

Sufficiency of Self-Documenting Data Exchange Standards

Felimon Gayanilo

Harte Research Institute, Texas A&M University Corpus Christi, Corpus Christi, TX 78412-5869 (j)Email: fgayanilo@tamu.edu

Matthew K. Howard

Department of Oceanography, Texas A&M University, College Station, TX 77843-3146 (Email: mkhoward@tamu.edu)

Abstract

Scientific data warehouses aggregate data and information to facilitate data use and reuse. Data provenance metadata describes the data's origins and any transformations that may have occurred as data move from raw forms to the processed forms available from the warehouse. Such metadata establishes usability and credibility for the data by providing attributions for people, institutions, methods and more. Metadata standards specify what auxiliary pieces of information are required and how to encode them. Many metadata standards exist and it may take more than one to completely express all that needs to be expressed and often, there is considerable overlap and duplication. The International Organization for Standardization (ISO) produced the ISO 191xx series of geographic information standards which are widely-used to document data that has strong geospatial content. The Open Geospatial Consortium (OGC) developed widely-used standards for publishing heterogeneous sensor data of all types for consumption primarily by automated data system. The U.S. Integrated Ocean Observing System (IOOS) has as its goal the automated and largely-unattended transfer of data from sensors to desktops. To support interoperability among various regional data systems and the Federal backbone, IOOS adopted two standards from the OGC Sensor Web Enablement (SWE) framework: the Sensor Observations Service (SOS) web interface and the Observation & Measurements (O&M) encodings. The SOS defines what parameters are available, their temporal coverage, the methods used to collect them, and the data originator. Because much of the IOOS data has a strong geospatial content, efforts are made to automate transformation of OGC SOS to the ISO 191 family of standards. This poster evaluates sufficiency of the OGC SOS inputs to *Digital Object Identifier* (DOI) requirements, ISO 191xx and information needed in data landing pages.

Discussion and Conclusion

Extensible Stylesheet Language (XSL) can be used (see below for an example from NOAA IOOS Program) to map and transform elements of the OGC SOS to ISO 191xx. A qualitative assessment of this mapping using the *Gulf of Mexico Research Initiative Information and Data Cooperative's* (GRIIDC) core elements (see table), using the following indicators was performed.

- - SOS provide more details than required
- - equivalent SOS data is available
- - alternate SOS data is available but may be insufficient or questionable
- - no equivalent SOS data

This exercise reveals the quality of the possible transformation from the self-describing data services OGC SOS to an ISO 191xx standard. An assessment was also made for mapping requirements to mint a *Digital Object Identifier* (DOI) and for required elements of a landing page such as those displayed using GRIIDC's evolving landing page format (e.g. <https://data.gulfresearchinitiative.org/data/R1.x134.073.0014/>) and NOAA *National Oceanographic Data Center* (NODC) dataset landing pages.

The resulting table indicates that self-documenting data service standards, such as the *OGC SOS GetCapabilities*, is sufficient to be in compliance to ISO standard. However, the table also reveals sections of the OGC SOS standard that are insufficient or may require review and revisions to completely meet the requirements of ISO. Most recently, the ISO has started to introduce changes to the ISO 191 family and it may affect the results presented in this document.

The results of this mapping study also leads to discussions as to how organizations, who serves near-real time data via the OGC SOS and other self-documenting standards such as *SensorML*, should submit data to national data centers for archiving or to data centers such as GRIIDC.

References and Links of Interest

IOOS Working XSL for SOS -> ISO
https://github.com/ioos/metadataTransformations/blob/master/SOS/SOS2ISO_ML.xsl

OGC SOS Specifications
<http://www.opengeospatial.org/standards/sos>

NOAA EDM: SOS->ISO Discussions
https://geo-ide.noaa.gov/wiki/index.php?title=Sensor_Observation_Service_Capabilities_and_ISO_Metadata

GCOOS SOS GetCapabilities in ISO
http://www.ngdc.noaa.gov/docucomp/page?xml=NOAA/IOOS/GCOOS/iso/xml/data.gcoos.org:8080_52nSOS_sos_kvpxml&view=xml2text/xml-to-text-ISO

Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC)
<https://data.gulfresearchinitiative.org/>

GCOOS Data Portal SOS
<http://data.gcoos.org:8080/52nSOS/>

Acknowledgement

This publication was made possible in part by a grant from BP/The Gulf of Mexico Research Initiative (GOMRI), and in part by NOAA Integrated Ocean Observing System (IOOS) and the Gulf of Mexico Coastal and Ocean Observing System (GCOOS).

ISO 191xx -> Open Geospatial Consortium (OGC) SOS Map Check List and Quality Assessment

GRIIDC Core Elements	ISO 191xx Path	OGC SOS 2.0 Equivalent	DOI Rqd.	ISO Rqd.	Ideal for Landing Page	Comments
General						
File identifier	//gmd:ML_Metadata/gmd:fileIdentifier	//gmd:fileIdentifier		●	●	Unique identifier remains essential in file management
Dataset Contact						
Name	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:individualName/*	//ows:ServiceProvider/ows:ServiceContact/ows:IndividualName(text)=1		●	●	19115-1 redirects CL_ResponsibileParty to CL_Responsibility and all mapping remains the same
Organization	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:organisationName/*	//ows:ServiceProvider/ows:ProviderName(text)=1		●		
Position	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:positionName/*	//ows:ServiceProvider/ows:ServiceContact/ows:PositionName(text)=1		●	●	
Role	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:role/gmd:CL_RoleCode(text)=1	(default =PointOfContact)		●	●	
Telephone Number	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:phone/gmd:CL_Telephone/gmd:voice/*	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Phone/ows:Voice/text)=1		●	●	OGC's OpenGIS Web Service (ows) Common is used in SOS schema (http://schemas.opengis.net/ows/2.0/ows15Subset.xsd is not for ISO 19115-1 but presence an equivalent)
Address	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:deliveryPoint/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:DeliveryPoint(text)=1		●	●	
City	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:city/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:DeliveryPoint/ows:City(text)=1		●	●	
State/Province/Region	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:administrativeArea/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:AdministrativeArea(text)=1		●	●	
ZIP/Postal Code	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:postalCode/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:PostalCode(text)=1		●	●	
Country	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:country/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:Country(text)=1		●	●	
Email	//gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:electronicMailAddress/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:ElectronicMailAddress(text)=1		●	●	
Dataset Information						
Title	//gmd:MD_DataIdentification/gmd:citation/gmd:CL_Citation/gmd:title/gco:CharacterString(text)=1	//ows:ServiceIdentification/ows:Title(text)=1		●	●	
Date	//gmd:MD_DataIdentification/gmd:citation/gmd:CL_Citation/gmd:date/gmd:CL_Date/gmd:date/gco:Date(text)=1	//sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering/gml:name(text=urn:ioos:network*all)/sos:time/gml:TimePeriod/gml:endPosition(text)=1		●	●	This value is expected to change for monitoring systems with continuous assimilation of data
Abstract	//gmd:MD_DataIdentification/gmd:citation/gmd:abstract/gco:CharacterString(text)=1	//sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering[gml:id=urn_..._ioos_network_.../ows:ServiceProvider/ows:ProviderName(text)=1_...all]/sos:observedProperty(*) and //sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering[gml:id=urn_..._ioos_network_.../ows:ServiceProvider/ows:ProviderName(text)=1_...all]/sos:featureOfInterest(*)		●	●	ISO defines abstract as a summary of the content but do provide detailed elements describing parameters and their definitions
Purpose	//gmd:MD_DataIdentification/gmd:citation/gmd:purpose/gco:CharacterString(text)=1	//ows:ServiceIdentification/ows:Abstract(text)=1		●	●	ISO defines Purpose as intent of the dataset which is defined in abstract
Suppl. Info. (Parameters and Units)	//gmd:MD_DataIdentification/gmd:supplementalInformation/gco:CharacterString(text)=1	//sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering/gml:name(text=urn:ioos:network*all)/sos:observedProperty/*		●	●	Many SOS implementations, utilizes a registry to define the parameters and served via MMI
Keywords						
Theme Keywords	//gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:type/gmd:MD_KeywordTypeCode/@codeListValue=theme and //gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyWordS(text)=1	//ows:ServiceIdentification/ows:Keywords/*		●	●	SOS do not make distinction between the Theme and Place keywords
Place Keywords	//gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:type/gmd:MD_KeywordTypeCode/@codeListValue=place and //gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyWordS(text)=1	//ows:ServiceIdentification/ows:Keywords/*		●	●	
Topic Category Keywords	//gmd:MD_DataIdentification/gmd:topicCategory/gmd:MD_TopicCategoryCode/*	(default=occean Climatology/Meteorology/Atmosphere)		●	●	In SOS, topical keywords appropriate for most ocean observing systems is 'ocean'
Data Extent						
Spatial Extent	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:geographicElement/gmd:EX_GeographicBoundingBox/gmd:westBoundLongitude and //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:geographicElement/gmd:EX_GeographicBoundingBox/gmd:eastBoundLongitude and //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:geographicElement/gmd:EX_GeographicBoundingBox/gmd:northBoundLatitude and //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:geographicElement/gmd:EX_GeographicBoundingBox/gmd:southBoundLatitude and/or //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:geographicElement/gmd:EX_BoundingPolygon/*	//sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering[gml:id=urn_..._ioos_network_.../ows:ServiceProvider/ows:ProviderName(text)=1_...all]/gml:boundary/gml:Envelope/gml:lowerCorner(text)=1) and //sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering[gml:id=urn_..._ioos_network_.../ows:ServiceProvider/ows:ProviderName(text)=1_...all]/gml:boundary/gml:Envelope/gml:upperCorner(text)=1)		●	●	
Temporal description	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:description(text)=1	(default= ground condition)		●	●	ISO is either 'ground condition' or 'modelled period'. In SOS, it is always assumed to be set to 'ground condition'
Start date	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:beginPosition(text)=1	//sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering[gml:id=urn_..._ioos_network_.../ows:ServiceProvider/ows:ProviderName(text)=1_...all]/sos:time/gml:TimePeriod/gml:beginPosition(text)=1		●	●	SOS provide start date for all platforms
End date	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:endPosition(text)=1	//sos:Contents/sos:ObservationOfferingList/sos:ObservationOffering[gml:id=urn_..._ioos_network_.../ows:ServiceProvider/ows:ProviderName(text)=1_...all]/sos:time/gml:TimePeriod/gml:endPosition(text)=1		●	●	SOS provide end date for all platforms
Distribution Information						
Data file size	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:distributorTransferOptions/gmd:transferSize/gco:Real(text)=1	(not applicable)		●	●	SOS does not provide a file size as it is dynamic in nature
URL	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:distributorTransferOptions/gmd:gmd:online/gmd:CL_OnlineResource/gmd:linkage/gmd:URL(text)=1	(//ows:OperationsMetadata/ows:Operations/ows:DCP/ows:HTTP/ows:Get/*) or (//ows:OperationsMetadata/ows:Operations/ows:DCP/ows:HTTP/ows:Post/*)		●	●	SOS provides dynamic extraction of data (spatial and temporal) and setting a temporal range to indicate time of submission is not an option
Distribution Contact						
Organization	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:organisationName/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ProviderName(text)=1		●	●	OGC SOS does not differentiate dataset point of contact information and distribution contact details. This a deficiency as the details for the point of contact may not necessarily represent the organization distributing the data.
Role	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:positionName/gco:CharacterString(text)=1	(default =PointOfContact)		●	●	
Telephone number	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:phone/gmd:CL_Telephone/gmd:voice/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Phone/ows:Voice/text)=1		●	●	
Address	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:deliveryPoint/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:DeliveryPoint(text)=1		●	●	
City	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:city/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:DeliveryPoint/ows:City(text)=1		●	●	
State/Province/Region	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:administrativeArea/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:AdministrativeArea(text)=1		●	●	
ZIP/Postal Code	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:postalCode/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:PostalCode(text)=1		●	●	
Country	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:country/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:Country(text)=1		●	●	
email	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:electronicMailAddress/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:ElectronicMailAddress(text)=1		●	●	
URL	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor/gmd:distributorContact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:onlineResource/gmd:linkage/gmd:URL/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ProviderSite(text)=1		●	●	
Metadata Contact						
Name	//gmd:contact/gmd:CL_ResponsibileParty/gmd:individualName/*	//ows:ServiceProvider/ows:ServiceContact/ows:IndividualName(text)=1		●	●	OGC SOS does not differentiate dataset point of contact to that of metadata point of contact. This is a deficiency of OGC SOS as the metadata point of contact may not necessarily be the same person as the portal's point of contact.
Organization	//gmd:contact/gmd:CL_ResponsibileParty/gmd:organisationName/*	//ows:ServiceProvider/ows:ProviderName(text)=1		●	●	
Role	//gmd:contact/gmd:CL_ResponsibileParty/gmd:role/gmd:CL_RoleCode(text)=1	(default =PointOfContact)		●	●	
Telephone Number	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:phone/gmd:CL_Telephone/gmd:voice/*	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Phone/ows:Voice/text)=1		●	●	
Address	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:deliveryPoint/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:DeliveryPoint(text)=1		●	●	
City	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:city/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:DeliveryPoint/ows:City(text)=1		●	●	
State/Province/Region	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:administrativeArea/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:AdministrativeArea(text)=1		●	●	
ZIP/Postal Code	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:postalCode/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:PostalCode(text)=1		●	●	
Country	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:country/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:Country(text)=1		●	●	
Email	//gmd:contact/gmd:CL_ResponsibileParty/gmd:contactInfo/gmd:CL_Contact/gmd:address/gmd:CL_Address/gmd:electronicMailAddress/gco:CharacterString(text)=1	//ows:ServiceProvider/ows:ServiceContact/ows:ContactInfo/ows:Address/ows:ElectronicMailAddress(text)=1		●	●	