



World Data System

World Data System and Canadian CoreTrustSeal Cohort Members Needs Assessment - Reviewer Instrument and Supplemental Information

March 2023

Caroline Lee
Research Associate for International Technology Office
ito-ra3@oceannetworks.ca

Sarah Gonzalez
Program Manager for International Program Office
sgonzal4@utk.edu

Karen Payne
Director for International Technology Office
ito-director@oceannetworks.ca

A REPORT BY THE

World Data System • International Technology Office (ITO) and International Program Office (IPO)
Funded under the Digital Research Alliance of Canada (The Alliance)

Cite this report (APA):

Lee, C., Gonzalez, S., & Payne, K. (2023). World Data System and Canadian CoreTrustSeal Cohort Needs Assessment Reviewer Instrument and Supplemental Information. World Data System.



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



World Data System –
International Technology Office



Ocean Networks Canada
University of Victoria
#100-2474 Arbutus Rd. Victoria, BC V8N 1V8
P 250 472 4527 | wds-ito.org

Executive Summary

From July 2022 to December 2022, the World Data System (WDS) International Technology (ITO) and International Program (IPO) Offices conducted a review of strategic plans and technical roadmaps of all current WDS members and the set of Canadian repositories that participated in the [Digital Research Alliance of Canada's CoreTrustSeal Certification Support and Funding Pilot](#) (Digital Research Alliance of Canada, 2022). The purpose of this review was to test out a new organizational assessment method designed to identify needs and challenges faced by the WDS and Canadian CoreTrustSeal (CTS) Cohort members. Additionally, incorporated into this assessment were criteria from the [Global Biodata Coalition \(GBC\)](#), which is an initiative created to support funders by identifying and prioritizing fundamental data resources.

This report serves as a supplemental file to the primary assessment, tentatively titled "Research Analysis: A World Data System and Canadian CoreTrustSeal Cohort Needs Assessment," which will be published at a later date. This report includes the reviewer instrument, a list of criteria used when reviewing strategic plans, technical roadmaps, and additional types of documentation by the WDS research team. Included are all 98 questions that comprised the reviewer instrument, including qualitative and quantitative questions designed to capture as much information as possible about the reviewed organization. Additionally, this report provides the Global Biodata Coalition (GBC) Indicators that were used in conjunction with additional prioritized criteria to create the reviewer instrument. Finally, incorporated into this report is a list of specific standards, policies, and practices mentioned and utilized by the reviewed WDS and Canadian CoreTrustSeal Cohort members. The primary intention behind this list is to demonstrate the varying standards and protocols used by the organizations included in this assessment. While also reflecting the type and variety of information the World Data System was able to glean from the review of strategic plans and technical roadmaps utilizing this organizational assessment method.

Table of Contents

Reviewer Instrument	5
Global Biodata Criteria and Our Evaluation	12
List of Specific Standards, Policies, and Practices Utilized by the 95 WDS and Canadian CTS Cohort Members Reviewed	20
References	27

Reviewer Instrument

1. Name of Reviewer
2. Date of Review
3. Name of repository/organization being reviewed
4. WDS Member or part of the Canadian Cohort (in the name of the file)
5. Document Name (from the file text)
6. Document Type
7. Number of pages:
8. Is this plan owned/maintained by the data repository or WDS member? Or is the plan owned/maintained by the host organization of the data repository or WDS member? (For example, data repository = University of Ottawa Dataverse and host organization = University of Ottawa Libraries. In this example, the strategic plan is owned by the host (libraries) not by the repository itself (Dataverse). In this case, use the host organization as the owner of the plan document. If in doubt, refer to the verified plans spreadsheet.)
9. Coverage dates for the plan. Please include the start date and the end date
10. Start date: date, month, year (or any combination - use what is in the document)
11. End date: date, month, year (or any combination - use what is in the document)
12. Domain: Domains served by the repository/organization (based on CRDC 2020 https://www.statcan.gc.ca/en/subjects/standard/crdc/2020v1/index)
13. Mission: Does the plan include a mission statement?

14. Mission: What is the Mission statement?
15. Vision: Does the plan include a vision?
16. Vision: What is the vision?
17. Goals: Does the plan include long-term goals? These should be specific and separate from their priorities or values. Can also be "long-term goals."
18. Goals: What are the goals? These should be specific and separate from their priorities or values.
19. Guiding Principles: Does the plan include guiding principles?
20. Guiding Principles: What are the guiding principles?
21. Strategic Priorities: Does the plan include strategic priorities?
22. Strategic Priorities: What are the strategic priorities?
23. Keywords of focus areas. <i>List these at the discretion of the reviewer as you identify focus areas (drones, in-situ, vocabulary services, visualizations, etc.):</i>
24. Activity areas/projects:
25. Review Stage:
26. Polar: Do they have a reference to Polar data/resources?
27. Polar: What are the polar references/activities?
28. Obstacles/known challenges:

29. Technical Requirements:
30. Strategies employed/implementation:
31. GEO: Does the plan reference GEO or any GEO activities? GEO = Group on Earth Observations
32. GEO: What are the GEO references/activities?
33. Status of the document?
34. Last update?
35. How extensive is the document?
36. Additional Links?
37. Does the document contain references to Indigenous groups, advancing the Calls to Action for indigenous groups, or building towards reconciliation with indigenous groups?
38. Copy and paste references to indigenous groups here:
39. Does the document contain references to advancing diversity, equity, and/or inclusion? (or related terms)
40. Copy and paste references to diversity, equity, and/or inclusion here:
41. Does the document reference FAIR, TRUST, CARE, OCAP?
42. Copy and paste references to FAIR, TRUST, CARE, OCAP:
43. Does the document reference sustainability and long-term funding?

44. Copy and paste references to sustainability and long-term funding:
45. Name of host organization:
46. Is the host organization an academic institution, government entity, commercial organization or other (please describe)?
47. Country:
48. Is the repository run by an international consortium?
49. Description of users or target audience:
50. Description of the size of staff/faculty:
51. KPI/Performance Metrics/Targets?
52. Description of KPI/Performance Metrics/Targets:
53. Partners including other key research services, funders, and in-kind service providers in the plan:
54. Does the document include planning/outreach for obtaining new partnerships for research services, funders, or in-kind service providers?
55. Copy and paste planning/outreach for obtaining new partnerships for research services, funders, or in-kind service providers?
56. Does the plan refer to the use of PIDs?
57. Description/text from use of PIDS:
58. What is the size of the holdings?

59. Does the document report uptime? If yes, report it as a KPI.
60. Does the document report web page response time? If yes, report it as a KPI.
61. Are there cyber-security references?
62. Details of cyber-security references:
63. Does the plan refer to standards for metadata?
64. Outline standards for Metadata:
65. Does the plan refer to standards for data?
66. Outline standards for data:
67. Does the plan refer to data curation?
68. Outline data curation:
69. Does the plan refer to the ability to provide provenance to data holdings?
70. Outline data provenance:
71. Does the plan refer to data quality?
72. Outline data quality:
73. Products: what data services are referred to in the plan?
74. Does the plan refer to data formats?

75. Outline data formats:
76. Does the plan refer to helpdesk/technical support?
77. Does the plan refer to mechanisms for user feedback?
78. Outline user feedback mechanism(s):
79. Does the plan refer to any training programs/training materials?
80. Outline training programs/training materials:
81. Does the plan refer to mechanisms for user outreach, engagement, and updates?
82. Outline mechanism(s) for user outreach, engagement, and updates:
83. What languages does the repository serve?
84. Does the plan refer to any governing or advisory bodies?
85. List governing or advisory bodies:
86. Does the plan refer to archival services?
87. List archival services:
88. Does the plan refer to data licensing?
89. List data licensing guidelines:
90. Does the plan refer to a user privacy policy?

91. List guidelines for the user privacy policy:
92. Does the plan refer to ethics resources?
93. List ethics resources:
94. Does the plan refer to any success stories?
95. List success stories:
96. Does the plan refer to a counterfactual? (Consequences if the data and the repository services did not exist or ceased to exist)
97. Paste the counterfactual from the plan here:
98. Please list any observations on the plan review that may be pertinent to this research. This may include feedback on how the instrument was used for this review or facts about the plan outside of the scope of the review instrument (outliers):

Global Biodata Criteria and Our Evaluation

Categories for Global Core Biodata Resource Indicators	Indicators for Global Core Biodata Resource Identification	Sub-Indicator When Applicable	GCBR Indicator Description	Our Interpretation of GCBR Criteria
1. Scientific focus and quality	1a. Deposition database and/or Knowledgebase		The distinction between deposition databases and knowledge bases is important contextual information. Deposition or archival databases receive and archive de novo data sets and well-structured metadata deposited by scientists. Knowledge Bases are added-value databases which are built on archival data and add substantial value through expert curation, annotation of metadata, sophisticated data processing and/or data integration. It is possible to be both a deposition database and a knowledge base.	N/A
	1b. Scope statement		Describes the scientific focus/domain covered by the resource, including	Domains served by the repository.

			factors such as nature of the primary data item (e.g., nucleic acid family, species occurrence, protein interaction, gene, biological sample, image, metabolite, dataset), experimental methods represented and characteristics that distinguish the resource from other biodata resources of related focus.	
	1c. Global dimension	1c i. Operation	Describes the global characteristics of the biodata resource with respect to: (i) the operation of the resource, e.g., is it run via an international consortium? Is it funded by agencies in different countries?	Academic Institution Nation International Umbrella / Parent / Host Organizations
		1c ii. Users/contributors	(ii) the geographical distribution of both its users (including the basis on which this is known) and contributors, e.g., does it take data submissions from a globally distributed range of nations?	Description of users or target audience
	1d. Staff effort			Description of the size of staff if provided.
2. Community	2a. Data resource		Describes usage of the biodata	KPIs / performance

	usage - quantitative data		resource over time, including access via a web browser in terms of number of visits/hits per month, unique IP addresses (which are a proxy for unique visitors) and sessions/page views, as well as data downloads per month in terms of hits/requests/requesters (unique IP addresses), data transfer volume, and global distribution of users.	metrics / targets? KPIs / Metrics / Targets Details
	2b. Usage in research as measured through data resource citation in the scientific literature		Describes citation in the scientific literature in terms of the frequency of citation of the resource name and the resource-specific data items (for example, via data accession numbers) over time.	Size of the data holdings (# or total bytes)
	2c. Citation of key publications describing the data resource		Key publications that describe the biodata resource, for example articles in the Nucleic Acid Research "Database" issue, with the numbers of times they have been cited in the scientific literature.	Document name Document link Document Full Citation Document Type
	2d. Connections to other data resources		Describes how the biodata resource is embedded in the ecosystem of biological, life science, and	Partners or other key research services mentioned in plan.

			biomedical data resources. Includes data exchanges between biodata resources and the direction and nature of those exchanges.	
3. Quality of Service	3a. Identifier use		Describes the system used to generate and implement persistent and unique identifiers, with identifier resolution services/mechanisms employed, if relevant.	Does the plan refer to the use of PIDs? If so, which ones and in what context
	3b. Data volume		Describes in quantitative terms the cumulative total number of entries, records processed, depositions, assays, etc., as relevant to the biodata resource, as well as and total data volume, in gigabytes etc., over time.	Any references to the size of the holdings?
	3c. Technical performance:	3c i. Uptime	Describes: i., percentage availability per month for a sample of indicative web pages and/or search functions over the past 12 months	Does the document report uptime? If so, report it as a KPI
		3c ii. Response times of key web pages	ii. response times for web pages that represent the typical web-based use case	Does the document report web page response time? If so, report it as a KPI

		3c iii. Back-up and disaster recovery	iii. the strategy for ensuring adequate back-up/disaster recovery for the data housed within the data resource.	Cyber-Security References? Cyber-Security Details
	3d. Use of standards		Describes community interoperability standards used for metadata and data housed in the biodata resource, and/or requested as part of a data submission protocol.	Does the plan refer to any standards for metadata and data? If so, which ones?
	3e. Documentation	3e i. Data Curation	Describes the provision of i. documentation of the data curation process/deposition workflow	Does the plan refer to data curation?
		3e ii. Provenance and Evidence	ii. links to the primary scientific literature for provenance of and/or evidence for data statements or biological context	Does the plan refer to the ability to add provenance to data holdings?
		3e iii. Quality Assurance	ii. versioning and/or evidence trails for modifications to datasets or data/metadata statements	Does the plan refer to data quality? If so, in what context
	3f. Data availability	3f i. Data sharing services	Describes the options in place for sharing data from the biodata resource in terms of: i. the services that facilitate sharing	Products: what data services are referred to in the plan?

		3f ii. Data sharing formats	ii. the formats in which the data is made available.	Does the plan refer to particular data formats? If so, which ones?
	3g. User support	3g i. Helpdesk	Describes support to users in terms of i. helpdesk provision/access	Does the plan refer to helpdesk support?
		3g ii. User feedback	ii. opportunities provided for user feedback	Does the plan refer to mechanisms for user feedback? If so, how?
		3g iii. Training	iii. training materials/opportunities	Does the plan refer to any training programs or materials? In what context?
		3g iv. Communications	iv. notification methods employed for updates and announcements	Does the plan refer to mechanisms for user outreach, engagement, and updates? If so, which ones?
		3g v. Language	v. language(s) in which the resource is made available	What languages does the repository serve?
4. Funding, governance, and legal infrastructure	4a. Funding		Describes funding secured by the biodata resource over the previous	Repository Funding Source.

			five years, current funding, and future committed funding.	
	4b. Scientific Advisory Board		Describes the composition, function and activities of the Scientific Advisory Board, or other equivalent advisory body.	Does the plan refer to any governing or advisory bodies?
	4c. Data preservation		Describes the planning by the biodata resource for data preservation in the long term.	Does the plan refer to archival services?
	4d. Open Science		Describes the licensing arrangements in place for the biodata resource that support open science.	Does the plan refer to data licensing? If so in what context?
	4e. Privacy policy		Describes the policy under which user personal data is collected and employed in the provision of the biodata resource services to the user and how security around that data is managed.	Does the plan refer to a user privacy policy? If so, in what context
5. Impact stories	5a. Accelerating science		Describes the ways in which the biodata resource has made specific contributions that have potentiated scientific progress or discovery or facilitated scientific methodologies. This may include, for example setting	Does the plan refer to any success stories? If so, provide a brief recap

			and promoting the use of metadata standards, actively promoting re-use of data or software, extending technical products.	
	5b. Counterfactual		Describes the consequences for the biodata resource ecosystem, the scientific community and primary scientific research were the biodata resource to cease to exist and its data, services, and functions not be replaced.	Does the plan refer to a counterfactual? If so, provide a brief recap
Source: (Global Biodata Coalition, 2022).				

List of Specific Standards, Policies, and Practices Utilized by the 95 WDS and Canadian CTS Cohort Members Reviewed

Metadata Standards			
Darwin Core	International Virtual Observatory Alliance Technical Specification	CMDI - Component Metadata Infrastructure	WDC-RRE Metadata Standard (V1.0)
Dublin Core	ISA - Tab	ESDIS Metadata Requirements - Base Reference of NASA Earth Science Data Products (423-RQMT-003)	NOAA WDS-Paleo tailor-designed JSON Metadata Format
DataCite Metadata Scheme	ISO 19139	UK GEMINI Metadata Standard	
DDI - Data Documentation Initiative	ISO 19115	Directory Interchange Format (DIF)	
Data Standards			
Water Quality Exchange (WGX) schema - uses EPA data standards	NASA Earth Science Data Systems (ESDS) Standards	Flow Cytometry Standard (FCS)	<i>INTERMAGNET Data Standards</i>
OSAP (Ontario Stream Assessment Protocols)	Open Geospatial Consortium (OGC) Standards	Chartered Institute of Archaeologists (CIfA) Regulations and Standards	<i>ILRS Data Standards</i>
DDI - Data Documentation Initiative	Climate and Forecast (CF) netCDF Conventions	<i>Odum Institute Data Archive Collection Development Policy</i>	<i>Reference to Open Data Standards</i>

<i>ImmPort Data Standards</i>	Human Immunology Project Consortium (HIPC) Standards	<i>WGMS Data Standards</i>	
Note: Organizations that have stated they use specific data standards that they have created for their organization are in <i>italics</i> .			
Persistent Identifiers			
DOI	Handle	URN: NBN	World Data Center - Renewable Resource and Environment (WDC-RRE) Organization-Specific Data Identifiers
PURL	<i>Flowing Waters Information Systems (FWIS) Organization-Specific Data Identifiers</i>	ORCID	
Note: Organizations that have stated they have specific persistent identifiers that are used by their organization are in <i>italics</i> .			
Data Licenses/ Conditions of Use			
Creative Commons - CC	BSB License	BGS Data License	CLARIN ACA+BY+NORED
Creative Commons - CC0	ADA Data License	NSSDC's Measures for Data Management and Open Sharing	Creative Commons - CC-BNC-SA-4.0
<i>Public Domain License</i>	<i>Licensed Data Access</i>	Norwegian License for Open Government Data (NLOD)	<i>Standard Access - Login Required</i>
<i>Non-Commercial Use</i>	<i>Restricted Data Access</i>	Creative Commons - CC BY 4.0 (Attribution 4.0 International)	INTERMAGNET Conditions of Use
GeoNB Open Government License (New Brunswick, Canada)	ImmPort User Agreement	Creative Commons - BY-NC-SA 3.0 and 4.0	The DOBES Code of Conduct v.2
Apache License 2.0	NASA Earth Science Data and Information Policy	NERC Data Licensing and Charging Policy	CLARIN RES+BY+NORED
Open Government License (Canada)			

Data Formats			
<i>Images</i>	ASCII	VOTable	FLAC
<i>Audiovisual data</i>	FITS	PDBx/mmCIF	AIFF
<i>Statistical data</i>	SAV	BMRB Core Archive NMR-STAR format	WAVE
<i>Plain text</i>	DVL	PDBML/XML	IMFV1.23 GIN Dissemination Format
<i>Source code</i>	RSF	UTF-8	IMFV2.83 Satellite Transmission Format
JSON	DFT	Open Office formats (.doc, .docx, .xls, .ppt)	INTERMAGNET Archive Format
CSV	MCS	ESRI Grid	ImagCDF
HDF4 and HDF5	VEL	Shapefile	IYF INTERMAGNET Year-Mean File
NetCDF	IAGA2002 Data Exchange Format	KMZ/KML	IBF INTERMAGNET Baseline File
TXT	RINEX	DXF/DWG	GIF
GeoTIFF	JPG	AVI	PDF
M-JPEG2000			
Note: Formats that do not indicate a digital file format, such as " <i>Images</i> " are in <i>italics</i> .			
Mechanisms for Ensuring Data Quality			
Metadata inspection and checks	Calibration and validation checks of research instruments	Quality control checks to ensure data is in the preferred formats	Filtering of data using automated approaches such as machine learning to eliminate spam or

			irrelevant information from datasets
File auditing	Quality control checks to inspect deposited data for completion, inconsistencies, errors, and accuracy	Automated quality control checks	Quality control checks to ensure that data meets required data standards
Require that any issues found during quality control checks be addressed before data can be added to the repositories data holdings			
Cyber-Security Mechanisms			
Blockchain Technology	Virtual Private Networks (VPN)	Firewalls, anti-virus programs and spam-filters	Backups and version control of all software programs
ISO-27002 International Standard on Information Systems Security	Backup storage of data holdings in a separate secure location	Monitoring of network activity	Employment of contingency/ disaster preparedness plans for data
Password use, encryption, and secure logins	Daily, weekly or monthly backups of data holdings	Two-Factor Authentication	Regular security scans of systems and servers to detect system vulnerabilities
Mechanisms for Data Curation			
Creation and review of metadata	Preservation and protection of original datasets	Ensuring that appropriate credit is given to the authors/creators of deposited data, and data citations are documented with the dataset when shared	Creating thumbnails for data documentation entries
Assigning PIDs	Quality Control Checks	Ensure that any changes to original datasets are only made on copies of the dataset	Checking of data licenses to ensure data is distributed appropriately

Linking research data to other relevant sources	Enhancement of metadata, data, or additional documentation for completion and/or accuracy	Digitization of physical data products to ensure preservation of data at risk of degradation	
Conversion of original datasets into alternative formats	Secure storage of deposited data before public data sharing	Ensuring data holdings are accessible and available to data users, including ensuring data is searchable using keyword vocabularies	
Mechanisms for Community Outreach			
Workshops	Public education events	Recruitment of underrepresented minorities for employment	Publication of annual meetings to provide updates to the designated community of the activities of the organization
Lectures	Conferences	Making use of social media and online networking tools	Engage with students at the post-secondary level
Events	Presentations	Provide educational resources for elementary and secondary education	Working Groups
International professional activities, such as collaborating with international colleagues	Scientific Publications	Partnering and collaborating with Indigenous communities	Newsletters distributed to the user community
Scholarship Programs	Recruitment of early-career researchers for employment	User training sessions and courses	
Mechanisms for Obtaining New Partnerships			

Making direct contact with potential partners for funding or research opportunities	Development of projects that enable cross-disciplinary collaboration with other organizations	Provide information through programs, publications and other communication channels that inform potential partners and stakeholders about the value of the organization, and the activities and work they are involved in	Develop frameworks and models to aid in the facilitation of attracting external funding sources
Develop programs for facilitating collaboration with potential partners in the academic community	Participation in international initiatives and collaboratives	Collaborate with academic institutions to promote education and the importance of research databases	Utilize working and interest groups to help align organizational priorities with the priorities of industry or NGO organizations
Participation in outreach and community engagement	Joint involvement in research projects with like-minded organizations to address joint goals or research questions	Engage in communication and dialogue with the designated community, to form community partnerships for research opportunities	
Mechanisms for User Feedback			
Annual user satisfaction surveys and publication of the results	Comments section on organization's website	Obtaining user feedback through social media channels	Direct email contact for user feedback
Helpdesks	Contact form or feedback form on organization's website	Virtual meetings with users	Online forums
Obtaining user feedback directly at events or conferences			
Ethical Practices Employed			

Adherence to privacy and confidentiality laws, such as HIPPA	Require ethics approval for all data depositors	Use of a privacy or ethics policy that outlines how the organization handles research procedures, confidentiality, data disclosure or non-compliance	Adherence to the principles of fully and openly accessible data
No disclosure of personal information related to a dataset or project without consent	Require that employees sign and adhere to a declaration of confidentiality agreement	Ensure the security and protection of sensitive or restricted datasets by guaranteeing they are not publicly available, through the use of confidentiality agreements or data license agreements	Utilize checks for plagiarized data
Removal of all personal or non-scientific data when a dataset is deposited into the repository	No storage of personal or sensitive information	Adherence to all copyright and intellectual property laws and regulations	Ensure that users are informed of the ethical procedures of the organization when using their data products through the use of a data policy, user agreement, or terms of use document

References

Digital Research Alliance of Canada. (2022). CoreTrustSeal Certification Support Cohort & Funding. *Digital Research Alliance of Canada*.
<https://alliancecan.ca/en/coretrustseal-certification-support-cohort-funding>

Global Biodata Coalition. (2022). Global Core Biodata Resources: Concept and Selection Process. <https://doi.org/10.5281/zenodo.5845116>