





Environment Canada Data Catalogue

Paul Paciorek

Manager - Data Management Information Management Corporate Services Branch

February 19, 2013



Context: Environment Canada's Data Landscape

- A department with a wide variety of science-based program areas
 - Ecosystem Sustainability / Science & Technology
 - Biodiversity Wildlife and Habitat
 - Water Science
 - Atmospheric Science
 - • •

Weather Environmental Services

- Weather Observations, Forecasting and Warning
- Climate Information, Predictions
- WES Services for Targeted Users (NavCan, DND, ...)

Environment Protection

- Substance & Waste Management
- Climate Change & Clean Air
- Compliance Promotion & Enforcement





EC Data Landscape

- 3 Dimensions: Variety, Volume, Velocity

1. Variety (range of data types, sources)



2. Volume (amount of data)

Exponentially Increasing Size & Frequency



Environmental Monitoring, Research, Modeling, Prediction, Permitting, Regulating Sensor Networks Mobile Sensors Satellite/Radar Imagery Supercomputer Distributed Databases Scalable Storage Systems Component-Based, Service Oriented Architecture

3. Velocity (speed of data)



Context: EC Data Management Program (ECDMP)

- Action Plan at a Glance

- Five interdependent, foundation projects:
 - 1. Data Governance and Architecture
 - 2. Data Catalogue Data Discovery
 - 3. Data Portal Data Access and Sharing
 - 4. Data Consolidation
 - 5. Data Integration & Preservation



- Data management services in support of programs
- Foster a data management culture.





ECDMP – Target State

Incrementally implement the target EC Data Architecture.



• Key Principle: Act Local, Think Global, Progress Incremental







Spotlight: Data Catalogue - Why?

If we cannot find our data...

- We cannot access/use/reuse it.
- We risk collecting it more than once.
- We cannot share it.
- We cannot publish it.
- We cannot preserve it.
- We cannot manage it. ٠
- We cannot cite it / get credit for it. ٠
- We cannot verify it.
- ... Cannot leverage it to its full potential!

Can you imagine:

vironment

Canada

- A public library without a library catalogue?
- A DVD store without categorized sections?

Canada

- A grocery store without categorized and organized aisles?
- A science-based organization without a Data Catalogue?









Spotlight: Data Catalogue

- 1. Data Search & Discovery
- 2. Data Inventory & Preservation
- 3. Data Publishing & Sharing ("Interoperable")
 - Internal & External
 - GC Open Data, Partners, Science Departments, Federal Geospatial Platform

4. Compliance

- TBS Policy: TBS Standard on Geospatial Data, RecordKeeping Directive, ...
- Audits





- How it works? > "Describe, Publish, Discover"



1. Describe







EC Data Catalogue Canadă Teame exerch exerce teato to total Type Search keywords here EC Data Catalogue Search Advanced Search

Government Gouvernern of Garacia du Caracia Data Stewards use standards-based metadata creation features to quickly & easily create metadata that makes their data searchable and discoverable

✓ Datasets Metadata - ISO19115-NAP
 ✓ [Monitoring Site Data] - OGC SensorML

Data Stewards use standards-based publishing features to publish metadata to:

- EC Data Catalogue (intranet)
- External portals/applications (internet) (e.g. GC Open Data Portal)

Users search & discover environmental & scientific data via the Data Catalogue's search interface or external applications/portals.

- How it works? > Standardized approach for Data Publishing

GC Open Data (Data.gc.ca)	Geoconnections (geodiscover.cgdi.ca)
Français Home Contact Us Help Search canada.gc.ca Search by Keywords (e.g. weather, immigration,) Search	None Start Gatague Start Rap Note Office Find Data or Services dep Enter a search term and dick the Search button to perform a search. pair water quality Search Advanced Search >> @
	Federated Search/Harvest (via standards)
	External (Internet) Data Catalogue Interface (API) EC Data Catalogue
	"Describe, Publish, Discover" Internal (Intranet)

Other applications, departments, partners, research organizations, World Meteorological Organization, etc...

Slide 9

- Key High-Level Requirements

"Describe"

- Web-based, bilingual application (compliant with GC Web Standards)
- Ability to create and manage standards-based metadata
 - Support GC metadata standard for geospatial data (ISO 19115 NAP)
 - Ability to define custom metadata forms, profiles and templates
 - Ability to bulk import metadata
- Ease of use for non-technical users
 - Basic/Advanced Metadata Editor View; Metadata templates

"Publish" (interoperability)

- Ability to publish metadata to the internal and external applications/portals
 - Cataloguing standards (federated search/harvesting)
- Ability to manage metadata publication workflow processes
- Ability to harvest metadata from other catalogues/repositories

"Discover"

- Ability to perform effective basic & advanced searching to find metadata.
 - Basic Search, Advanced Search, Facetted search, Location-based search
- Ability to provide reporting functionality on content and usage statistics.





- Technology Used

GeoNetwork pensource

http://geonetwork-opensource.org/

- Addressed mandatory requirements
- Strong support for international standards ("interoperability")
- Flexible configuration/customization options
- Active development community
- Successfully deployed in a number of large organizations:
 - ON Ministry of Natural Resources (<u>Land Information Office</u>), Natural Resources Canada, World Health Organization, United Nations, GEOSS GEOportal, Dutch National GEO Registry, ...







EC Data Catalogue - Implementation Schedule

Phase 1 Apr/12 – Aug/12			
GC Web Interface (Bilingual+Accessible)	Pilot Aug/12 – Dec/12		
Metadata Editor Enhancements – Dataset, Monitoring site	Collect metadata for an EC Program.	Phase 2 Nov/12 – Mar/13	Deployment
Montoning one	320+ records collected.	Publishing Approval Workflow	Mar/13 – Apr/13
Basic Search/Discovery Enhancements		Metadata Editor	Soft Launch (Mar/13)
Compliance to Metadata Standards (TBS, ISO, OGC)		 Improvements Biological template (CWS) 	User Acceptance Testing
(, -,)		Data Discovery	Departmental Launch (Apr/13)
	4	 Faceted Search 	
		•View Metadata •External Data Catalogue Interface	Communication & Engagement
		(API)	CSB Service Catalogue Update
		User Interface Sizzle!	
		 Home page design/layout Look and feel 	Operational Service
		Data Catalogue Training Centre	

The Potential of "Interoperability"

- Federated Data Discovery Network

 Federated Data Catalogue tools that apply <u>common standards</u> are promising for scientific data discovery across Canadian research institutions.



Key Elements:

- Metadata Standard: ISO19115 North American Profile (TBS Geospatial Standard)
- Data Catalogue Interface Standard: CSW 2.0, OAI-PMH, ...



- Interoperability: Data Catalogue Interface Standards

• Standards which define a common interface to discover, browse, and query metadata about data, services and other potential resources.

• Metadata Harvesting:

- the process of periodically collecting remote metadata and storing them locally for a faster access.
- Not just an import: local and remote metadata are kept in sync.

Examples:

- OGC Catalogue Service for the Web (CSW2.0.2)
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Z39.50





- Interoperability: Metadata (1)

- Dataset Metadata
 - North American Profile of ISO 19115:2003 Geographic Information Metadata
 - Government of Canada Standard
 - Standard on Geospatial Data: <u>http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16553§ion=text</u>
 - GC NAP-Metadata Website: <u>http://nap.geogratis.gc.ca/metadata</u>

Other types of metadata/data implemented:

- Biological template of ISO19115 NAP
- Monitoring Site Data (based on OGC SensorML standard)





- Interoperability: Metadata (2)

- Common misconception about Geospatial Metadata:
 - Too complex/advanced; For technical GIS experts only; ...
- Developed an <u>EC Metadata Profile</u> that identifies core metadata elements
 - Example of core elements: Title, Date, Abstract, Keywords, Time Period, Geographic location, Online Resource, …
 - Can be applied to all types of data
 - Basis for proposed/draft TBS Open Data Metadata Profile
- Metadata Editor allows toggle between:
 - 1. Basic View: Core elements in 1 simple form
 - 2. Advanced View: Full ISO standard broken down into several sections (for advanced users)





EC Data Catalogue - Interoperability: Metadata (3)

Mandatory	Defi	-													
ree Text			n Options												
Section	0.0	- COW	T Spanns												+
Metadata Re	ecord		Dataset Identifi	cation	Thesaurus (repeat.)	Descrip.	Keyords	Contact Info		Extent	- 22	Constraints	Disti	oution Info.	
		-				-		_	-						
File Identifier	11		Title		Subject / Topic	Keyword		Individ. Name		Time Period		Legal Access	Online		
							6. 1.					Constraints	Reso		
Date Stamp			Date		Title	Type		Org. Name			Begin				
	-	-		-					-		Position	LegalUse		Linkage	
Date Modified			Date Modified		Date			Position Name			End Position	Constraints			
				5										Transfer Size	
anguage			Date Type		Date Type	-		Tel. Number	Ц.	Geo. Bound. Box	_			Protocol	_
Dataset URI	11		Abstract		Organization Name			FaxNumber		-				PTOROCOL	
											North			Description	
French French			Status			-		Civic Address			Bound Lat				
fierarchy Level			orantus			-		Crivic Avaaress	-		South			Format Name	
											Bound Lat.				
			Language					City							
Drganization											West Bound. Lat.			Format Version	
	-		Character Set					State / Prov.	<u> </u>		L.P.				-
Contact				-							East Bound	2			
204.030 M					(a.)			-			Lat	2 M 10			
Metadata	-		Maintenance and			-		Postal Code	_	Geographic	-				-
Standard Name			Update					Country	<u> </u>	Region Name					
			Frequency												
Catalog Type								Email							
		-	Supplemental Info.			-		Online	-		_				-
								Resource							
				Program URL											
								Hours of Service							
	++			Data				ORIVIOR	-						
		_		Dictionary		_									
		-		General				Roles							
				Contract of the local data		-									
				Additional											
		_	-	Metadata		-									
	-		Data Series					-							-
				Dataset											
				Series											
		-		Name	-										-
			-	Issue											
		_		Identificatio		_									
				Data Series											-
				UFIL											

DEMONSTRATION



