attributes of information and its delivery, including what is working well and what needs improvement.
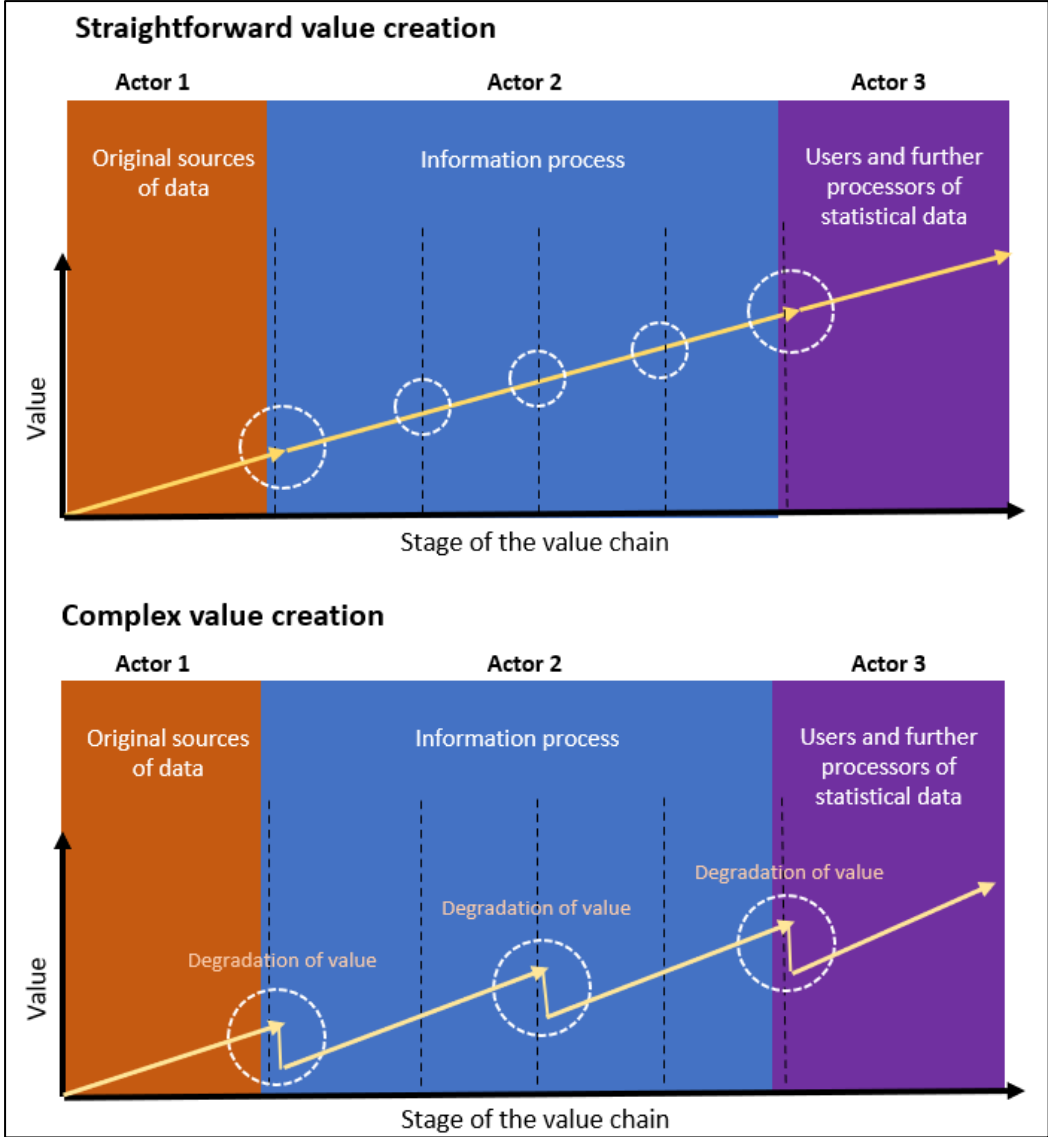


Figure 1: Straightforward and complex value creation. The value of information increases as it progresses through the information value chain. In a straightforward value chain, information advances without quality degradation, allowing it to be directly usable for the next actor or in the next process stage. In a complex value chain, the increase in the value of information is preceded by a decrease in value, which requires data processing. Through the editing processes, the value of information increases again.

## 5　Value Chain and Quality Framework - Examples from Statistical Geospatial Data Production

The recently (2020-2021) renewed Geocoding process of Statistics Finland's Business Register has proven be an excellent tool for defining value streams and chains of statistical information. Traditionally only the quality of ready statistical products is measured and by doing so everything might appear to be great. National Statistical Institutions also readily accept any administrative data sources available to produce statistics and resort to expensive direct inquiries when other the other sources aren't sufficient.

By describing the example process as a value chain, it is apparent that the insufficient administrative data sources cause a drop in information value when past from the source and received by National Statistical Institution. This drop in value has to then be compensated by National Statistical Institution by implementing inefficient expensive direct inquiries and editing processes. Furthermore, the drop in value is multiplied if the insufficient administrative data source is widely used across the public sector.

## 6　Data Quality Management in Information Processes

In the ongoing project, work has been initiated to understand how to create conditions for comprehending and managing the aforementioned data value chain in the information process and more broadly within the information ecosystem. At this stage of the work, three key areas have been identified as means to address data quality management.

Guidance, definition of information needs, and agreement: Strong commitment is needed between various stakeholders in the processes to understand quality and enhance the flow of information. This requires a willingness to view quality as part of the overall process and to understand the roles of different actors within it. Process management should be committed at all levels, from guidance to practical actions, so that practical details can be transformed into agreements at the steering level.

Monitoring of key points and measurement points: Processes are constantly changing. Information, technologies, and responsibilities evolve, so monitoring the flow of information is necessary to ensure that the relevant factors are measured despite changes.

Automation of quality assessment in processes: Quality measurement needs to become an established practice. In practice, only through automation can we ensure that quality measurement is an integral part of the process of utilising information.

# 7 Conclusions

Statistical processes can benefit from value chain approach as the focus is on collaboration between different organisations and understanding the needs of different actors in the value chain. The value chain approach can help to improve data sources and efficiency through better understanding and collaboration.

## Acknowledgment

## References

GSGF in Finland (2024). *Data Value Chains and Quality in the Utilisation of Geospatial Information for Statistical Production*. Project work document (in Finnish, ongoing project).