

DATA, INFORMATION AND KNOWLEDGE MANAGEMENT

Data, Information and Knowledge Management

Chapter **14**





14.1 Introduction

This chapter covers the management of data, information, and knowledge in a national statistical office (NSO). Statistical data are the major asset and *raison d'être* of an NSO. Producing statistics information, and knowledge is the core work of an NSO, but NSOs are also leaders in the national statistical system (NSS). With increasing amounts of data becoming available from NGOs, businesses, and other actors, the NSO's role in managing data has expanded. To ensure that data are available to the people who need them, in the right format, and at the right time requires appropriate and well-functioning data systems, information systems, and knowledge management systems. These systems cut across many domains involving information governance, information management, information security, records management, information access and customer information management.

The following activities are covered in this chapter:

- Data and metadata management.
- Information management.
- Knowledge management.
- Management of archives.

Data and metadata management deals with the requirements of the statistical data processing life cycle from collection through dissemination to archiving. Information management covers the ways to organize, analyse, and retrieve data and information and how information management systems can make data and information searchable, accessible, easily retrievable and as widely available as possible within an NSO.

Knowledge management concerns how collective information, knowledge and expertise can be used to be more effective as an organization. It covers the management of data flow, the sharing of “know-how”, knowledge retention, collective information, knowledge and expertise. The management of archives covers policies for data archiving and data retention.

The **Generic Activity Model for Statistical Organizations (GAMSO)** describes and defines the activities that take place within a typical NSO and provides supplementary text for this chapter of the most essential activities needed to manage the data, information, and knowledge in an NSO. In particular, the GAMSO section ‘Manage information and knowledge’ describes the activities covering the ownership or custody of records, documents, information and other intellectual assets. It also covers the governance collection, arrangement, storage, maintenance, retrieval, dissemination, archiving, and destruction of information. It includes the activities needed to maintain the policies, guidelines, and standards regarding information management and governance.

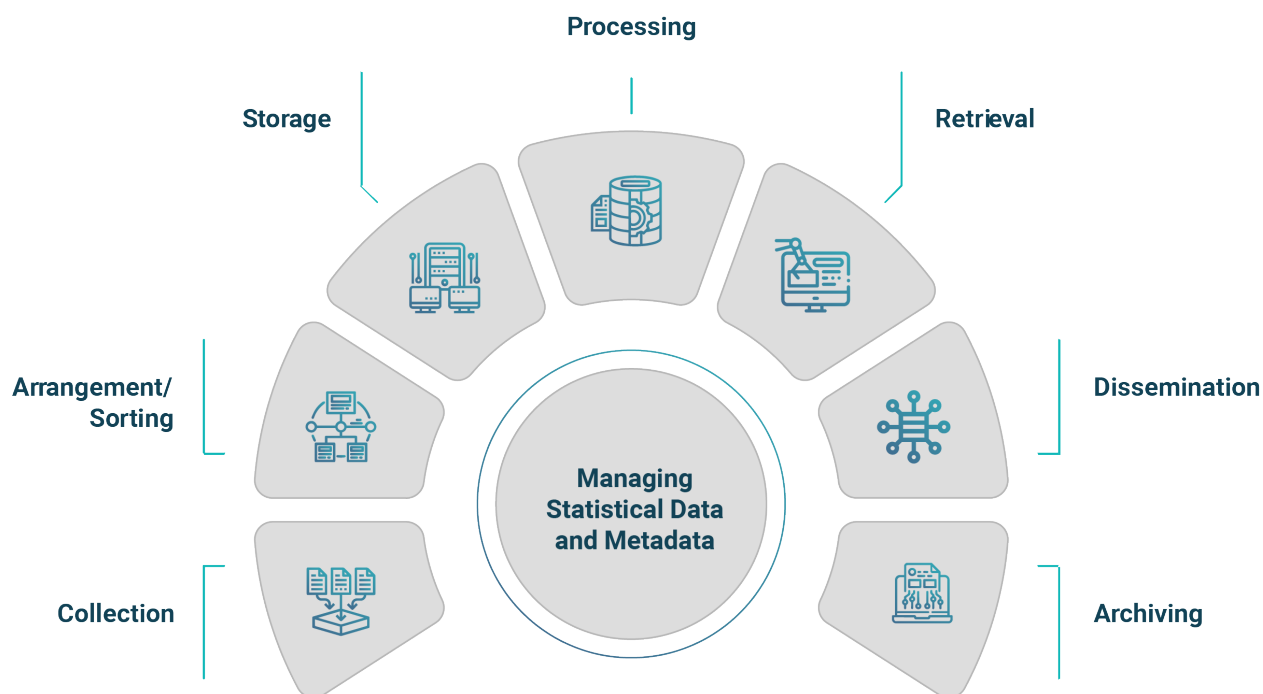
14.2 The relationship between data, information, and knowledge

The relationship between data, information, and knowledge can be described as follows: data consist of collections of concrete facts, measurements or observations. Once data have been categorised, analysed, interpreted and summarised to give them structure and meaning, they become information (statistics). When human experience, expert opinion, and insight are applied to information, it is transformed into knowledge. Data are the basic facts; information is data with context; knowledge is processed information with meaning. Taken a step further, wisdom is knowledge coupled with judgment or insight.




As an example, taking the temperature of a patient generates a number: this is data. Checking the temperature and comparing it against other facts indicates that the patient has a high temperature: this is information. Based on this and other information, a doctor uses expertise to conclude the patient has a particular type of fever: this is knowledge. The doctor then uses wisdom to prescribe an appropriate cure.

14.3 Managing statistical data and metadata

Data and metadata management involves the management of documents and records, archiving, managing knowledge, standards and access rights as well as metadata and data management. A data and metadata management policy should consider the entire statistical production process from data collection to dissemination and archiving as described below (derived from [GAMSO](#)):



- **Collection** covers surveys and censuses (see *Chapter 9.2 — Sample surveys and censuses*) and administrative data. This is typically the most labour-intensive phase of the statistical process. The organizing of data collection operations will need to satisfy a number of strategic goals. The management of collection includes the effective planning and coordination of data collection across statistical domains, deciding what data are needed and what data items need to be collected directly from respondents. It should ensure the continuous improvement of efficiency of data management. It should work towards the effective reduction of response burden and costs for both respondents and the statistical authority and follow the principle of “collect once, use many times” to avoid duplicate requests being made to those providing data. It is fundamentally important to keep individuals’ and businesses’ information secure and confidential, regardless of whether it has been collected from a survey or non-survey sources.

- **Arrangement/sorting** covers the organization of data into collections of data. It includes data modelling, which is the process of creating a data model for the data to be stored in a database. It should, where possible, use existing standards such as SDMX Data Structure Definitions (see [Chapter 15.4.5 — Statistical Data and Metadata Exchange \(SDMX\)](#)) for efficiency, reusability and interoperability.
- **Storage** covers database management systems (DBMS), which are designed to define, manipulate, retrieve and manage data in a database. Such database systems vary according to the capacity of an NSO and can range from simple spreadsheet-based systems to highly sophisticated (hierarchical) database systems (see [Chapter 15.6 — Basic IT infrastructure needs and skill requirements](#)).
- **Processing** covers the sub-processes of the GSBPM process phase which are: integrate data, classify and code, review and validate, edit and impute, derive new variables and units, calculate weights, calculate aggregates and finalise data files.
- **Retrieval** covers the production and dissemination systems an NSO has in place to locate and retrieve data, information and knowledge.
- **Dissemination** covers the systems an NSO has in place to disseminate data. This covers the GSBPM sub-processes update output systems, produce dissemination products, manage release of dissemination products, promote dissemination products, manage user support (see [Chapter 11 - Dissemination of Official Statistics](#))
- **Archiving** covers the systems to archive data and metadata, and in accordance with the rules and regulations removing data (both physical and digital) in line with relevant legislation.
- *Statistics Canada - Data collection planning and management* (.
- *Philippines Statistics Authority (OpenSTAT platform)*, a national data portal that now serves as the central database for publicly available information. (.
- *Philippine Statistics Authority Statistical Survey Review and Clearance System (SSRCS)*. This mechanism aims to: ensure sound design for data collection; minimize the burden placed upon respondents; effect economy in statistical data collection efforts; eliminate unnecessary duplication of statistical data collected; and, achieve better coordination of government statistical activities.
- UK Office of National Statistics - Data use and management policies (.

14.3.1 Managing statistical data

An NSO manages data from many sources including censuses, surveys, administrative data and, increasingly, new, non-traditional data sources often referred to as Big Data (see [Chapter 9.5 — Big Data](#)).

Common Statistical Data Architecture

Guidance for data management is provided by a reference architecture for data, such as the *Common Statistical Data Architecture (CSDA)*. A data architecture is composed of models, policies, rules or standards that govern which data are collected, how they are stored, arranged, integrated, and used in data systems. Well-designed data architecture can result in improved timeliness and more disaggregated statistics at higher frequencies.

The CSDA can support an NSO in the design, integration, production and dissemination of official statistics. It acts as a reference template for an NSO in the development of data architecture (see [Chapter 15.2.13 — Common Statistical Data Architecture](#)) that can guide IT staff in the development of systems to be used by statisticians in the production of statistical products. The CSDA shows how an NSO can organize and structure its processes and systems for efficient and effective management of data and metadata in order to help in the modernization and the improved efficiency of statistical production processes. It can help an NSO manage the newer types of data sources such as Big Data, scanner data, citizen-generated data and web scraping.

The CSDA covers the following:

- How information is managed as an asset throughout its lifecycle.

- Accessibility of information.
- Describing data to enable reuse.
- Capturing and recording information at the point of creation/receipt.
- Using an authoritative source.
- Using agreed models and standards.
- Information security.

CSDA is focused on capabilities related to data and metadata. These include data input; data transformation; data integration and provisioning; metadata management; data governance; provenance and lineage; and security. CSDA is a “data-centric” view of an NSO’s architecture, putting emphasis on the value of data and metadata, the need to treat data as an asset and to focus on the way an NSO could treat their data and metadata.

The Modernisation Maturity Model

The companion to CSDA is the Modernisation Maturity Model (MMM). The MMM and its roadmap focus on how to build organizational capabilities through the implementation of the models and standards identified as key to statistical modernisation, such as the Generic Statistical Business Process Model (GSBPM), and its extension the Generic Activity Model for Statistical Organizations (GAMSO), the Generic Statistical Information Model (GSIM) and the Common Statistical Production Architecture (CSPA). See *Chapter 15.4 — Use of standards and generic models in an NSO*.

The MMM is a self-evaluation tool to assess the level of organizational maturity against a set of pre-defined criteria. There are multiple aspects of maturity in the context of modernisation, and there are several distinct dimensions. Within each of these dimensions, an NSO may have different levels of maturity. A maturity self-assessment should be carried out by a cross-cutting group involving members of the corporate planning, statistical production, information, methodology, applications and technology functions in order to ensure a comprehensive review.

The MMM allows statistical organizations to evaluate their current level of maturity against a standard framework. This assessment will provide a clear picture of the organizational maturity level, which can then be compared between organizations, and between statistical domains/business units within an organization.

The Modernisation Maturity Roadmap

The MMM has an associated roadmap, the Modernisation Maturity Roadmap (MMR), that provides clear guidelines on the steps to take to reach higher levels of organizational maturity more quickly and efficiently. The roadmap includes supporting instruments to help statistical organizations, at different maturity levels, to implement the different standards. The MMM and the roadmap can help an NSO regardless of its capacity level. They acknowledge that there can be different maturity levels depending on the statistical domain or part of the NSO. The roadmap addresses the needs of an NSO, particularly those in the earlier stages of modernisation, to have clearer information about how to progress in the most efficient way.

- *CSDA Reference Architecture* (.
- *Modernisation Maturity Model* () and the Roadmap for Implementing Modernstats Standards.


14.3.2 Managing statistical metadata

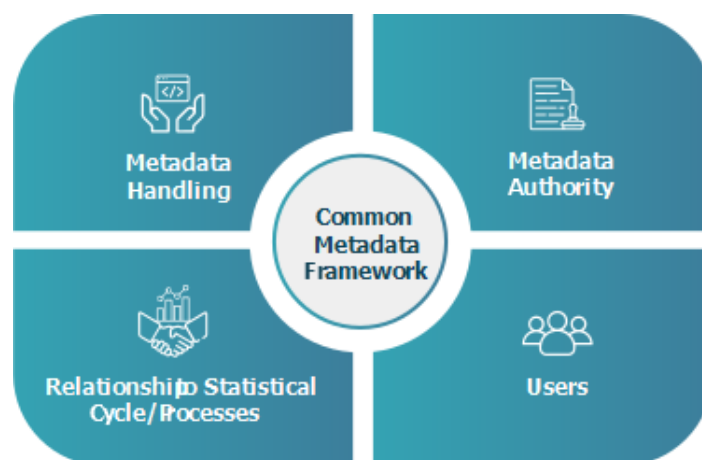
Principle 3 of the United Nations Fundamental Principles of Official Statistics (UNFPOS) states that to facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics. NSOs must be fully transparent for all users about their methods and provide comprehensive metadata linked to statistical data that is publicly accessible. Many NSOs have standard metadata templates that accompany each release of statistics.

An NSO must ensure that its users are properly informed regarding the location of data, how data were defined and compiled, the level of quality assigned to the data, and what related data can be used for comparison or to provide context. NSOs are obliged to describe accurately and openly the strengths and weaknesses of the data they publish and

to explain how much inference the data can support. Thus, metadata provide information to enable the user to make an informed decision about whether the data are fit for the required purpose. Metadata management is an overarching GSBPM process. Some of the metadata serve as internal guidance on statistical production, and some support the user of statistics.

Statistical metadata consists of data and other documentation that describe data in a formalised way, and describe the processes and tools used in the production and usage of statistical data. Metadata describe the collection, processing and dissemination of data as well as relating directly to data themselves. All published or released statistics should be accompanied by metadata. Many NSOs have implemented metadata-driven systems that use metadata as input to configure processes so that a common process can serve despite the differences between statistical domains (and organizations). Using metadata as input parameters in this way can create highly configurable process flows. Good metadata management is essential for the efficient operation of statistical business processes. Metadata are present in every phase of a statistical production process and should be captured as early as possible and stored in each phase. Metadata management systems are vital and have the goal to make it easier for a user or application to locate specific data. This requires designing a metadata repository, populating the repository and making it easy to locate and use information in the repository. A number of data and metadata management systems are available to an NSO (see [Chapter 11.4 — Metadata - providing information on the properties of statistical data](#)).

The *Common Metadata Framework* () is a repository of knowledge and good practices related to statistical metadata and identifies the following principles for metadata management:



- **Metadata handling**

- Manage metadata with a focus on the overall statistical business process model.
- Make metadata active to the greatest extent possible to ensure they are accurate and updated. Active metadata are metadata that drive other processes and actions.
- Reuse metadata where possible for statistical integration as well as efficiency.
- Preserve history of metadata by preserving old versions.

- **Metadata authority**


- Ensure the registration process and workflow associated with each metadata element is well documented, so there is clear identification of ownership, approval status, date of operation, and so forth.
- Single source: Ensure that a single, authoritative source exists for each metadata element.
- One entry/update: Minimize errors by entering once and updating in one place.
- Standards variations: Ensure that variations from standards are tightly managed, approved, documented, and visible.

- **Relationship to statistical cycle/processes**

- Make metadata-related work an integral part of business processes across the organization.
 - Ensure that metadata presented to end-users matches the metadata that drove the business process or was created during the process.
 - Describe metadata flow with the statistical and business processes, data flow, and business logic.
 - Capture metadata at their source, preferably automatically as a by-product of other processes.
 - Exchange metadata and use them for informing both computer-based processes and human interpretation. The infrastructure for the exchange of data and associated metadata should be based on loosely coupled components, with a choice of standard exchange languages, such as XML.
- **Users**
 - Ensure that users are clearly identified for all metadata processes and that all metadata captured will create value for them.
 - Metadata is diverse. Different views correspond to different uses of the data; users require different levels of detail; and, metadata appear in different formats depending on the processes and goals for which they are produced and used.
 - Ensure that metadata are readily available and useable in the context of the external or internal users' information needs.
 - Ensure feedback is collected from the knowledge management process and technology user.

14.3.3 The Statistical metadata system

Statistical metadata systems play a fundamental role in NSOs. Such systems comprise the people, processes and technology used to manage statistical metadata. Statistical metadata systems generate and manage metadata. It is an integral part of an NSO and is cross-cutting by nature and needs the involvement of managers, subject-matter statisticians, methodologists, information technology experts, researchers, respondents and end-users. Their needs and obligations vary according to whether they participate in the system as metadata users, metadata suppliers, designers, developers, producers, administrators and/or evaluators.

As noted in the UNECE publication *'Statistical Metadata in a Corporate Context: A guide for managers'* () the statistical metadata management system can effectively support the following functions:

- Planning, designing, implementing and evaluating statistical production processes.
- Managing, unifying and standardising workflows and processes, and making workflows more transparent by sharing work instructions.
- Documenting data collection, storage, evaluation and dissemination.
- Managing methodological activities, standardizing and documenting concept definitions and classifications.
- Managing communication with end-users of statistical outputs and gathering of user feedback.
- Improving the quality of statistical data, in particular consistency, and transparency of methodologies.
- Managing statistical data sources and cooperation with respondents.
- Improving discovery and exchange of data between the NSO and its users.
- Improving the integration of statistical information systems with other national information systems. Growing demands to use administrative data for statistical purposes require better integration and sharing of metadata between statistical and administrative bodies, to ensure coherence and consistency of exchanged information.
- Disseminating statistical information to end-users. End-users need reliable metadata for searching, navigation, and interpretation of data. Metadata should also be available to assist post-processing of statistical data.

- Improving integration between national and international organizations. International organizations are increasingly requiring integration of their own metadata with metadata of national statistical organizations in order to make statistical information more comparable and compatible, and to monitor the use of agreed standards.
- Developing a knowledge base on the processes of statistical information systems, to share knowledge among staff and to minimize the risks related to knowledge loss when staff leave or change functions.
- Improving the administration of statistical information systems, including administration of responsibilities, compliance with legislation, performance and user satisfaction.
- Facilitating the evaluation of costs and revenues for the statistical organization.
- Unifying statistical terminology as a vehicle for better communication and understanding between managers, designers, subject-matter statisticians, methodologists, respondents and users of statistical information systems.
- *The OECD data and metadata reporting and presentation handbook* (🔗).
- *Metadata Management - GSBPM and GAMS* (🔗).
- *UNECE Terminology on Statistical Metadata* (🔗).
- *Metadata Concepts, Standards, Model and Registries* (🔗).
- A basic framework for the role of the SMS in statistical organizations as defined in the UNFPOS (🔗).
- Core principles for metadata management, UNECE (🔗).
- *UNECE Common Metadata Framework – A guide for managers* (🔗).
- (Eurostat) publishes a large amount of metadata. This approach follows principle 15 of the European statistics code of practice (🔗), which states that data should always be made available with supporting metadata.
- *UNECE Statistical Metadata in a Corporate Context: A guide for managers* (🔗).
- Philippine Statistics Authority – Inventory of Statistical Standards in the Philippines (ISSiP).

14.4 Information and knowledge management systems

14.4.1 Information management

Statistical information is constantly sought after within an NSO and by its external users. With the ever-increasing amount of information, we receive in our daily lives, finding relevant information can be a time-consuming challenge. Information can be scattered across a multitude of disconnected systems, repositories, databases and data warehouses, and access to it can also be restricted. This often leads to the creation of ‘information silos’ that evolve, reflecting the way organizations are structured, in that different departments have different priorities, responsibilities and visions. This can lead to inefficiencies in communication and collaboration between business units, making it difficult to find and retrieve information. This in turn, contributes to duplication of information and inconsistencies, thereby reducing efficiency and productivity, as well as leading to frustration.

Fostering a culture of information sharing and the use of information management tools can improve the ways to find, retrieve and deliver information in order to facilitate better sharing of information and knowledge, thereby maximising the use of statistical data.

This can help in the identification of new trends and opportunities, improve the sharing of information with other agencies and help protect vital information. It brings together data, information, technology, information systems and business processes with the goal of providing information to users according to their needs.

14.4.2 Knowledge management

Knowledge management concerns how collective information, knowledge and expertise are used in order to be more effective as an organization. It covers the management of data flow, the sharing of “know-how”, knowledge retention, collective information, knowledge and expertise. The creation of knowledge systems can help filter out what content is important to a user, encourage collaboration across boundaries and break down organizational silos and base organizational actions on knowledge, sharing experiences and learning from them, so that the wheel is not constantly reinvented. With the increasing demand for transparency and accountability from governments, knowledge management is a key component of a strategy to efficiently manage knowledge in a multi-stakeholder environment like an NSO.

Knowledge management systems refer to any kind of IT system that stores and retrieves knowledge, improves collaboration, locates knowledge sources, mines repositories for hidden knowledge, captures and uses knowledge, or in some other way enhances the knowledge management process.

To succeed, a knowledge management system requires both tools and a culture of collaboration. Technology is obviously important to implement a knowledge system, but the focus should be on people and processes rather than IT systems and on creating a culture of collaboration and knowledge sharing. Knowledge management systems are software that specialise in the way information is collected, stored and accessed, combined with an open and collaborative approach.

Knowledge management requires the expertise to organize information and knowledge in a systematic way, and when properly applied, it can reduce the time needed to find information, prevents the loss of knowledge, and allow people to work faster in a more efficient and agile manner. Simply put, knowledge management can make an NSO more effective. An NSO can benefit from a knowledge management system using collective information and knowledge irrespective of its capacity level. Such a system can range from basic process documentation to more sophisticated systems using knowledge management software and work practices.

14.4.3 Information and knowledge management tools

There are a number of tools and best practices for implementing information and knowledge management systems. These include:



Content Management Systems manage the creation and modification of digital content. They typically support multiple users in a collaborative environment. They support the separation of content and presentation and are widely used for organizational content management and web content management. Main functions include web-based publishing, format management, history editing and version control, indexing, search, and retrieval.



Groupware is collaborative software which can be used to exchange knowledge and expertise. It is designed to help people involved in a common task to achieve their goals. Groupware refers to programs that help people work together collectively while located remotely from each other.



Search engines are information retrieval tools that connect knowledge seekers with experts and answers.



Knowledge portals are web-based applications that provide a single point of access to organizational knowledge, integrating knowledge repositories, expert directories, collaboration tools, and other knowledge-intensive applications.



Communities of practice are networks of individuals working with a common, shared purpose grouped together to facilitate knowledge building, idea creation and information exchange.



Enterprise social media focuses on the use of online social networks or social relations among people who share business interests and/or activities.



Mentoring and knowledge transfer is the practice of retaining and transferring knowledge within an organization by mentoring and late-career knowledge transfer to develop talent, skills and careers.



Post-mortems, lessons learnt is a systematic way of recording experiences lessons learnt that can be subsequently made findable throughout the organization with the use of taxonomies.



Transfer of best practices is a systematic way of recording best practices that can be subsequently made findable throughout the organization with the use of taxonomies.



Classification of content using taxonomies within a content management system. Taxonomy can refer both to the hierarchical structure into which content is authored as well as the metadata elements and vocabularies created for tagging content.

- Examples of Knowledge Management Software ()

14.5 Managing archives

Data archiving is the process of moving data that is no longer actively used to a separate storage device for long-term retention. Archive data is older data that remains important to the organization or must be retained for future reference or regulatory compliance reasons.

In GSBPM, archiving has been incorporated into the process of data and metadata management, to reflect the view that archiving can happen at any stage in the statistical production process, so archiving is not considered to be necessarily a final stage.

Many archives exist for the purpose of making data actively available, whereby data and metadata are systematically stored, protected, and made available according to agreed rules. Laws may require content to be kept for specific periods. Internal and external audits may require document retention. Archiving ensures the continued viability and usability of data now and in the future and provides access to these data within the framework of the national legislation by ensuring confidentiality and protecting privacy.

Microdata sets contain information on individual persons, households or business entities collected through a census, survey, interview or administrative recording systems. NSOs and other producers of official statistics generate vast amounts of microdata from surveys and censuses, and these data represent a significant investment by an NSO and have considerable value for both existing and future users for the production of national statistics and research. An NSO should therefore have a policy and the mechanisms in place to ensure that microdata and metadata can be archived.

When archiving data, the need to keep information online should be balanced with the fact that keeping too much old information available online consumes valuable storage which could be better used for newer information. It can also increase the number of irrelevant search results returned and adds to the effort required to maintain, migrate, and re-classify content. Long-term storage of this material adds to storage costs and to the security burden of safeguarding the confidentiality of the records.

Archives should be indexed and searchable so that files can be easily located and retrieved. A microdata catalogue and a document management system can help to manage the creation, capture, indexing, storage, retrieval, and disposition of the records and information assets of an NSO. Records management addresses the issues of knowing what data the NSO has, where it is stored, how long it should be kept and how secure it is.

Archiving rules for specific statistical processes depend on the general statistical legislation and any archiving policy of the NSO. These rules include consideration of the medium, location of the archive, and the requirement for keeping duplicate copies. They should also consider the conditions (if any) under which data and metadata should be disposed of.

The Generic Longitudinal Business Process Model (**GLBPM**) has been derived from the GSBPM and provides a generic model that can serve as the basis for informing discussions across organizations conducting longitudinal data collections, and other data collections repeated across time. The model is intended to serve as a reference model against which implemented processes are mapped, for the purposes of determining where they may be similar to or different from other processes in other organizations. It may also prove useful to those designing new longitudinal studies, providing reminders of steps which may need to be planned.

GLBPM covers the following phases:

- Evaluating and specifying needs;
- Design and re-design of data collection instruments;
- Build and re-build of data collection instruments;
- Data collection;
- Processing and analysis;
- Archive/preserve/curate;
- Dissemination and discovery;
- Research and publish;

- Retrospective evaluation.
- The NADA Microdata Cataloguing Tool ([🔗](#)).
- *UK Office of National Statistics Code of Practice Protocol on Data Management, Documentation and Preservation* ([🔗](#)).
- *ICPSR Guide to Social Science Data Preparation and Archiving* ([🔗](#)).
- Philippine Statistics Authority – PSA Data Archive ([🔗](#)).
- Philippine Statistics Authority Data Archive ([PSADA](#)).

14.6 Policy on retention of data and related information

A data retention policy is part of an NSO's overall data management strategy. A comprehensive data retention policy outlines the business reasons for retaining specific data as well as what to do with it when targeted for disposal.

A data retention policy, or records retention policy, is an organization's established protocol for retaining information for operational or regulatory compliance needs.

A data retention policy should treat archived data differently from backup data. Archived data is no longer actively used by the NSO but still needed for long-term retention. An NSO may need data shifted to archives for future reference or for compliance. An NSO's backup data can help it recover in the event of data loss. A backup policy is important to make sure the NSO has the right data and that the right amount of data is backed up. Too little data backed up means that any recovery needed after a data loss will not be as comprehensive as needed, while too much can be difficult to manage. Achieving a balance between these conflicting requirements is the objective underlying a data retention policy.

The policy should cover all technologies that are used to obtain data and cover a variety of formats such as paper forms, Computer Assisted Telephone Interview (CATI) records, Computer Assisted Interview (CAI) records, electronic administrative data, data streams, scanned images and faxes.

The policy should cover legislative responsibilities for the data it collects, publishing what is collected, and doing so in a manner that will not enable identification. For proper creation and implementation of a data retention policy, especially regarding compliance, the IT team will need to work with the legal team. The legal team will know how long data must be retained by law while IT will carry out the actual implementation of the policy.

Defining a retention schedule depends on the type of data, the data collection cycle and will be according to the specific needs and policy of an NSO, and the legislation of each country. Each data collection should be retained according to its merits and can be destroyed after all processing, and likely provider queries have been resolved. Some NSOs have adapted their backup software archiving functionality to automate data disposal. In some countries, a central government agency has responsibility for the storage and archiving of important documents and files.

- IHSN Archiving ([🔗](#)).
- Australian Bureau of Statistics Policy on Retention of Collection Forms and Administrative Material ([🔗](#)).

14.7 Skills needed

Specific skills needed to implement data, information and knowledge management systems include the following:

- Content management systems, groupware and social media.
- Database technology and decision support systems.
- Subject matter experts for constructing taxonomies.

For more information see *Chapter 13 - Human Resources Management and Development*.