Office of the Chief
Information Officer

BRITISH
COLUMBIA

DATA ADMINISTRATION STANDARDS

Architecture, Standards and Planning Branch
Office of the CIO ● Province of BC
People ● Collaboration ● Innovation

Version 1.0

February 12, 1998

Data Administration Forum

A subcommittee of AC/IM

TABLE OF CONTENTS

1. INTRODUCTION

This set of standards arises from two of the original Data Administration Framework Task Force (DAFTF) recommendations, developed in fall 1996 under the auspices of the British Columbia government's Chief Information Officer. Recommendation #4 stated, in part:

Put in place the organizational structures needed for effective government-wide data management, specifically:

b) Ensure that all data within government fall under the mandate of a properly staffed Data Administration function

Recommendation #7 stated, in part:

Design and implement a data management education, training, development and support program

The DAFTF, a limited time-frame group with the delivery of a report as their specific purpose, converted into the Data Administration Forum (DAF) in early 1997. DAFTF members formed DAF to ensure ongoing improvement of data administration practices across government. DAF members consist of Data Administrators from government ministries who are cooperating to bring consistent changes in how government staff manage information.

Within DAF, there is a subcommittee that develops standards relating to information management. At this writing, the Standards sub-committee membership includes:

Andrew Faulkner, MELP	Rob Gretchen, Agric/Fish/Foods
Elaine Dawson, MELP	Jeremey Janzen, Forests
Karen Lee, ITSD	Krishan Saxena, Health
Mike Zimmer, Health	Rick Robertson, Trans & Highways

There are many "standards" available and under development. The Data Administration Forum means to continue a collective, cooperative, and ongoing effort, with improvements to standards being published as they are available. Since paper documents go out of date so quickly in such an environment, we have included web site addresses where appropriate, that either list current government-wide standards or show how single ministries have implemented specific standards.

The Data Administration Forum government-intranet web site address is

http://gww.acim.gov.bc.ca/   (then click on "DAF")

A current copy of this document is available from the above site. It is updated as needed by the Data Administration Forum. A BCGOV1 userid/password is required to access this site.

01.1.Purpose

The specific beginning-level standards outlined in this document were developed by the Standards sub-committee of DAF, and are meant to assist any BC government ministry or agency that is starting a Data Administration function. DAF expects that ministries with existing Data Administration functions will already have standards in place that, at minimum, meet those outlined here. DAF publishes these standards with the consent of the Advisory Council on Information Management (AC/IM).

These standards set minimum criteria for data management that apply to data within government, and are for use by program ministries as they see fit. The intent is to encourage all ministries to follow best practices in information management, while making a minimum set of standards mandatory. For example, this document may specify that each ministry should have "data element naming standards" and describe in general what they are; however, it may not actually specify detailed standards, because some ministries may not have the infrastructure developed that could action those standards *at that level of detail*. So, we have left it open for ministries to develop to their capability, so that at least **some** naming standards are followed. The idea is that a little consistency is better than none, even if "a little" isn't perfect.

11.2.Intended Audience

Data Administrators in each ministry play a key role as a negotiator, facilitator, and validator within a successful information management environment. Therefore, this document's primary intended audience is Data Administrators in each ministry, or those setting up or overseeing the Data Administration function. However, it can also serve as a reference to all internal (government) and external (contracted) resources involved in the management of data or the development and/or maintenance of systems which act on government data.

21.3.Scope

The scope of the standards outlined in this document apply to all ministries. DAF's intentions are to advertise and promote these standards as being mandatory in all government ministries, however, DAF does not have the resources to enforce standards compliance. To begin to effectively manage information, **a ministry should implement these standards as a minimum**, through a ministry data administration function.

The focus of this document is to add value to business data through careful and intelligent application of reasonable standards (i.e., not defining standards for standards' sake). Ministries that adopt these standards will facilitate the government-wide management and sharing of information.

31.4.Standards Creation

One of DAF's objectives is to develop modelling standards and techniques at a level suitable for cross government adoption. The Data Administration Forum provides ministry data administrators with an opportunity to present concerns and initiatives to the larger community of government data administrators. The members of the forum are committed to the incorporation of information management standards as a minimum baseline in the execution of the DA function in their ministry.

2. DATA ADMINISTRATION VISION

Data administration within government is guided by the corporate government principles surrounding information management and data administration that were developed by the Data Administration Framework Task Force and recorded initially in the DAFTF Report.

42.1.Principles

(The full text of these principles is included in Appendix A of this document.)

- Data and information are the essential linkages between an organization and its systems.
- Data is a corporate resource that needs to be managed as an asset.
- Data needs to be managed across its entire life cycle.
- Data can be, and where possible should be, shared.
- The whole of an organization has primacy over the parts.
- Modeling of data requirements yields sound and stable data structures.

52.2.The Context of Standards

Business areas can use the standards that Data Administration promotes to:

Improve Sharing of Information: Standards enable sharing indirectly through acceptance of modelling as a common practice and directly through development of common data structures;

Leverage Corporate Resources: Standards enable reuse and multiple application of corporate data resources;

Realize Cost Savings: Introduction of standards should generate real cost savings to ministries, through reduced duplication of effort, reduced translation costs, and broader applicability of data resources.

The standards identified in this document have been specifically focused on the baseline needs of government. In the past, most systems in government were built without following modelling standards. Modelling standards may need to be applied or implemented differently depending on where they are used:

Data Planning: In order to reinforce the linkage between business planning and information management, standards are needed that enable the early identification of data integration and sharing needs and opportunities.

New Systems and Data: Standards are needed that ensure optimal integration and sharing of data by new systems.

Existing Systems and Data: Standards are needed for consistent cataloguing and description of existing systems and data, so that a consistent and comprehensive understanding of all government data resources is available to everyone.

In order for standards to become accepted within an organization, they must be perceived as pragmatic and useful. Without organizational acceptance, it is impossible for any standard to be successful. For these reasons, the following categories of standards have been specifically selected to provide the greatest benefit for government:

Modeling Standards: management of the metadata and techniques for representing data and its relationships.

(more to come later)

3. MODELING STANDARDS

63.1.Introduction

There are no provincial government standards for modeling at this time. However, modeling should be undertaken for all systems development projects to ensure that the activities the business performs are well understood.

DAF recommends that data and process models be done for system development projects to ensure that the activities the business performs and the information used by those activities are adequately documented. **Remember!** *Modeling standards are directly related to the methodology used* — this creates some limitations. This factor needs to be kept in mind when reading Section 3.

73.2.Data Models

A data model is a graphical (Entity-Relationship Diagram or ERD) and textual (data dictionary) representation of the business data deemed of interest to an organization. It is a representation of data objects that can be shared and reused across application systems, organizational boundaries, and different functional areas. Models can be effectively used as:

- a source of information about the stated interests of an enterprise;
- a communications tool to facilitate improvements in strategies, tactics, and operations;
- a basis for information systems database design;
- a basis for accuracy and integrity of information;
- a communication tool to provide the understanding of data that leads to the identification of data sharing possibilities so that redundant data entry and unintentional replication of data can be reduced.

The model itself has a requirement for metadata to assist the Data Administrator in the management of the modeling environment -- all data model metadata is to minimally include:

Model Name: a full name can be provided within the modeling tool however the directory structure that contains it must conform to the operating system environment in which it is being used.

Model Title: a short meaningful business oriented phrase to identify the model.

Description: a description of the area represented by the model.

Business Area: identification of the business area being modeled.

Organization: the name of the primary Ministry or Government organizational structure for which the model was created.

Model Creator: name of the person responsible for creation of the model.

Model Creation Date: date the model was approved by the business area and the Data Administrator.

Validating Data Administrator: name of the Data Administrator that validated the model.

Last Modified Date: date the model information was last updated.

The basic components of a data model are:
- Entity type
- Attribute type
- Relationship Type

ENTITY TYPE

An object about which the business wishes to collect information; a person, place, thing, event, or concept of importance to the enterprise that is singular, exclusive, and uniquely identifiable. The minimum acceptable metadata for an entity is:

Entity Name: always a noun, named according to its definition and representing a single concept. Names must be:
- meaningful;
- self-documenting (by looking at the name, the reader should have a good idea what the name means without having to read the description);
- always singular (denoting the single pattern for all instances);
- derived from the business use or purpose; and
- repeatable (different people from different areas of the organization reading the name at different times should have the same understanding of what the name means -- a corporate-wide use of the name).

Unique Identifier: one or more attributes that may be used to uniquely identify an instance of an entity type.

Entity Definition: describes what it is rather than how, where, and when it is used. Entity type definitions should clearly describe what business information they record, using:
- precision (they resolve ambiguities and qualify imprecise terms);
- completeness (all terms are defined);
- clarity (plain English, few if any buzzwords);
- brevity (brief and to the point);
- examples (relevant to the business);
- compatibility (with other definitions).

ATTRIBUTE TYPE

A property or characteristic of an entity type or entity class. For example, color, weight, and gender are attributes that could describe an entity type. One or more of these attributes are used to uniquely identify an instance of an entity type. There are two types of attributes: key and non-key. The minimum acceptable metadata for an attribute are:

Attribute Name: (see comments under Entity Name)

Attribute Definition: (see comments under Entity Definition)

Data Type: defines the composition of the kind of fact represented by this Attribute Type, e.g., date, time, integer, decimal, alphanumeric string, binary array, etc.

Data Length: defines how many character positions are allowed for the kind of fact represented by an Attribute Type. Data length may be variable in some cases (e.g., in textual descriptions); however, a maximum length must be defined.

Data Format: describes how the kind of fact represented by the Attribute Type is expected to be portrayed when seen or heard (e.g., audio, video, display, etc.). This includes text justification, and placement of decimals and symbols where appropriate.

Data Domain: indicates the values that are permissible for the attribute. This can be expressed as a range or list of valid values.

Optionality: indicates if the attribute is mandatory (must have a value) or optional (can be left with no data entered).

RELATIONSHIP TYPE

A relationship is an association between two entity types that has meaning to the business. The minimum acceptable metadata for a relationship are:

Relationship Name: Always a verb or verb phrase given to a relationship. The name of the relationship reflects the activity or function that takes place between two entities. When read in sequence (entity-optionality-relationship name-cardinality-entity), an English statement is made about business operations such as "The employee sometimes delivers many services". (E.g. "The EMPLOYEE [entity type name] *sometimes* [optionality] DELIVERS [relationship name] *many* [cardinality] SERVICE [entity type name]". Note that entity type names are always singular to denote a pattern.)

Cardinality: A statement of the number of entity type instances that may or shall participate at each end of a relationship. Expressions include one (1), many (N or M), or a predetermined number. Entity type relationships are also described as one-to-one, one-to-many, or many-to-many. Specific numbers (two through infinity) are optional. Note: enforcement of specific cardinality (i.e., two through "less than many") must be done through programming.

Optionality: Expresses whether the association between an entity type occurrence and another entity type occurrence is mandatory or optional; also referred to as obligatory or non-obligatory.

Business Rules: Only a limited segment of an organization's business rules can be expressed in terms of an E-R diagram. They may need to be described textually if the business rule is to be applied through process rather than through the data.

83.3. Process Models

Process modeling is part of business modeling. It is a tool and technique used to formally document what activities the business performs in a structured way. Process models provide a powerful framework for analyzing the business information requirements and the transformations required. Process models:

- provide the basis for confirming and understanding business activity.
- provide a means to define business requirements before system is designed and built
- provide a framework for analyzing and improving current business processes and for business re-design.

- document the results of the analysis of business information requirements
- can be used to verify the data model
- can be used in analyzing organizational responsibilities

The analysis techniques chosen should allow modeling of each of the following key ideas to some degree:
- Business Activities
- Hierarchical Organization and Decomposition
- Data Objects
- Data Usage (i.e., association matrices)
- Activation, Sequencing, and Termination
- Governing Rules and Conditions

There is much more variety in the methods available for process modeling than is to be found in data modeling. Given this diversity, it is not feasible to recommend one approach over the other. Not every set of modeling techniques will employ all of these ideas, but most will reflect a fair sized subset of the notions presented here.

BUSINESS ACTIVITIES

The terms *process* and *function* are used by most modeling methods to describe those things that are performed by the business. The two terms are often used as synonyms, and the usage is inconsistent from method to method. Given the lack of agreement on these notions within the various methodologies, the term *business activity* could also be used. This notion is meant to convey the idea that the business does things in order to accomplish its work, and these may be referenced by name. Business activities should be defined in the terms or phrases that the business uses.

HIERARCHICAL ORGANIZATION AND DECOMPOSITION

Most activity modeling methods allow business activities to be decomposed into constituent activities, in a hierarchical fashion. This allows us to present various levels of detail and conversely, various levels of abstraction.

DATA USAGE

DAF recommends that the logical data model be validated against the process model. Data are created as output of a business activity and may become input to other business activities. This is often referred to as "data flow". Methodologies that utilize the concept of data analysis record the usage of data by a given business activity according to the "CRUD" matrix method. CRUD is an acronym for "Create, Retrieve, Update, and Delete". For each entity type represented in the data model, the pattern of data usage for each business activity is documented in a matrix, using the appropriate CRUD information.

ACTIVATION, SEQUENCING, AND TERMINATION

A description of how business activities or events are initiated or triggered, what the outcomes are, and what rules govern the sequence of activation should be included.

Office of the Chief
Information Officer
BRITISH COLUMBIA

GOVERNING RULES AND CONDITIONS

Even though previous techniques give a rich model of business behavior, the hardest part is to document the rules and conditions that govern the business activities. These are now commonly referred to as "business rules". There are a number of formal systems that have been proposed for documenting these various rules and conditions in a declarative fashion, as logical assertions. Most of these are difficult, and none have become a de facto standard. Any effort to capture business rules is worthwhile; however, at this writing DAF is not recommending anything beyond what is in the data and process sections.

93.4. Spatial Data Considerations

The traditional relational model is not quite sufficient for adequately handling spatial constructs. Relational modeling of spatial-temporal data would be difficult and the resulting implementation may be subject to performance limitations. Data interchange of spatial data is also a problem in part because there is not a standard vendor independent approach to modeling spatial data constructs that has been adopted, nor has the same rigor been applied to modeling spatial data that there has been to administrative, financial and socio-economic data.

The relational model envisions data as consisting of a number of relations (i.e. tables), each of which has a name and is composed of a set of tuples (i.e. rows). All rows in a table have the same structure, defined by a sequence of values corresponding to a given sequence of attributes. The possible values for an attribute constitutes the attribute's domain, which in turn comes from a given data type. Only a fixed number of data types are supported, typically including integer, real number, Boolean (true or false), character string, and date. Relationships between features represented in the data are based on the attributes' values and are defined independently by each application. This is done by matching values for one or more key attributes in a table to their corresponding values in other tables. The basis for doing this is a join operation.

Geo-spatial data brings together location, elevation, topology and features which may or may not change over time. Although a relational design can be implemented for geo-spatial data, it would lead to unacceptably poor performance because of the number of table joins required and the large size of the tables. Also, the type of indexing available in relational data bases only works well with simple or one-dimensional data. No support is available for true multidimensional data, which is a fundamental characteristic of geographic data.

Geographic data also has several other characteristics of note. Many queries are concerned with topological relationships of features to one another such as which provincial parks are adjacent to the ocean. Many types of metric relationships are also pertinent to basic queries, such as how much prime timber is within 100 meters of a stream. Another confounding characteristic of spatial information is that to be meaningful, it must be referenced to a co-ordinate system defined in the context of a geometric framework tied to the earth. The relational model was not designed to handle these characteristics.

SPATIAL MODELING AND INTERCHANGE GUIDELINES

The Spatial Archive and Interchange Format (SAIF, pronounced "safe") is (i), a language for modeling geographic data, and (ii), a vendor-neutral format for archiving and distributing such data. Developed as a means of sharing spatial and spatial-temporal information, SAIF is designed to facilitate interoperability, particularly in the context of data exchange. Its concepts can also be used, regardless of whether data is ever held in the SAIF/ZIP format. It is officially recognized as a national standard in Canada through the Canadian General Standards Board.

SAIF follows a multiple inheritance, object oriented paradigm. Constructs such as arcs, adjacency, datum and projection can be handled directly. In addition to supporting a standard set of such constructs, users can also define their own, as required. These new standards and specifications include a set of base types which encompass basic spatial or spatial-temporal concepts and which can be extended by the user. Consequently, the semantics of the data can be represented far more effectively, and the mapping between a conceptual model and the physical model is far more direct. As important, mappings between different models of the same or similar geographic constructs become possible.

A subset of SAIF is SAIF/Lite, which is designed for operational use. SAIF/Lite is supported directly by the Feature Manipulation Engine (FME (commercial)) and FMEBC (freeware) software products, as are a large number of proprietary GIS formats.

SAIF/Lite is endorsed by DAF as the guideline to be followed when modeling spatial-temporal data.

The Ministry of Environment Lands and Parks, Geographic Data BC is the custodian of the SAIF standard and FMEBC. For more information on SAIF, SAIF/Lite and FME and FMEBC visit: *http://www.env.gov.bc.ca/gdbc/fmebc* and *http://www.safe.com*.

103.5. Package Solutions

Package solutions from suppliers capture and maintain business information. The purchase of a package solution for a business system does not eliminate the need for a data model. A data model should be an integral part of the package software evaluation and installation processes. Depending on the methodology of the organization, any one or all of the following models may exist for the package solution.

BUSINESS REQUIREMENTS DATA MODEL
- Defines the data that the package must support
- May or may not be defined to the standards as expressed in this document
- Could be included in the RFP document that is issued to suppliers
- Used in the evaluation process to assess gaps and impact to the business of a package solution

PACKAGE SOLUTION DATA MODEL
- A data model provided by the supplier that represents the data scope of the package
- Not likely to conform to standard

- The RFP process can require suppliers to provide data model if possible to the minimum standard as expressed by this document, as part of the supplier's response to the RFP
- Ideally the supplier can be requested to relate the package solution model to the Business Requirements Data Model as provided in the RFP

PACKAGE IMPLEMENTATION MODEL
- The data model that represents the business data captured and maintained in the installed instance of the package solution
- This model is mandatory and must conform at a minimum to the standards expressed in this document

4. OTHER RELATED DOCUMENTS

114.1. Data Administration Forum *Terms of Reference*

124.2. Y2K Task Force Date Format standard

134.3. IT Framework (AC/IM)

144.4. Data Management Roles and Responsibilities

154.5. Land Data BC — Corporate Data Modeling Standards and Guidelines

164.6. GMOP
- Chapter 8 -- in the process of being rewritten by ISTA.

174.7. DAF Web Site
- *http://gww.acim.gov.bc.ca/*
  (a BCGOV1 userid/password is required to access this site)

5. REFERENCES

18Ministry of Forests Systems Development Guides
- S7: *Data Modelling Standards for Relational Applications*
- S19: *Data Naming Standards*
- S21: *Use of Standard Code Tables in Relational Applications*
- S35: *Management Guide to Custodianship*
- S38: *Data Administration Standards*
  (The above guides are available from the Forests' Data Administration web site at *http://www.for.gov.bc.ca/isb/datadmin*)

19Ministry of Transportation and Highways
- *Data Administration User Guide* (version 5.0)

*20*Ministry of Environment standards

LANDDATA BC
LandData BC Corporate Data Modelling Standards and Guidelines, Issue 1/1, Interim (1996/1997), Government of BC, September 30, 1996.

LandData BC Corporate Data Model Framework, Issue 1/1, Government of BC. September 30, 1996.

Spatial Archive and Interchange Format: Formal Definition, Release 3.2, January 1995, Surveys and Resource Mapping Branch (now Geographic Data BC), B.C. Ministry of Environment, Lands and Parks.

21Web Sites
- Data Administration Forum:
  http://gww.acim.gov.bc.ca/
    (a BCGOV1 userid/password is required to access this site)
- Ministry of Forests, Data Administration Section:
  http://www.for.gov.bc.ca/isb/datadmin/

6. APPENDIX A — GUIDING PRINCIPLES

*(from the Data Administration Forum's **Terms of Reference**)*

Data and information are the essential linkages between an organization [the *people*] and its systems.

The habit of managing data properly acts to unify systems development and business improvement activities, and reduces costs (see "modelling" below). The connection and similarity of data requirements across business areas must always be considered whenever business and system improvements or changes are planned. Expert data management/information management representation (normally a Data Administrator) is needed in all systems development and business improvement activities.

Data is a corporate resource that needs to be managed as an asset.

Big-picture thinking about data will need to be injected into virtually all changes and improvements in government. Data used in the operation of government are assets of the government. Good data management practices yield data that enables good government.

Data needs to be managed across its entire life cycle.

Data quality can only be achieved when it is integral to all aspects of the data management life cycle. This will require clear definitions, and implementation of accountabilities, processes, roles and responsibilities for data, beginning with planning.

Data can be, and where possible should be, shared.

The usefulness of data depends on the underlying form of the data, the original purpose for the data, and the degree of confidence that the data are representative of real world phenomena. The value of data increases with the extent that data is shared.

The whole of an organization has primacy over the parts.

The connection and similarity of data requirements across business areas should be considered whenever business and system improvements or changes are considered. This is not generally recognized or rewarded in current political and organizational systems, particularly at the corporate government level.

Modelling of data requirements yields sound and stable data structures.

Modelling adopts a rational, disciplined and systematic approach consistent with semantic and relational data management theories; this produces data structures that match business information needs. The information needs of an organization are generally more stable than the organization's business processes. Different (and changing) business processes can use unchanged data structures as long as those structures match the business information needs of the organization.