BC Geographic Warehouse

A Guide for Data Custodians & Data Managers

Last updated November, 2013

DataBC

A Guide for
Data Custodians & Data Managers

TABLE OF CONTENTS

INTRODUCTION

PURPOSE

This is a guide for Data Custodians and Data Managers. It explains why these roles are important, and their associated responsibilities. It also describes how business areas can use the BC Geographic Warehouse (BCGW) to make their data holdings accessible to clients.

AUDIENCE

If you are a Data Custodian or Data Manager involved with managing spatial data holdings that are widely used and vital to Ministry business, then this guide is for you.

CONTENTS

The guide covers these topics:
- Getting the most out of the data we collect and manage
- The role and obligations of Data Custodians and Data Managers
- Using the BCGW to distribute data
- Where to find more information.

IT'S ALL ABOUT INFORMATION

This guide is about managing information within the BCGW and making it available for decision making.   The following points provide some context.

1. **We use a lot of information -** The Province collects and uses a wide variety of information. Much of this information is collected by business areas that have a responsibility for inventory and providing access to this information.

2. **Information is used for making decisions -** It's the reason we collect and provide it. The wider its access, the greater its potential.

3. **Information integration adds value -** The power of information for decision-making increases significantly when it is integrated with other related data. Base mapping and other foundation data play a crucial part in providing a common geographic context for analysis and decision making.

4. **Information standards are important -** Standards provide the basis for information processing, integration and access using common sets of tools. Lack of adherence to standards severely limits data use.

5. **Information is costly -** The Province spends considerable money, time and resources to collect information and manage it.

6. **Information needs to be understood -** Information used properly can provide immense value. However, it can have the opposite effect if used inappropriately. Users need information about the data they intend to use in order to assure themselves of the data's fitness for purpose.

7. **Information needs to be managed -** Data are not static. Over time, they are added to, updated, deleted, and reformatted, and so on. Keeping information in usable form is not a trivial task. It requires active management.

8. **Information needs to be protected -** In many respects, information is a fragile resource. It's easily corrupted and misused, and loses value if it goes out of date or is not properly described. Some information is confidential, and needs to remain so. Someone has to set the rules for using information and, having done so, step up to making sure the information is made available according to those rules.

WHAT IT MEANS TO BE A DATA CUSTODIAN

WHAT IS A DATA CUSTODIAN?

A Data Custodian is the person who

- ◆ protects and promotes the use of data holdings under his or her care;
- ◆ sets policies, and is accountable for defining the appropriate use of the data;
- ◆ is provider of the authoritative version of the data; and
- ◆ is ultimately accountable for issues related to definition, collection, management and authorized use of the data.

Every data holding should have one and only one Data Custodian - usually someone at the Director level.

AM I A DATA CUSTODIAN?

> *If you are a director with responsibility for the business (i.e., if you have authority based on legislation or policy), then you are probably also the Data Custodian.*

Unless this custodianship role is formerly assigned to someone else, you have the obligations and accountability described here.

OBLIGATIONS OF A DATA CUSTODIAN

If you are a Data Custodian you have these obligations:

1. You must manage the data as a valuable business asset through its entire lifecycle.

2. You must ensure the data is fit for the purposes for which it is collected and provided.

3. You must stay in touch with the users of your data and respond to their needs. This must occur at two levels:

    a. Understanding the business environments of users and the information requirements of the business; and

    b. Understanding the individual needs of users for access to information.

RESPONSIBILITIES OF A DATA CUSTODIAN

If you are a Data Custodian, you have these responsibilities:

1. **Data planning** - You decide what data is collected, when and how, its definition and purpose, what policies and standards apply, and what resources are utilized to manage and maintain it.

2. **Data content** - You decide user access rules, how the data is stored, promoted, distributed and retired. (Users for their part must respect the applicable-use policies set down by the custodian.)

3. **User advocacy** - You decide how to respond to users' business needs. You must understand the needs of the entire community of users. Of course, this doesn't mean you have to meet all user requirements. Rather, you should try to maximize the value of your data to users as a whole, balancing their respective needs.

The Data Lifecycle
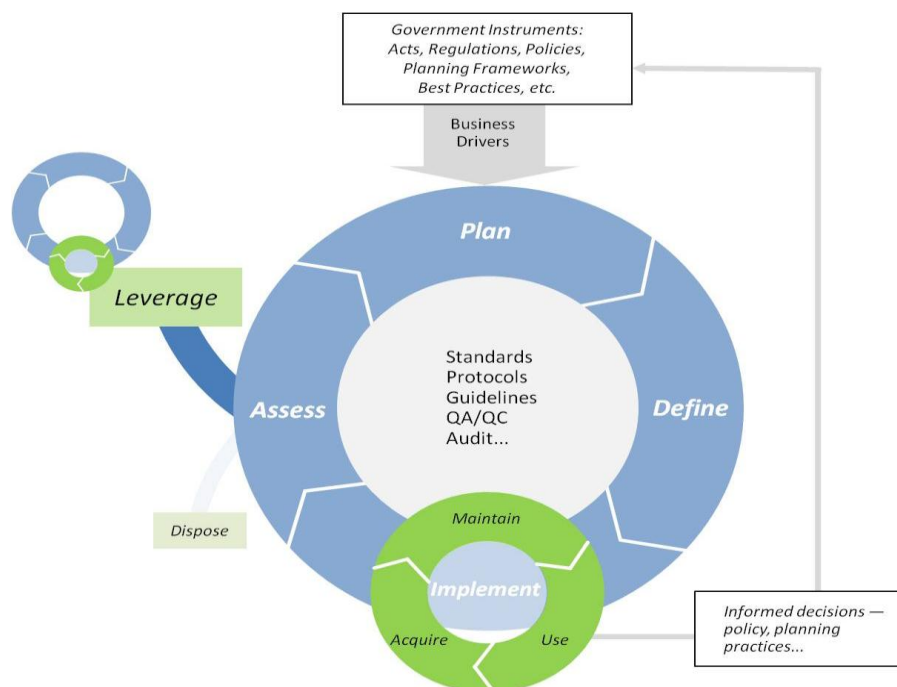
RELATED ROLES

THE DATA MANAGER

While the Data Custodian concentrates on issues of governance, the day-to-day management of data is the responsibility of the Data Manager. In effect, the Data Manager is the person who carries out the policies and directions set by the Data Custodian. There could be several Data Managers for a data holding, and the role can be delegated.

This is discussed further in the next section.

THE DATA STEWARD

Sometimes the Data Custodian may not have the resource available to fulfill his or her responsibilities, or it may be more efficient to conduct activities in concert with another group. In these cases, a Data Steward may become involved.

These are the characteristics of a Data Steward:

- Usually at the director level, like the Data Custodian
- Ability and resources to provide a set of services needed by the Data Custodian

Don't confuse the Data Steward with the Data Manager. The steward has an agreement with the Custodian to provide a specific set of custodial duties on behalf of the custodian.

The data Custodian and Data Steward should sign a written *Data Stewardship Agreement* that lays out the specific responsibilities that are being delegated to the steward by the custodian

Note that the involvement of a Data Steward doesn't change the obligations of the Data Custodian. The Custodian is still accountable for the governance and use of the data.

OTHER ROLES

- A Data Administrator is responsible for setting policies and standards related to managing and protecting data.
- A Discipline Authority is a business expert who understands the business relevance of data standards, and supports their application and use to meet an organization's needs.

WHAT IT MEANS TO BE A DATA MANAGER

WHAT IS A DATA MANAGER?

- A person appointed by a Data Custodian to manage a specific data set according to policies, plans and standards defined by the custodian.
- May take direction from a Data Steward designated by the custodian.
- Responsible for day-to-day management of the data. May coordinate operations staff.
- Usually has technical knowledge of the data, its collection & storage, and ways in which the data is commonly used.

AM I A DATA MANAGER?

> *You are a Data Manager if you have been directed by a Data Custodian to have the day-to-day responsibility for managing a data set.*

OBLIGATIONS AND RESPONSIBILITIES OF A DATA MANAGER

The main obligation of a Data Manager is to ensure that the Data Custodian's directions are properly fulfilled.

You will typically be responsible for these activities (though perhaps not all of them):

1. Data capture and storage
2. Data quality assurance
3. Metadata capture and maintenance
4. Define business rule for security and access to data
5. Data population and distribution
6. User support

Data population and distribution may involve the BCGW, and this is discussed further in the next section.

THE BC GEOGRAPHIC WAREHOUSE (BCGW)
The BCGW is a structured repository of priority data that's needed by Government, industry and the public in order to make decisions.

WHY THE BCGW

The BC Geographic Warehouse (BCGW) is managed by DataBC.   It supports the Province's electronic services strategy as it is designed and operated for the purposes of hosting and cataloguing sets of data (mostly spatial data) on behalf of Custodial ministries and other government Custodial agencies. The BCGW's access tools include ArcGIS for direct GIS access to spatial data, iMapBC and other web-based mapping tools, Discovery and Distribution Services for locating and downloading data.

The main purpose of the BCGW is to make data available to users in a form that best supports data integration, sharing and exchange, analysis and decision-making. (In general, this is the purpose of any data warehouse.)

Specifically, the BCGW accomplishes the following:

1. **Wide access** - In principle, any authorized person can gain access from anywhere on the internet, and browse, view, order and download data.
2. **Suited for purpose** - By means of data transformation, the BCGW allows data to be formatted and presented in a form useful to different audiences.
3. **Integration** - It allows data to be viewed, combined and analyzed with other data.
4. **User empowerment** - It includes some simple but effective data viewing and analysis tools.

By virtue of these features, the BCGW can provide substantial benefits for both Data Custodians and Data Managers. In particular it responds to Data Custodians' obligations to provide wide access to current, complete and authoritative data. It does this in a controlled manner, allowing graduated access from generalized browsing and viewing through to in-depth content access.

HOW DOES THE BCGW WORK?

The BCGW is somewhat similar to a distribution warehouse that you find in a manufacturing supply chain.

The BCGW's data publishing and distribution process works as follows:

1. Government ministries and agencies (the data providers) produce data, which they publish to the BCGW. (This is like component manufacturers creating goods and shipping them to a warehouse.)

2. At the warehouse, the data are

   ◆ repackaged for consumer use - i.e., for the broad audience that uses the BCGW;

   ◆ checked for valid format and structure; and then

   ◆ stored and made available for browsing and viewing via an online catalogue.

3. Data consumers browse and select data products from the warehouse, which are distributed to them, generally using download via the internet.

Information Management Through the BCGW
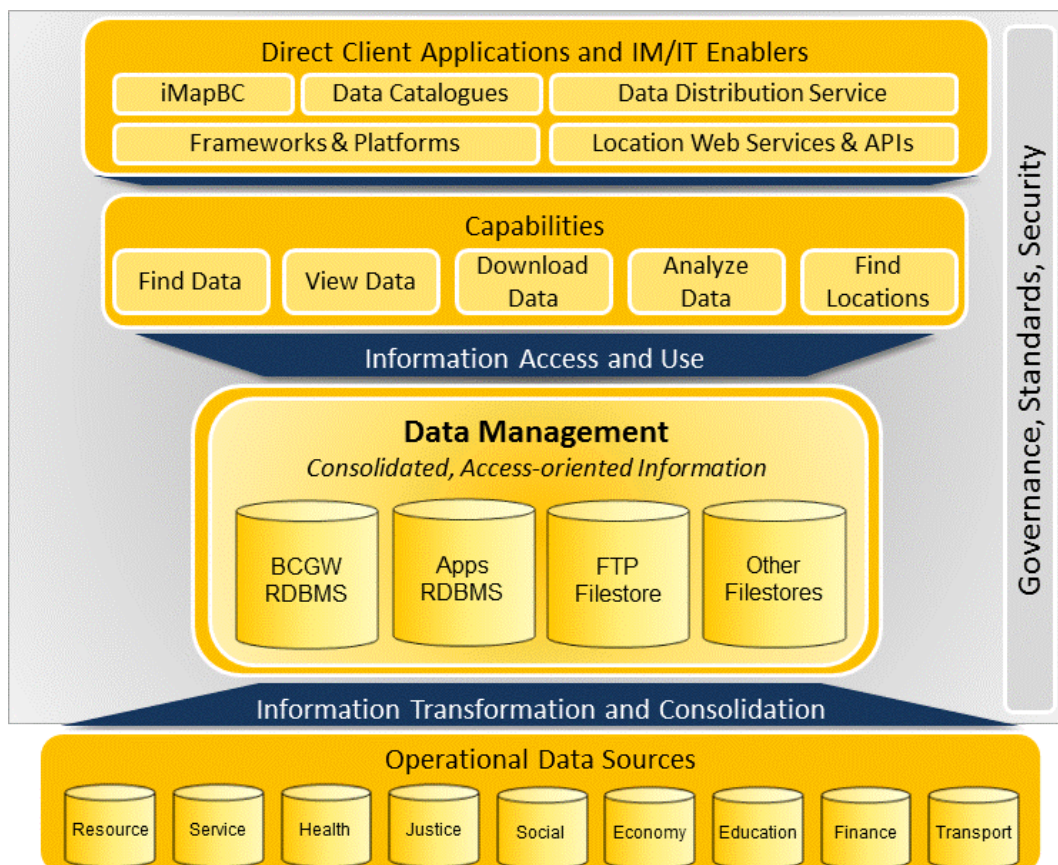
MAKING USE OF THE BCGW

BENEFITS

The BCGW provides the opportunity and means for Data Custodians to present their data to a wide audience of users.
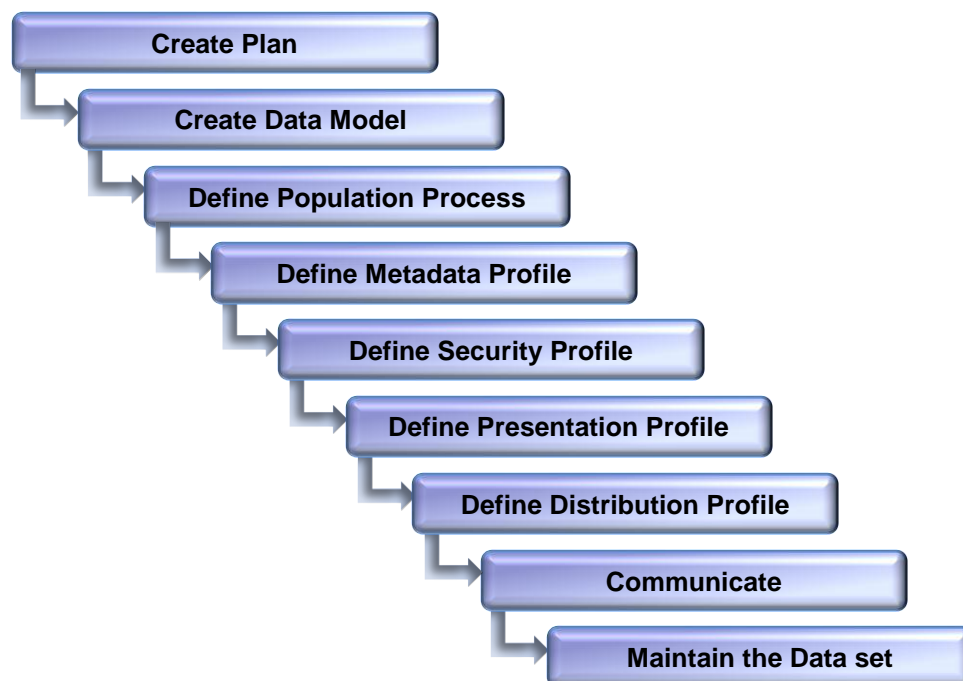
In this respect, the BCGW is enabling technology. It provides two sets of services - one set to assist Data Custodians to present their data, the other set to help user access and use the data.

These are the benefits:

- ◆ relieves Data Custodians and managers of many of the costs and headaches of making their data widely accessible
- ◆ user access is easier to manage by virtue of a standard security model
- ◆ ability to structure, integrate and package data to better suit user needs
- ◆ improves the user experience by offering a standard user interface and set of familiar tools
- ◆ allows easier viewing and integration with other BCGW data

STEPS FOR PLACING DATA IN THE WAREHOUSE

The effort required to utilize the BCGW depends on the nature and complexity of your data set. The steps described here can be "right-sized" to fit your needs.

Create Plan

Create Data Model

Define Population Process

Define Metadata Profile

Define Security Profile

Define Presentation Profile

Define Distribution Profile

Communicate

Maintain the Data set

The Data Manager is usually responsible for preparing data for population and distribution using the BCGW. The process is fairly straightforward, but it does involve a number of steps.

DataBC, Strategic Initiatives Division, who operates the BCGW, has staff available to work with Data Managers to make the process as streamlined as possible.

If you are a Data Manager, these are steps that you will undertake in order to place your data in the BCGW.

START WITH THE END IN MIND

As a Data Manager, it's worth remembering that the point of this process is to end up not just with your data in the BCGW, but with it packaged and presented in a format that makes it easy for your clients to use it for planning, analysis and decision-making.

Thus, although the steps presented here are in a BGW population sequence, in many ways it's better to think first about the final steps of the process (i.e., how will I present my data and what data products will I provide?), and then work backwards. This will then inform your security, metadata and other modeling requirements.
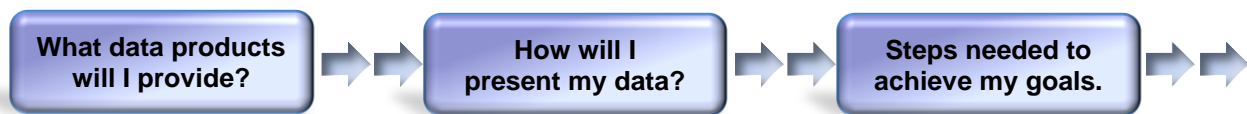
CREATE A PLAN

Why is this Required?

Depending on the size and complexity of your data, the steps required to prepare the data and make it accessible using the BCGW may take days or weeks, and will require participation of your staff. Therefore, the first step is to create a work plan and ensure you have the required time and resources.

Activities

1. Contact DataBC, who will help you get started and provide advice about resourcing. In particular, you will probably need a Business Analyst to facilitate the process.

2. Your Business Analyst will set up a whiteboard session that will be attended by you, technical staff, and any vendors who may be involved in developing and maintaining your operational system. The purpose of the whiteboard session will be to develop a plan that defines all the resource required, role and responsibilities, issue and timelines.

3. This plan must be approved by you before work proceeds.

4. Your Business Analyst will then implement and manage the plan to completion.

5. Based on need, a Stewardship agreement may be developed.

6. The process to establish the licensing requirements for the data is also initiated at this point using the Open Data Assessment Checklist.

CREATE THE WAREHOUSE DATA MODEL

Why is this Required?

When your data is stored in the BCGW, it may have to be structured quite differently from how it's organized in your operational system. This is because it will be used for potentially different purposes, serving a different audience who will use the BCGW's tool set for browsing and accessing the data. Unless the data conforms to certain design constraints, the BCGW's tools and services won't work properly.

A data model is a formal description of the structure of a data set - that is, the types of data, their properties and relationships. A data model has to be created to describe how your data will be stored in the BCGW.

The data model itself will be accessible to BCGW users. This is because it's an important element of metadata (see Define a Metadata Profile) that can help users utilize the data in their own systems environments.

Activities

As Data Manager, you are responsible for having your data modeled according to DataBC's standards and stored in the DataBC modeling environment. This requires the following:

1. Create a logical data model that describes the structure of the data as it will be stored in the BCGW.

2. Once this has been reviewed and approved, the physical data model must be created.

3. Create and provide the resulting scripts that will be used to build the tables, layers and views in the warehouse.

DEFINE THE POPULATION PROCESS

Why is this Required?

Population refers to the process of retrieving data from the Data Custodian's source system, transforming it as required and storing in the BCGW.

A number of different methods already exist, so it's a matter of selecting the most appropriate one and then configuring it to the specific needs of your data. Although this is the responsibility of the business area (who must also provide the funding), DataBC is available to you for consultation and support.

Activities

1. Initiate and fund a small project to develop the transformation process that will move the operational data into the warehouse.

2. Test the transformation process with a sample set of data to confirm that the warehouse model is correct. This is an important task. If problems are encountered then the data model will have to be fixed.

3. The next step is to test the transformation process with a full set of data to confirm that the process itself works correctly. Again, if there are problems, then the data transformation process will have to be corrected.

So by now, we have the data correctly represented in the warehouse and the process in place for populating the data from the operational system. But the data is not yet ready for access by BCGW users.

DEFINE A METADATA PROFILE

Why is this Required?

Metadata is information that describe data. A metadata profile is a collection of information that describes a wide variety of data characteristics, including such topics as geographic extent of the data, when the data was collected, who collected it, who owns it, its format, quality, version number, rights of use, how often it's updated, its intended usage, and so on.

Creating a metadata profile is important for the following reasons:

♦ It provides the context for the data and allows it to be described and managed in a consistent way, similar to other datasets.

♦ It supports user self-service. Users can browse the metadata profile to understand the purpose and characteristics of a data set and decide whether it's right for their needs.

♦ It supports sustainment in that it embodies knowledge about the data that might otherwise remain in a person's head, and not be captured.

There are international standards for metadata that DataBC follows. This provides interoperability with other data service utilities, meaning that descriptions of your data can be widely published and accessed (assuming your security profile allows this).

Visit the DataBC site for more information about DataBC's [metadata management tool](#).

Activities

1. You or your Business Analyst will use DataBC's metadata management tool to enter the metadata that describes your data. You may need to get the necessary permissions set up with DataBC. Also, if necessary, DataBC will provide training on using its metadata management tool.

2. This will then need to be reviewed and approved by DataBC. (As Data Manager, you will have authority to edit your metadata profile in order to keep it up to date.)

DEFINE A SECURITY PROFILE

Why is this Required?

A security profile defines who may access your data and what they can use it for. The security profile (which is part of the metadata) is very important because it ensures that user access is consistent with your policies for data security, confidentiality and appropriate use.

Activities

1. Define a security profile by defining access rules and restrictions that cover browsing, viewing, ordering and downloading your data via the BCGW. This is done using the metadata management tool.

2. Work with DataBC staff to ensure that your security profile is complete, addressing all of your data and all types of user access.

3. Create any required data use agreements. (Note that it is your responsibility to ensure compliance with Freedom of Information and Protection of Privacy policies.)

DEFINE A PRESENTATION PROFILE

This step usually requires substantial input from the business area. You can expedite the process by identifying who will make decisions about presentation and then arranging one or two focused meetings to reach agreement.

Why is this Required?

The presentation profile is very important. Its purpose is to present data in a consistent manner that best supports users' needs for analysis and decision-making. In effect it represents a business view of the data from the users' perspective, rather than the operational view of the source system.

Ideally, the presentation profile should make invisible the complexities and technicalities of the underlying physical data structures. Typically, it includes elements such as these:

- ◆ user-friendly names for data layers, tables and columns, e.g.,
  - o VEG_VEGETATIVE_COVER_POLYGON becomes Vegetative Land Cover,
  - o PRI_UTIL_LEVEL_CD becomes Primary Utilization Code
- ◆ translation of codes into understandable descriptions
  - o e.g., {1,2,3} becomes {high, medium, low}
- ◆ colours, symbols and fonts to distinguish different data types
  - o e.g., blue for rivers
- ◆ grouping of records according to logical category

Activities

1. DataBC will contact you to discuss how you want your data named and presented.

    a. You may also provide an ArcGIS layer file. If you do this, DataBC will review and approve the file

b. Alternatively, you may ask DataBC to develop a layer file on your behalf. In this case, DataBC will develop a first version according to your instructions and then send to you for review and approval.

2. Once the layer file is defined, DataBC will translate it for use with in various warehouse access channels for use with web mapping tools such as iMapBC. You will then be asked to review and approve the translation within a test environment.

Visit the DataBC site for more information about the various tools available to view and analyze and connect to the data.

DEFINE DOWNLOADABLE DATA PRODUCTS

Why is this Required?

Data products are predefined packages of data that BCGW users can browse and download via the BCGW's data catalogue.

Data products are important because they specify how your data can be used to support analysis and decision-making. You can define them based on types of use of the data (for instance, it may help users if you apply some pre-processing to the data to save them having to do it), or based on how your data is capable of being accessed (e.g., your data model may not support fine-grained access to the data).

Activities

1. Through discussion with users, develop an understanding of how they would like data assembled or "pre-packaged".

2. Within the technical constraints of your data and the BCGW, define a set of data products. DataBC can help in defining the data products and translating the definitions into technical specifications.

3. It's important that you reflect the security requirements associated with data products explicitly in the security profile because this will govern who may access the data products.

Visit the DataBC site for more information about the downloading data via distribution services.

COMMUNICATE

Continued engagement with your clients is important. This applies both at the levels of their business and their individual needs.

Why is this Required?

Once you have completed the steps described above, your data will be ready for use. You now need to let your user audience know that your data is available for access.

Activities

1. Decide on a communication strategy. Who do you have to reach? What are the messages that you want to convey? Users will want to know about the characteristics of the data and data products, what they can be used for, if there are restrictions on who may use them.

2. Launch the strategy. Initial promotion may involve several channels and extend over a period of months.

3. Sustain the strategy. Communication must continue throughout the lifetime of the data. You will need to continue to promote your data, to inform users of changes to the data, and to solicit their ideas and feedback.

MAINTAIN THE DATA

If there are data structure changes needed to your tables, layers or views in the warehouse, DataBC will work with you to modify the models and transformation process to ensure data continues to be provided as expected to BCGW users.

Why is this Required?

Once you have the data available for access via the BCGW, you must keep the pipeline flowing smoothly. You will enter into the maintenance phase, which involves less day-to-day effort, but is nevertheless important.

Over time the nature and application of a data holding may change. It's up to you as Data Manager to periodically update the models and profiles described above. You must also decide when obsolete data should be retired and removed from the BCGW.

Activities

1. Establish procedures for maintaining the data. These must address items such as
    a. who is responsible for day-to-day operations
    b. how updates to the data are handled
    c. how quality assurance is conducted (quality assurance is not DataBC's responsibility)
    d. how changes to profiles and models are handled
    e. how records management is done

2. Provide user support. This includes assigning resources to answer users' business-related questions and resolve issues (technical support will be handled through DataBC and Shared Services)

MORE INFORMATION

CONTACT DATABC

This guide explains the most important aspects of being a Data Custodian or Data Manager, and how the BCGW can be part of your approach to fulfilling the responsibilities of these roles.

If you have questions about this guide, or would like to talk more about using the BCGW, please contact DataBC at Data@gov.bc.ca