

E: Reports of Breakout Groups

CGKN Canadian Geoscience Data Model Workshop

Report from breakout group 1

Task: Develop a framework for coordination, funding, policy, and communication

Owner: Charles Roy

Facilitator: Franco Merlino

Participants: Trudy Curtis, Bryan Monette, Joan Waters, Lesley Chorlton, Phyllis Charlesworth

No report has been prepared by this group. The following was captured from the flip charts prepared by the group. Original Task to discuss policy, infrastructure, overall coordination, funding, and communication was redefined to "develop a framework for coordination, funding, policy, and communication", including such things as:

1. www presence - communicate
 2. Delivery mechanisms
 3. Technology transfer, training
 4. Mentoring - communicate
 5. Interaction with standards bodies
 6. Rewards and recognition
 7. Public policy, data sharing policy - communicate
 8. Language
 9. CGDI link and policy
 10. Client consultation; traditional and "new" methods
-
1. A) Re-state and B) clarify vision → Tailored Communication
 - Users
 - Developers
 - Managers
 - Political (Funding)
 - Prov\
 - > Nation
 - Fed /
 2. What are we going to produce?
 - 2a. Shared business model
 - 2b. Decompose to:
 - Products (deliverables)
 - Scope → Elephant - gazelle - mouse
 3. Plan for program
 - Coordination mechanism, e.g.
 - A) Metamodel
 - Meta - info - data --- web serve
 - Data models -- appl.

B) etc., etc.

1-3 Full timer, Program coordination, office! CGKN

Program coordination office

- Funding - proposals, stewardship
- Develop communication plan!
- Managing -> project delivery

National Geological Surveys Committee
1 member

CGKN - high profile
Program Office
Projects

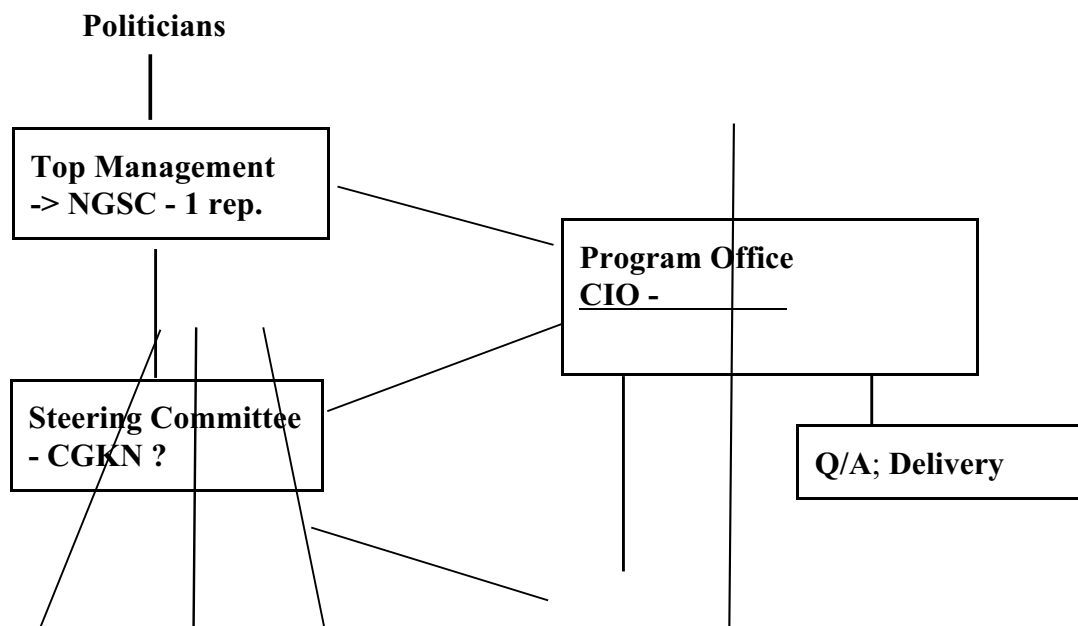
Short
Long

- clear structure
- clear accountability
- tie in to National Mapping program

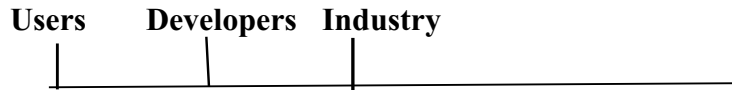
Tasks

- Input to policy - project
- cost recovery - licensing
- language
- standards
- GeoConnections - others

Program Office



**Academic
Research**



Communication Plan:

4. Web presence: internal
list server
chat room
shared repository
input and publish
Progress reports

- Web - external Beginning of portal
Marketing
Link to metacontent -- home for output task #2
Links to Geoscience projects
Links to CGDI tools
Links to framework (Task #3)
Publish - software (!!!)
 - Tutorials !
 - Standards !
 - Meeting minutes !

Project Charter: to get funding

Must have:

- Tech transfer; software, support
- conform to "NADM"
- training
- common standards
- mentoring
- network approach
- ongoing support → Plan for stewardship, ownership of tools
- identify support for shared business model

How to get money?

Endorse proposals to
CGKN

CGDI
MGDI
LTSP
NMP
Provinces

Marketing
Demonstration

Standards

- International standards
- must be mandated
- must be funded

Rewards and recognition

- Program = component
- Promote to NGSC to put items into Human Resource policy!
- Promotion Co
- Develop approach to effect promotion policies of effective agency!

Policy:

- Develop input into other policy mechanisms

Priorities:

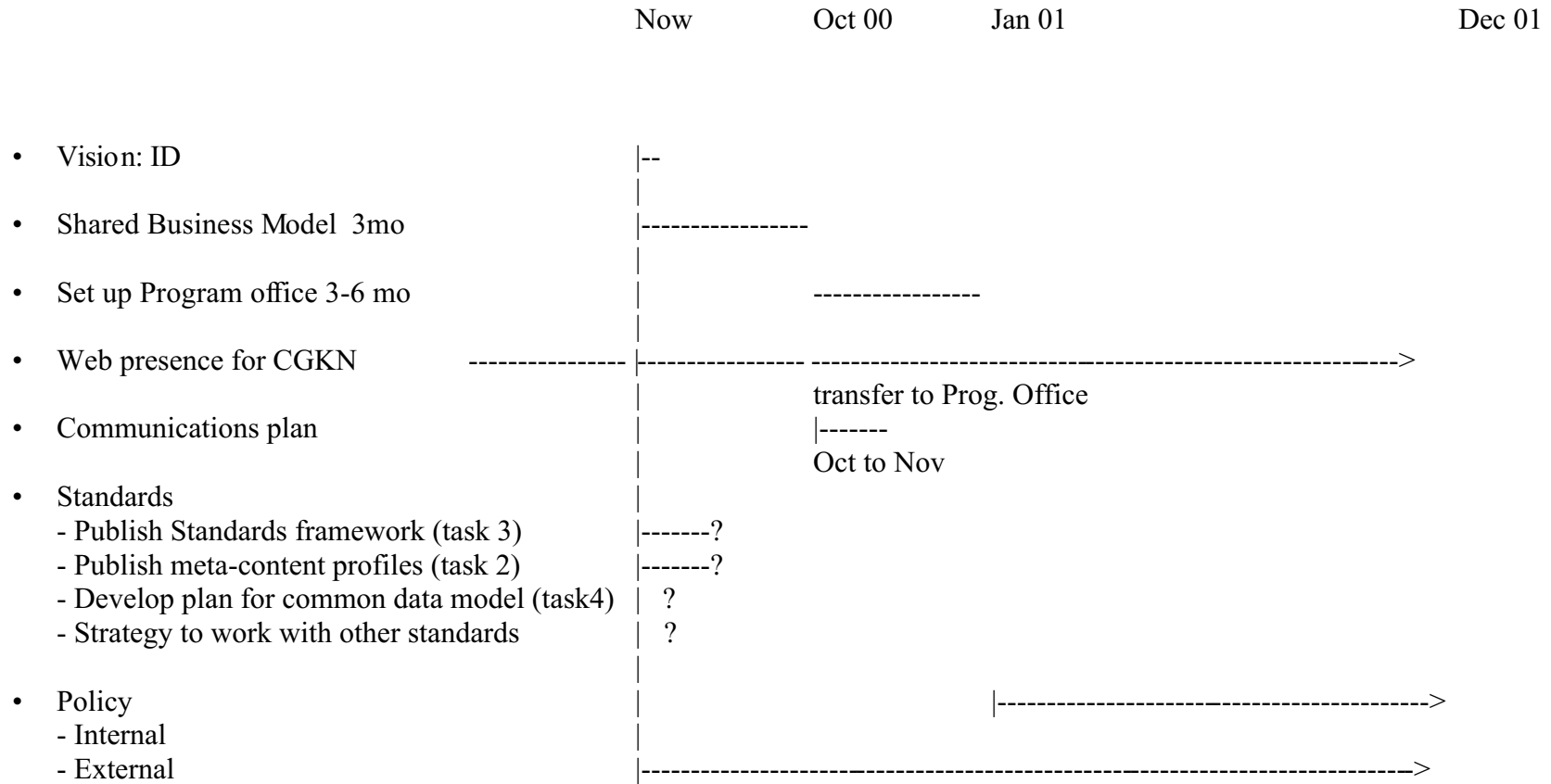
- Sharing data
- Citation
- Intellectual property
- Common licencing practice!
- Pricing
- Distributing!
- Copyright

Action Items

- Vision
 - CGKN subset
 - promote shared understanding
 - communicate, get funds, etc.
 - immediate - Now
 - 1 day
- Shared business model
 - CGKN
 - design governance and delivery structure; promote to

- now - concurrent vision
- 3 months
- Set up program office
 - funding search
 - endorse
 - start August to October
 - 2-3 full-time, October to Dec to develop main program, get plan rolling!
- Web presence
 - staged
 - status quo = CGKN
 - now → Oct.
 - expand → PCO - contributors from Network
 - Oct → TBD
- Communication plan: program office
 - Oct. - Nov. 2000
 - possibly - current staff + contract
 - build on existing
- Standards
 - publish standard framework; Task 3 - soon
 - publish meta info profiles; Task 2 - soon
 - develop plan for common data model; classification schemes, Task 4
 - strategy to work with
 - PPDM
 - NADM
 - ISO
 - Oct - Nov.
- Policy
 - a) Internal operational policy
 - CGKN – Prog. Office
 - Dec - Jan → ongoing
 - b) External policy
 - now - CGKN

Timelines, Task 1.



CGKN Canadian Geoscience Data Model Workshop
Working Group 2 Report:
The CGKN On-line Geoscience Data Catalog

Owner: John Broome

Facilitator: Phil Moir

Participants: Brian Fisher; Grant Abbott; Doug Irwin; Len Chackowski; Andre Lytviak; Jean-Marc Charbonneau; Bill Slimmon; Jamie Rupert; Mike Sigouin; Desmond Wynne

What:

Canadian government geological surveys have agreed that the Canadian Geoscience Knowledge Network (CGKN) should be developed to provide a single Internet portal to their geoscience knowledge and integrate into the Canadian Geospatial Data infrastructure (Dec. 1998; CGKN Workshop Report). This working group recommends establishing a project to develop a comprehensive catalog of Canadian government geoscience data and publications and an Internet search engine that will allow their discovery. The search engine should allow user-specified text or geographical searches of the catalog and direct the user to the source of discovered data. Efficient access to comprehensive and complete information describing available government data will be a valuable service to Canadians and encourage the utilization of geoscience data for a wide range of applications including resource exploration, identification of natural hazards, education, and environmental and climate change studies.

Why:

The group feels that one of the first priorities for CGKN development is a catalog containing consistent metacontent describing all government geoscience data and implementation of an Internet search engine. This project will solve the current difficulty in discovering what geoscience data and publications are available for a specific subject or geographic area from Canadian government geoscience agencies. During their history, geological surveys have invested vast sums of money in the collection of large volumes of geoscience data and in the interpretation of these data. By developing a complete and accurate catalog of Canadian Government geoscience data and an Internet tool to search this catalog, explorationists and geoscientists will be able to easily discover available geoscience data and how to obtain it. This project will encourage the utilization of geoscience for a wide range of existing and potential applications. The group feels the project will make Canada a more attractive mineral exploration target to both national and international companies by providing access to information that communicates the relative mineral potential of areas, aids in exploration planning, and communicates the existence of inexpensive data that can aid in exploration. The project will also encourage the utilization of government geoscience data for other important applications such as environmental and climate change studies, identification of natural hazards, support for education, and integration of geoscience data with other types of data.

Where:

The working group recommends that a team of federal and provincial experts develop the metacontent standard, the agency metacontent, documentation, tools, and the web site at agency

offices. The contents of the resulting data catalog should remain distributed at agency offices and maintained by agency staff.

When:

Short term:

- a) Within 18 months, create an on-line catalogue providing information about Canadian government geoscience publications and data holdings.
- b) Implement tools to support catalogue population and on-line text and geographic searches of the catalogue through the Internet.

Long term:

- a) Design/develop the catalogue and tools in a manner that ensures they are self-sustaining;
- b) Expand the catalogue and tools to encompass a broader range of data/information/knowledge.

Time Lines

(Refer to subproject table for activities)

Activity 1: "Model"

- | | |
|--|---------------------|
| 1. Establish working group | End of June 00 |
| 2. Model definition and Language | June to Sept. 30/00 |
| 3. Develop tools | July to Dec. 31/00 |

Activity 2: Preparation of metadata

- | | |
|--|---------------------------|
| 1. Establish readiness | Sept. 30 to Dec. 31/00 |
| 2. Establish allocations | End of Dec. 31/00 |
| 3. Support | Sept. 30/00 to Sept 30/01 |
| (self-supporting system after Sept/01) | |
| 4. Metadata preparation | Sept. 30/00 on |

Activity 3: On-line search tool

- | | |
|--------------------------------------|--------------------------|
| 1. Establish working group | End of June 2000 |
| 2. Inventory of existing tools | Sept. 30 to Dec. 31/00 |
| 3. Create new tools needed | Dec. 31/00 to Sept 30/01 |
| 4. Beta release | Sept. 30/01 |
| 5. Final release | Dec. 31/02 |

Who:

The working group recommends the following assignment of responsibilities:

Each NGSC agency will designate a contact person

- to be responsible for preparing metacontent describing their data and ensuring it is Internet accessible.
- Federal-provincial working group(s) will be created to guide the development of the project.
- A Project Leader will be designated by the CGKN Implementation Committee to lead the project and chair the working group(s)
- Contractors should be utilized where necessary to develop aspects of the project.

How:

Working group participants recommend the following approach. A project leader, working group, and agency participants should define a CGKN metacontent standard and assist each agency in bringing their metacontent to a common core standard through resource allocations and consultation. The metacontent standard should be a compatible subset of the CGDI metacontent standard. Agencies wishing to enhance their web presence beyond the core level will be given guidance on how to proceed but development of capability beyond the core (such as URL links to images of data or products) will not be resourced by the project. See the subproject activities table for more details.

How Much:

Although the resource requirements to create the geoscience data catalog and on-line search engine were not discussed in detail by the working group, there was agreement on the principles to be used to allocate resources to agencies. Since the goal of the project is to provide one-window Internet access to complete and comprehensive metacontent describing Canadian geoscience publications and data, resources will be allocated to agencies on the basis of need. Resource requirements could be determined through a survey of metadata readiness in each agency followed by a needs evaluation and resource allocation by the working group.

<p><u>Subproject 1:</u> “Model” for metacontent</p> <p>-----</p> <ol style="list-style-type: none"> 1. Establish champion who will organize a working group (WG) 2. Establish working group to: <ol style="list-style-type: none"> a) Define scope of profile b) Define subject classification Keyword pick list c) Define & recommend data priorities d) Resolve language issues (bilingual search capability) 3. Develop metadata creation tools and supporting documentation. 	<p><u>Subproject 2:</u> Agency preparedness of metacontent</p> <p>-----</p> <ol style="list-style-type: none"> 1. Establish readiness of agency metadata –contract? 2. Allocate funding support to agencies based on need to bring their metadata to a minimum level. 3. Provide support to agencies (communication & tools) 4. Agencies prepare their metadata and put it online. 	<p><u>Subproject 3:</u> On-line search tool adoption/development</p> <p>-----</p> <ol style="list-style-type: none"> 1. Establish working group 2. Working Group will: <ol style="list-style-type: none"> a) inventory online search tools b) develop single publications and data search web site c) support spatial and text queries. 3. Release Beta version 4. Release final version
--	--	--

CGKN Canadian Geoscience Data Model Workshop
Working Group 3 Report:
The CGKN Data Model Framework

Owner: Boyan Brodaric

Facilitator: Eric Grunsky

Participants: Don MacIntyre; Eric Boisvert; John Ernsting; Murray Journey; Kathryn Baker; Bert Struik; Jan Cramer; Paul Rennick; Dennis Chao

This breakout group considered the design and implementation of a common data model to be one of the most challenging tasks. The challenge lies in:

- 1) providing a model that accommodates diverse sets of geoscience data from the participants
- 2) developing of an interoperability standard to facilitate the exchange between the common data model and the participating agencies
- 3) implementing the data model, requiring the accommodation of diverse computing architectures and information dissemination mechanisms
- 4) developing a common data model and exchange mechanism that is reasonable in cost and requires minimum expense to implement and maintain

What: Build the data model framework

Building the data model framework requires three staged efforts:

- 1) Construction of the target model

Construction of the model requires that it be continuously reviewed, extended, and adopted by the participants. The target model must have a minimum metamodel with which stakeholders can effectively interact.

- 2) Addressing access and interchange formats and standards

The data model requires that a sufficient number of interfaces and protocols be established to insure realistic development and implementation. Features such as security and e-commerce capability should be considered in the design phase. Interoperability tools will form the backbone of enabling multi-agency participation.

- 3) Defining the requirements for a working model

Defining the requirements for the model will involve an evaluation-feedback process using a test implementation of the model. Scope and level of delivery will be defined during this process. Essential to the success of this initiative is effective communication between those who are developing the model and those who will be using the model. The model design should be scoped as to what can be realistically achieved by the participating groups.

Why: To enable access, usability and interoperability

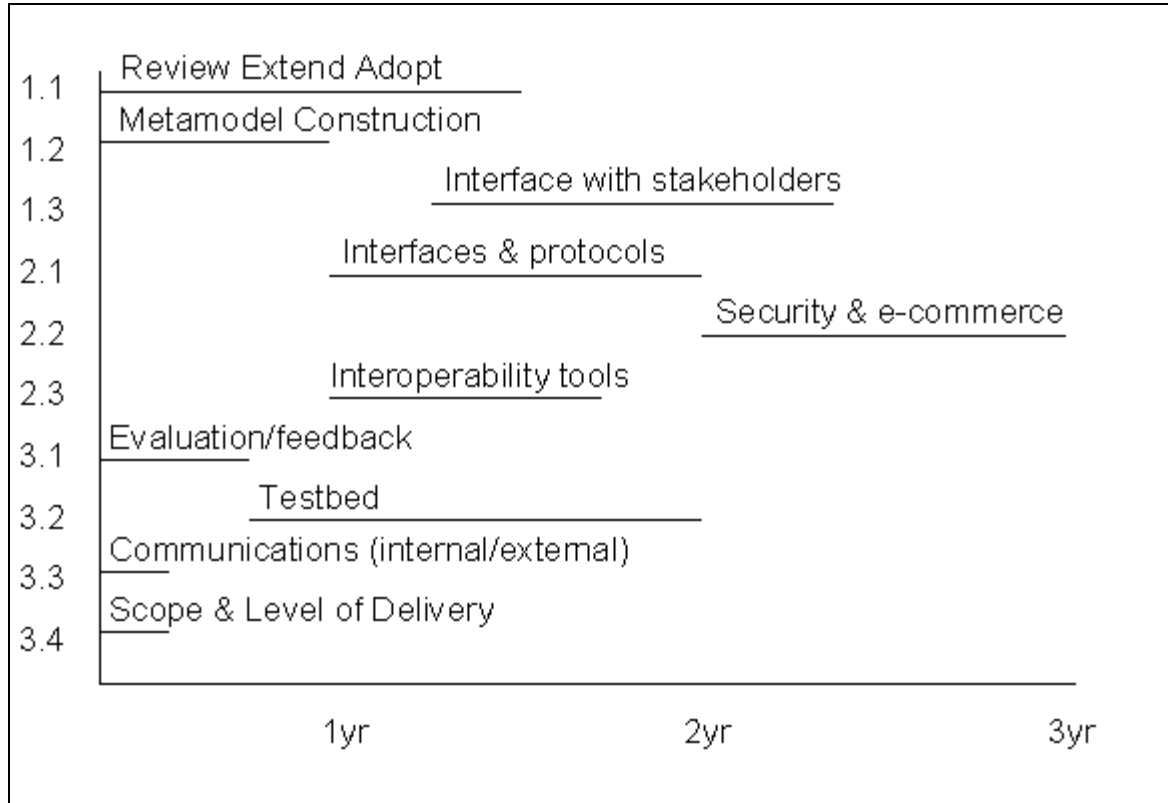
A universal data model will provide the framework for multi-agency data access which is considered a primary goal of the Canadian Geoscience Knowledge Network. The data model should be built to accommodate the interoperability requirements of the CGKN Data Model Working Group.

Where: Source: Canada and partners; Destination: Global

The existing models are currently being developed through the combined efforts of the provincial (Canada), state (U.S.) and federal surveys(GSC, USGS).

When: A three-year plan

The following chart provides the time line for specific tasks.



Who: CGKN Working Group and CGKN Data Model Working Group

Given the resources and knowledge required to consider the breadth of geoscience information the current sources of development should continue. Funding would be obtained from respective agencies based on local and international requirements to meet the goals of the CGKN.

Agreement must be reached as to who will be involved in the development of the model. Should the currently active participants (Boisvert, Journeay, Brodaric, Soller) continue or should there be additional members added to the development group?

How: Workshops, meetings

The creation of a data model will be a labour-intensive activity requiring extensive discussions, communications, workshops and meetings. Workshops and meetings can be combined with regular national meetings within Canada (PDAC, GAC, Cordilleran Roundup, etc.) and GSA meetings in the United States. A common website for the availability of the testbed site and associated tools is a necessary requirement. An agency/agencies need to be identified who will support the dissemination of the information and housing of the test site.

The group generally agreed that development of a common data model is the most challenging task facing the CGKN. It will also likely be the most costly. The challenge will be to accommodate local needs while keeping the model sufficiently simple.

How much: ~\$737,500 (CDN)

An estimate of ~\$737,500 (CDN) has been made based on the rough estimates of requirements made during the breakout session.

Tasks	Yr	Person	PY	
Review,extend, adopt	1.1	1.5	1	1.5
Metamodel	1.2	1	1	1
Interface stakeholders	1.3	1	1	1
Interface and protocol	2.1	1	1	1
E-commerce, security	2.2	0.5	1	0.5
Interoperability	2.3	0.75	1	0.75
Evaluation, feedback	3.1	0.5	1	0.5
Testsite	3.2	1.5	1	1.5
Communication	3.3	0.25	1	0.25
Scope	3.4	0.25	1	0.25
	Total	8.25		8.25
		Cost	\$70,000.00	\$577,500.00
		Testsite		\$100,000.00
		Workshop		\$60,000.00
				\$737,500.00

The CGKN thematic content requirements to develop data model extensions

Owner: Peter Davenport

Facilitator: Beth Sage

Participants: Ward Kilby, Andy Moore, Ross Kelly, Sonia Talwar, Celine Gilbert, Larry Nolan, John Glynn, Don Desnoyers, Dave Soller

What: Develop thematic or discipline-specific components of the geoscience data model. This resolved itself into establishing subgroups to model specific disciplines, and outlining a generic workplan for these subgroups to follow.

There was some discussion of how the task of developing an overall geoscience data model should be broken into manageable components. Suggestions included: by client requirements (i.e. mineral exploration, oil and gas, environmental applications, etc.), by spatial dimensionality (i.e. 2D surface map data, 3D sub-surface data), or by geoscience subdiscipline. Geoscience subdiscipline, or theme, crosses client requirements and dimensionality, but issues such as science language are subdiscipline-specific. A breakdown based on subdiscipline was confirmed as the best approach. Satisfying the business requirements of client groups will depend on ensuring interoperability between the discipline-specific components of the data model, which will be addressed in the overall data model design. Methods for satisfying the need for a third dimension will be dealt with once the key components of a 2D solution are determined.

The priority themes were identified as: 1. Geology, 2. Geochemistry, 3. Geophysics, and 4. Mineral Deposits. It was suggested that Surficial Geology and Bedrock Geology be combined into one Geology theme, but this idea was rejected in post-workshop discussions because of the distinct languages and classification schemes that have evolved in the two subdisciplines. Other themes that were recognized but considered as lower priority by the breakout group included: Curation (sample collections), Hydrology, Marine Geoscience, Oil and Gas, and Paleontology. Work may proceed on these latter themes if leaders emerge to carry them forward.

Why: The practice of Geoscience is divided into several subdisciplines (e.g. geophysics, surficial geology, geochemistry, etc.), each with its own data types and terminology. It is important to ensure that each is properly represented in the data model, which requires the involvement of specialists in these subdisciplines. In order to ensure interoperability, the science language used to describe features must be standardized and documented, and the choice of attributes to be recorded and their inter-relationships must be agreed upon. Work to develop discipline-specific data models is already underway in some CGKN agencies, so the challenge is to make sure that these components work together and share common design elements.

Where: Subgroups will draw members from government geological agencies in Canada having data holdings in each of the geoscience disciplines. Subgroups will also participate in parallel international groups such as the North American Data Model Science Language technical team.

When: The following generic timeline was proposed by the breakout group to guide the sub-groups formed for each discipline:

	Activity	Duration
1.	Identify leaders and committee members (from each survey), generate participation	1 mo
2.	Define end-uses and measures of success	1 mo
3.	A. Identify key attributes and prioritize B. Science language - in close contact with NADM and others (for at least one key attribute) C. Identify and examine existing data structures/models and interoperability between them	12 mo (A+B+C)
4.	Bring results of Activity 3 and defined requirements to overall data model group (see report from breakout group #3) and collaborate in integration of them into data model (extensions)	6 mo
5.	A. Define and perform projects using real data in physical databases B. Identify requirements for tools to populate the database (A. and B. in close cooperation with overall data model group and other subgroups)	open

Step 5 is open ended and will be dependent on resources available. Steps 3, 4 and 5 could occur simultaneously. It was agreed that for disciplines for which one or more models already exists, Step 5A had to take place before participants could properly judge the appropriateness of, or suggest changes to, any particular model.

Who: Each subgroup will have co-leaders (one federal, the other provincial/territorial), who will recruit members from the agencies that comprise CGKN. There was some suggestion that both co-leaders for Mineral Deposits be provincial. Leaders volunteered or were suggested for Geochemistry and Geophysics; others volunteered to identify and contact potential leaders for Geology and Mineral Showings. A leader for Surficial Geology has since emerged. Activities 3C, 4 and 5 might benefit from the participation of external contractors.

How: Activities 1, 2 and 3A and B will be addressed by subgroup members. For Activity 3B (Science Language) it was suggested that the most efficient approach be to attempt to use, or modify then use, an existing Language standard. Activities 3C (examining existing models), 4 (data model design) and 5 (data model implementation and population) could be done by a combination of survey staff and external contractors with expertise in data modeling and front-end tool development.

Two Bedrock Geology datasets and one Surficial were identified as ready to be converted to some physical form of NADM. The two bedrock projects (Yukon 1:250,000 compilation and Newfoundland GeoLegend data) were noted to be likely to proceed in some form with or without CGKN, and therefore could be enhanced by CGKN and the results shared with the group. It was later discussed that such projects might form the nucleus of a project proposal to take existing digital geological map data from up to four agencies, map them into a common data structure based on the North American Data Model (NADM), develop some common science language (or apply existing classifications), and deliver the data in a consistent format via the Internet. Such a project would make use of a potential core of the overall data model (NADM), develop group experience in a distributed, physical implementation of it, and further the development of database population, editing and distribution tools.

How Much: Until the subgroups have started their work it is not possible to estimate the resources that will be required. The primary roadblock is over-extension of staff who may be unable to devote as much time to this as it needs, because their agencies are unable to backfill whatever part of their jobs they would be abandoning. Travel costs for workgroup members to meet periodically are also an issue. Support of travel costs and alleviation of time constraints, perhaps by addition of temporary staff and/or by prioritization or redistribution of current duties, is essential for this endeavor to proceed expeditiously.

.