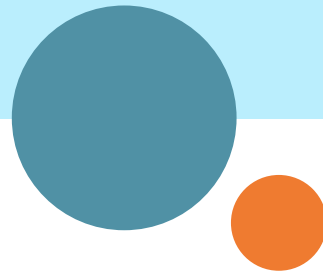


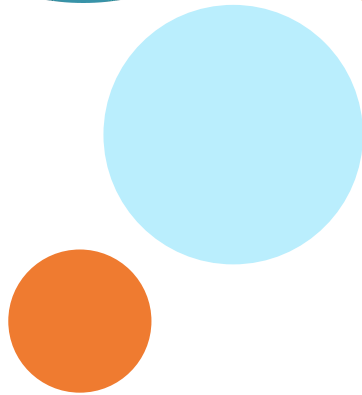


From Theory to Practice

Translating Schemas into Databases and Interfaces



Sarah Vela
MERIDIAN



Overview

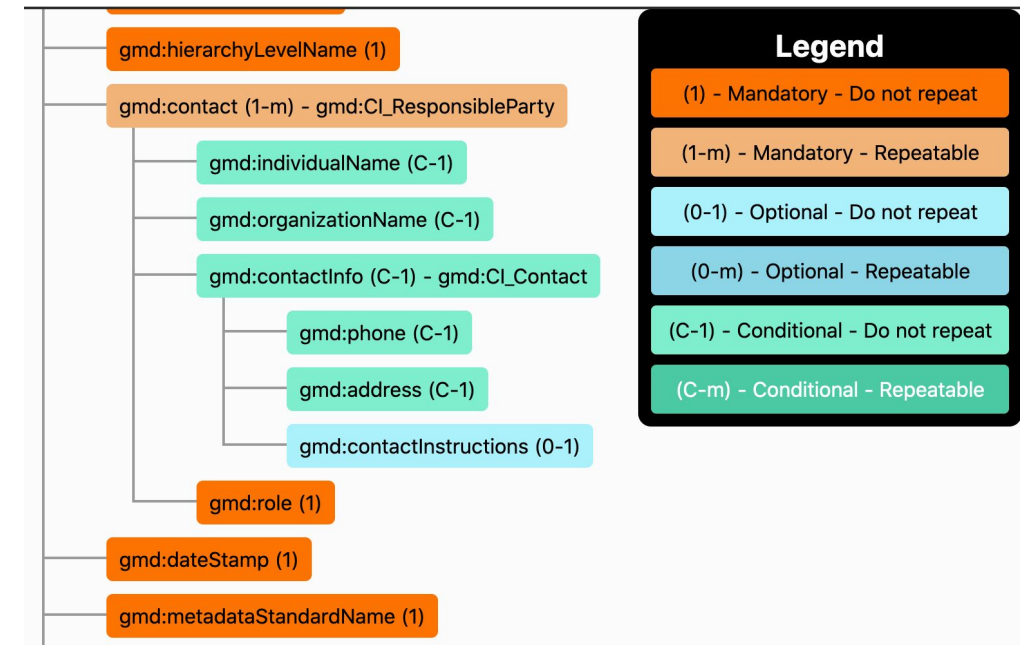


- Introduction to the example case
- Why do schemas differ from tools? Some thoughts...
- Translating a schema
 - How it works
 - Towards a schema translation language
- Examples from the MERIDIAN Discovery Portal development

MERIDIAN Metadata Profile



- Visual version of profile:
<https://docs.meridian.cs.dal.ca/metadata/>
- Tree structure indicates the hierarchy of classes, subclasses, and elements
- One of three requirement types:
 - Mandatory – must be included
 - Optional – can be included at user's discretion
 - Conditional – may be required depending on a given condition
- Repeatable or non-repeatable
 - How many times an element can be included



The MERIDIAN Discovery Portal and Submission Form



- Online tool meant to support marine resource reuse by helping users find relevant data, software, etc.
- Not a data repository, but a niche search engine
- Relies primarily on manual metadata submission by resource holders
- Five types of resources:
 - Software
 - Model Output
 - Vessel-tracking Data
 - Acoustic Data
 - Bioacoustic Data

This is a screenshot of the top portion of the MERIDIAN submission form. It features a vertical stack of tabs on the left side. The tabs are labeled: "Citation details" (highlighted in blue), "Resource details", "Extent", "Methods", "Access details", "Hardware details", "Other", and "Summary". At the bottom of this stack are two buttons: "Save" and "Submit". Below the "Submit" button is a small circular icon containing an information symbol (a lowercase 'i').

People or organizations involved in the product
You can either load an existing entity using the dropdown or a new entity. An entity may be an organization (in which case, name) or a person. You can drag and drop this field to reorder order used in the citation.

This is a screenshot of the input fields for the "People or organizations involved in the product" section. At the top, there are two small square buttons: one with a plus icon and one with a minus icon, followed by a text input field containing "Enter a new person...". Below this, there are two input fields side-by-side: "Personal (First) name" with the value "John" and "Initials" with the value "A.". Below these is a single input field for "Organization name (e.g., as an entity, or as a)" with the value "Government of Canada". At the bottom is an input field for "Email address" with the value "john@gov.ca".

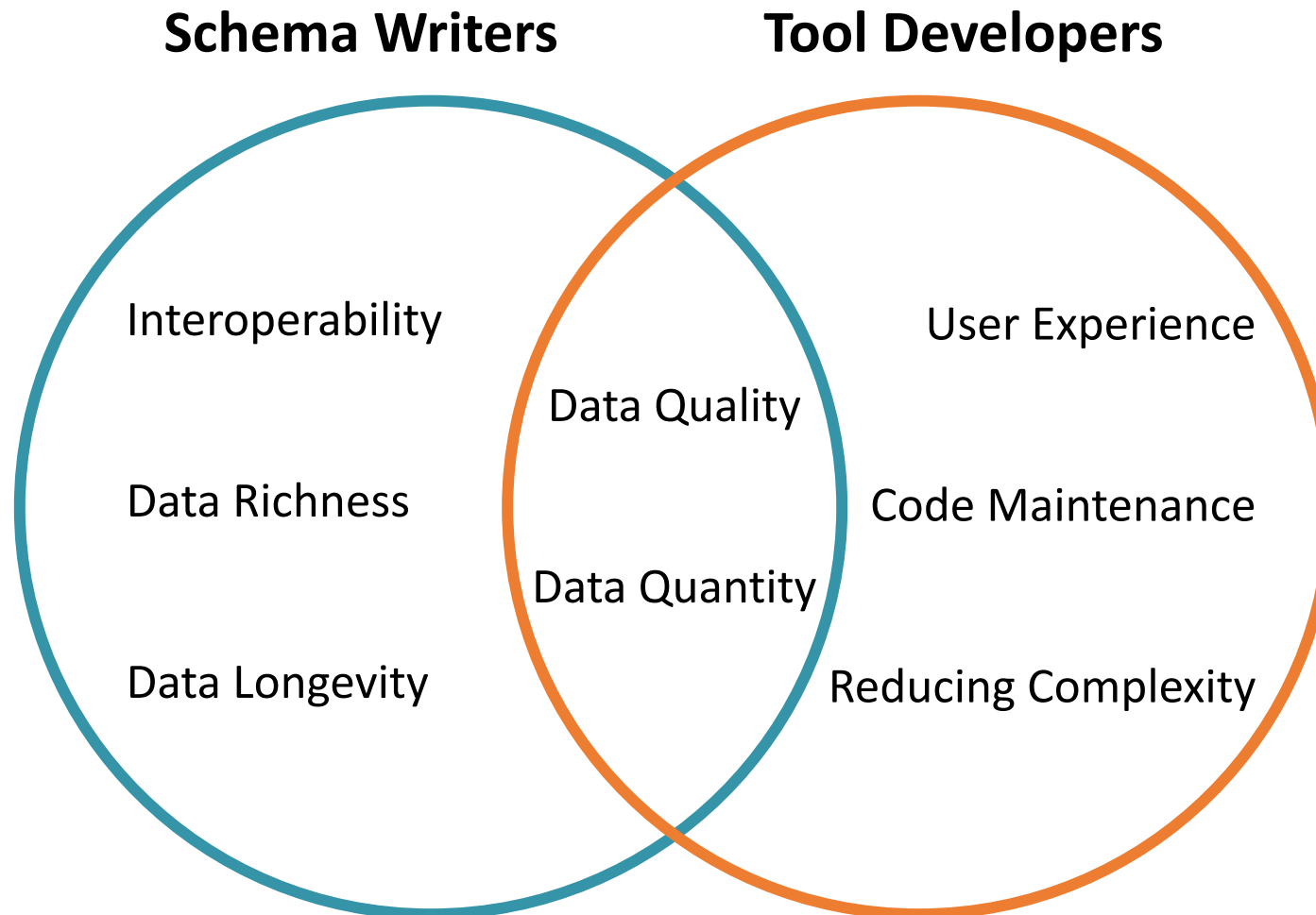
- To contribute: <https://discovery.meridian.cs.dal.ca/index.js>

Philosophical Positions



- Schemas are very important
 - Speed up development
 - Prevent costly errors
 - Ensure tools are as useful as they can be
- Schemas are often not made by developers
 - Made by information specialists and subject-domain experts
 - Collaborative or heavily negotiated – not looking for further input
- Schemas want to be blueprints
 - Often difficult to translate directly
 - Arguably not desirable to translate directly

Competing Priorities



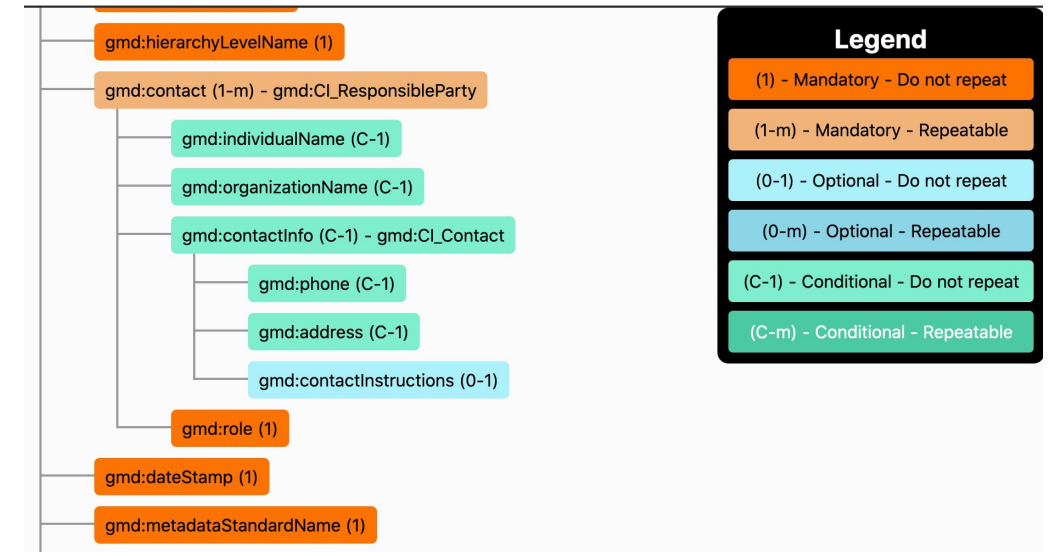
Design Philosophy



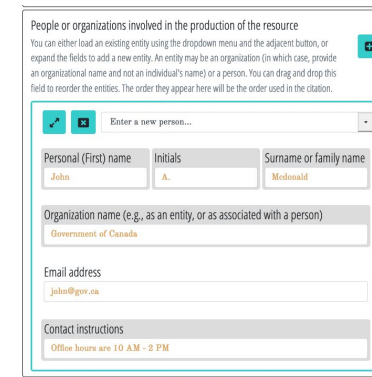
Translating Schema Language



- Relationships
- Repeatable classes and elements
- Requirement types
 - Mandatory
 - Optional
 - Conditional



- E.g. infinite hierarchies



Repeatable



- Defines database structure (one-to-one vs. one-to-many)
- JavaScript duplication of fields
 - Leads to overwhelming forms
- Multi-entry tools
 - Require instruction
 - More complicated validation and post-processing

The screenshot shows a web form with multiple sections. The top section has a label "is delivered in" and a text input field with the placeholder "sion format, or physical format if only physical access is available". Below this is a red-bordered box containing a label "red in." and a text input field with the placeholder "Enter format version". The text "Broadcast Wave Format" is entered in this field. Below this is another identical red-bordered box. At the bottom, there are two sections for selecting subjects and keywords. The first section is titled "Select some subject(s) relevant to the resource" and includes a text input field with the placeholder "Click in the field to see some available options. These subjects are taken directly from ISO". Below this are two buttons labeled "Inland waters X" and "Disaster X", followed by a text input field with the placeholder "Enter text here to search. You may add 1". The second section is titled "Select or enter some keywords relevant to the resource" and includes a text input field with the placeholder "To enter keywords, simply type. You must press enter or semicolon (;) to end a keyword.". Below this are two buttons labeled "offshore structure X" and "river station X", followed by a text input field with the placeholder "Enter text here to search, or typ".

Repeatable – Form Complication

- 2 hydrophones
 - Each calibrated once
 - 1 minor data corruption issue
- Attached to a rig deployed by 3 grad students

Describe the instrument(s) used in data collection

Type of instrument used
Click to see suggestions, or enter your own value.

Name or identifier
Provide a unique identifier for the instrument.

Hydrophone HYDRO394A

Description of instrument, including details necessary to understand the resource
Sound Fish ASF-1 MKII Hydrophone

Describe events which occurred to the instrument during data collection
This may help others understand periods with high error rates or missing data.

Select the general category in which this event falls
calibration

Description of the individual event
instrument re-calibrated at halfway point

Select the general category in which this event falls
dataLoss

Description of the individual event
Corruption of recording from 3:45:52 to 3:46:44

Type of instrument used
Click to see suggestions, or enter your own value.

Name or identifier
Provide a unique identifier for the instrument.

Hydrophone HYDRO394B

Description of instrument, including details necessary to understand the resource
Sound Fish ASF-1 MKII Hydrophone

Describe events which occurred to the instrument during data collection
This may help others understand periods with high error rates or missing data.

Select the general category in which this event falls
calibration

Description of the individual event
instrument re-calibrated at halfway point

Provide details on the platform on which instruments were mounted. Note: If information is provided here, at least one instrument must be provided.

Identifier of platform
Buoyed platform

Description of the platform, including details necessary to understand the resource
An anchored buoy holding a 6-foot wide metal rod at a fixed position. Hydrophones are attached to each end of the rod for data recording.

People or organizations involved in the deployment or development of the platform as used to gather the resource
You can either load an existing entity from the dropdown, or expand the fields and create a new entity. An entity may be an organization (in which case, provide an organizational name and not an individual's name) or a person. These fields can be reordered using drag-and-drop.

Enter a new person...

Personal (First) name Initials Surname or family name
John A. Smith

Organization name (e.g., as an entity, or as associated with a person)
Dalhousie University

Email address
jsmith@dal.ca

Contact instructions
Office hours are 10 AM - 2 PM

Enter a new person...

Personal (First) name Initials Surname or family name
Dave K. Johnson

Organization name (e.g., as an entity, or as associated with a person)
Dalhousie University

Email address
d_johnson@dal.ca

Contact instructions
Office hours are 10 AM - 2 PM

Enter a new person...

Personal (First) name Initials Surname or family name
Anne A. Moore

Organization name (e.g., as an entity, or as associated with a person)
Dalhousie University

Email address
amoore@dal.ca

Contact instructions
Office hours are 10 AM - 2 PM

Mandatory



- Form validation
- Database schema requirement
- Gold for search design
- What if the value is not known?
 - Can placeholder values be used?
- Can any of the values be pre-filled?
 - Contextually set values
 - Presumed values (overwritable default)

Title

This will be the title used to cite your resource.

Abstract or description

This should be a brief summary of the resource. Remain general in this field - more specific fields follow.

Resource status

Select the status most appropriate to the resource.
You can hover over options to see more details.

Ongoing

Optional



- Usually indicated with an mark (*) or visual difference
- Limited search use - risks not returning relevant results
- Relational databases – empty fields increase size
- Code has to accommodate value maybe not existing
- Too many optional fields crowd forms and dissuade users

Select or enter the license applied to th
Free text input is allowed if the license you use is

Enter text here to search, or hit enter to

Resource version identifier

Most relevant if substantial changes may occur be

2.1

Conditional



- The nature of the condition is paramount
- Simple conditions:
 - Different forms
 - Responsive forms
 - Responsive search
- Complicated conditions:
 - Provide clear instructions
 - Catch what you can (e.g. custom validation)
 - Hope

Whenever gmd:CI_Citation is used,
there must be at least one
gmd:CI_Citation with at least one
gmd:contactInfo containing an electronic
address



You must provide either an
email address or phone
number for at least one person

A Translation Vocabulary



Schema Term	Development Term	Description
Mandatory	Mandatory - Default	Value is predictable (with or without override option)
	Mandatory	If they don't know, we don't want the data
Mandatory/Conditional	Mandatory - Unknown	If they really don't know, we still want the data
Conditional	Conditional – One Of	One field from a set is mandatory, but not each
	Conditional – If/Then	If a field has a value, others are mandatory
Optional	Optional - Compromise	We would make this mandatory if we could
	Optional - Other	Could be collapsed into one field with example values
	Optional - Fishing	Very low likelihood of user completion, deprioritize

Thank you!
Any questions?

