OGC Maritime Limits and Boundaries Pilot

Engineering Report

Publication Date: 2020-07-29

Approval Date: 2020-06-23

Submission Date: 2020-05-27

Reference number of this document: OGC 20-013r4

Reference URL for this document: http://www.opengis.net/doc/PER/MLB-pilot

Category: OGC Public Engineering Report

Editor: Jonathan Pritchard

Title: OGC Maritime Limits and Boundaries Pilot: Engineering Report

OGC Public Engineering Report

COPYRIGHT

WARNING

LICENSE AGREEMENT

Permission is hereby granted by the Open Geospatial Consortium, ("Licensor"), free of charge and subject to the terms set forth below, to any person obtaining a copy of this Intellectual Property and any associated documentation, to deal in the Intellectual Property without restriction (except as set forth below), including without limitation the rights to implement, use, copy, modify, merge, publish, distribute, and/or sublicense copies of the Intellectual Property, and to permit persons to whom the Intellectual Property is furnished to do so, provided that all copyright notices on the intellectual property are retained intact and that each person to whom the Intellectual Property is furnished agrees to the terms of this Agreement.

If you modify the Intellectual Property, all copies of the modified Intellectual Property must include, in addition to the above copyright notice, a notice that the Intellectual Property includes modifications that have not been approved or adopted by LICENSOR.

THIS LICENSE IS A COPYRIGHT LICENSE ONLY, AND DOES NOT CONVEY ANY RIGHTS UNDER ANY PATENTS THAT MAY BE IN FORCE ANYWHERE IN THE WORLD. THE INTELLECTUAL PROPERTY IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE DO NOT WARRANT THAT THE FUNCTIONS CONTAINED IN THE INTELLECTUAL PROPERTY WILL MEET YOUR REQUIREMENTS OR THAT THE OPERATION OF THE INTELLECTUAL PROPERTY WILL BE UNINTERRUPTED OR ERROR FREE. ANY USE OF THE INTELLECTUAL PROPERTY SHALL BE MADE ENTIRELY AT THE USER'S OWN RISK. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR ANY CONTRIBUTOR OF INTELLECTUAL PROPERTY RIGHTS TO THE INTELLECTUAL PROPERTY BE LIABLE FOR ANY CLAIM, OR ANY DIRECT, SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM ANY ALLEGED INFRINGEMENT OR ANY LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR UNDER ANY OTHER LEGAL THEORY, ARISING OUT OF OR IN CONNECTION WITH THE IMPLEMENTATION, USE, COMMERCIALIZATION OR PERFORMANCE OF THIS INTELLECTUAL PROPERTY.

This license is effective until terminated. You may terminate it at any time by destroying the Intellectual Property together with all copies in any form. The license will also terminate if you fail to comply with any term or condition of this Agreement. Except as provided in the following sentence, no such termination of this license shall require the termination of any third party end-user sublicense to the Intellectual Property which is in force as of the date of notice of such termination. In addition, should the Intellectual Property, or the operation of the Intellectual Property, infringe, or in LICENSOR's sole opinion be likely to infringe, any patent, copyright, trademark or other right of a third party, you agree that LICENSOR, in its sole discretion, may terminate this license without any compensation or liability to you, your licensees or any other party. You agree upon termination of any kind to destroy or cause to be destroyed the Intellectual Property together with all copies in any form, whether held by you or by any third party.

Except as contained in this notice, the name of LICENSOR or of any other holder of a copyright in all or part of the Intellectual Property shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Intellectual Property without prior written authorization of LICENSOR or such copyright holder. LICENSOR is and shall at all times be the sole entity that may authorize you or any third party to use certification marks, trademarks or other special designations to indicate compliance with any LICENSOR standards or specifications.

This Agreement is governed by the laws of the Commonwealth of Massachusetts. The application to this Agreement of the United Nations Convention on Contracts for the International Sale of Goods is hereby expressly excluded. In the event any provision of this Agreement shall be deemed unenforceable, void or invalid, such provision shall be modified so as to make it valid and enforceable, and as so modified the entire Agreement shall remain in full force and effect. No decision, action or inaction by LICENSOR shall be construed to be a waiver of any rights or remedies available to it.

None of the Intellectual Property or underlying information or technology may be downloaded or otherwise exported or reexported in violation of U.S. export laws and regulations. In addition, you are responsible for complying with any local laws in your jurisdiction which may impact your right to import, export or use the

Intellectual Property, and you represent that you have complied with any regulations or registration procedures required by applicable law to make this license enforceable.

Table of Contents

Chapter 1. Introduction

The Open Geospatial Consortium (OGC) Maritime Limits and Boundaries (MLB) Pilot is a project designed to advance the implementation of a data model, architecture, and prototypes for use with the creation, management, integration, dissemination, and onward use of official data for maritime baselines, limits, zones, and boundaries. It also aims to enable, by extension, the "reuse" of such data within far broader geo-regulation contexts.

The Pilot has emerged from activities of the sponsoring organizations; namely the Canadian Hydrographic Service, Natural Resources Canada, Geoscience Australia, and the United Kingdom Hydrographic Office, who have participated in the development of the International Hydrographic Organization (IHO) standard IHO S-121. The OGC has a general Memorandum of Understanding with the IHO covering collaboration on development of standards and matters of common interest. Under the MOU the OGC seeks to bridge gaps relating to marine standards with other groups and domains.

The original reasons why the Pilot came into being were to build a detailed implementation of the S-121 standard, to test productionization of its many different facets, and to demonstrate its potential in a broader context. The Pilot was designed to build on the progress made in baselining the S-121 standard and to test the practicality and potential of its features. The Pilot has been driven by its sponsoring organizations and built advanced implementations alongside participating industry representatives using OGC standards, architecture and best practices.

These activities use the scope of S-121 as a foundation but seek to engage the broader georegulation community and contribute to the evolving standardization efforts in these broader domains. The tangible advantages of taking a lead in this evolution is the prevention of rework and the harmonization of features between S-121 (as defined in the IHO Geospatial registry) with those from other domains which may be similar in nature but not specifically marine in nature.

The Pilot is being carried out by the OGC employing its own internal defined approval process for deliverables including a review and approval cycle with interested Domain Working Groups within the OGC Technical Committee. Delivery to sponsors, by the OGC, will then take place to the sponsoring organizations to meet their stated requirements as defined in the Call For Participation. The IHO S-121 Project Team, a subgroup of the IHOs S-100 working group, have been involved through the reciprocal participation of multiple project members and sponsors and their respective organizations and have all contributed to the defined Pilot goals. The sponsoring organizations have been responsible for defining the requirements and deliverables and are the prime customers of the Pilot.

This is the first Pilot to have been run in this way within the OGC and is unique in many ways. It has taken a standard defined by another global standards development organization (SDO), and integrated it with an OGC baseline, opening up its modeled content to the full range of OGC web services and architectures.

There are two goals of the S-121 standard, first: to have an open, international, coordinate-based, representation of maritime boundaries and their associated rights, restrictions, and responsibilities (RRRs) that are authoritative and easy to interpret. The second goal is to facilitate the deposit of maritime boundary claims with the Division for Ocean Affairs and the Law of the Sea (DOALOS) in

a spatial coordinate-based format that fulfills legal requirements under UNCLOS. The Division for Ocean Affairs and the Law of the Sea (DOALOS), of the Office of Legal Affairs of the United Nations, approached the International Hydrographic Organization with a request to identify appropriate technical standards. After consultations, the S-121 project team was formed by Member States of the International Hydrographic Organization and, accordingly, S-121 was created.

As an IHO Standard within the S-100 framework, and by sharing ISO19152 components, the standard contributes to both the goals of open standards and facilitating depository obligations as well as contributing to broader goals in the Marine Spatial Data Infrastructure (MSDI) community and the IHOs allied organizations including the evolving field of marine georegulation.

The MLB Pilot initiative thus begins with an implementation model, encoding, architecture, and prototypes for sharing Maritime Limits and Boundaries (MLBs) between digital systems compatible with the structures of S-121. Techniques and samples of natural extensions to the encoded data are also proposed to contribute to the longer-term roadmap. This was accomplished by enabling reuse of data through OGC standard web services and enabling use of ISO19152 conceptual model and structures in the related domains.

Chapter 2. Executive Summary

This document comprises the Engineering Report (ER) and documents the activities under Phase 1 and Phase 2 of the OGC Maritime Limits and Boundaries (MLB) Pilot.

This Engineering Report details the activities undertaken by participants in the pilot, the data supplied, transformed, and used to demonstrate the pilot's objectives, and the results of the various interoperability tests performed within the pilot. Also documented here are the various outputs from the pilot activities, where they are directed, and where updates or clarifications are recommended to external standards or other elements of the broader ecosystem.

The ER documents the journey taken within the project, from its origins as an architectural statement of intent, through to the definition of its core elements (the GML application schema which lends a common language to the data) and the implementation of that schema within COTS, open source and bespoke software elements. The engineering report also summarizes the outputs from the process, any unresolved issues, and potential enhancements for the future.

2.1. Overview of Pilot Activities

The MLB Pilot is about data, the standards used to define it, its content definition, and its path from originating producers to a variety of end users.

The main activities of the Pilot were the following.

- Productionizing S-121 data production by creating an initial Application Schema which enables the encoding of S-121 data in Geographic Markup Language (GML), a dialect of XML specifically designed for encoding geospatial data. The existing S-121 standard does not contain an encoding of this nature, only a text encoding designed for its core use case.

- Transformation of representative datasets into the GML encoding defined.

- Onward use of the GML data within several OGC Standard web services and its transformation to a text form specific to S-121 use cases.

- Extension and incorporation of S-121 data alongside data from other domains.

The flow of this engineering report reflects the flow of data through the various artefacts, tools, and systems defined above and used in the Pilot. The idea is to illustrate each of the components in detail as major sections within the ER. The diagram below thus shows the data flow through the Pilot as defined in detail by the architectural model in the call for participation.

The pilot's activities are a significant achievement in several ways:

- It is the first systematic implementation of S-121 (and extensions to it) using a variety of different organizations' data in different forms;

- It exercises all the different facets of the S-121 standards including geospatial feature realization and defining LADM structures in conformance with S-121's implementation of ISO19152;

- It successfully blends IHO feature content and encoding with OGC standards and web services

broadening dramatically the number of onward uses and participants for the data content; and

• It shows how multiple domains can be combined in GML encoded content.
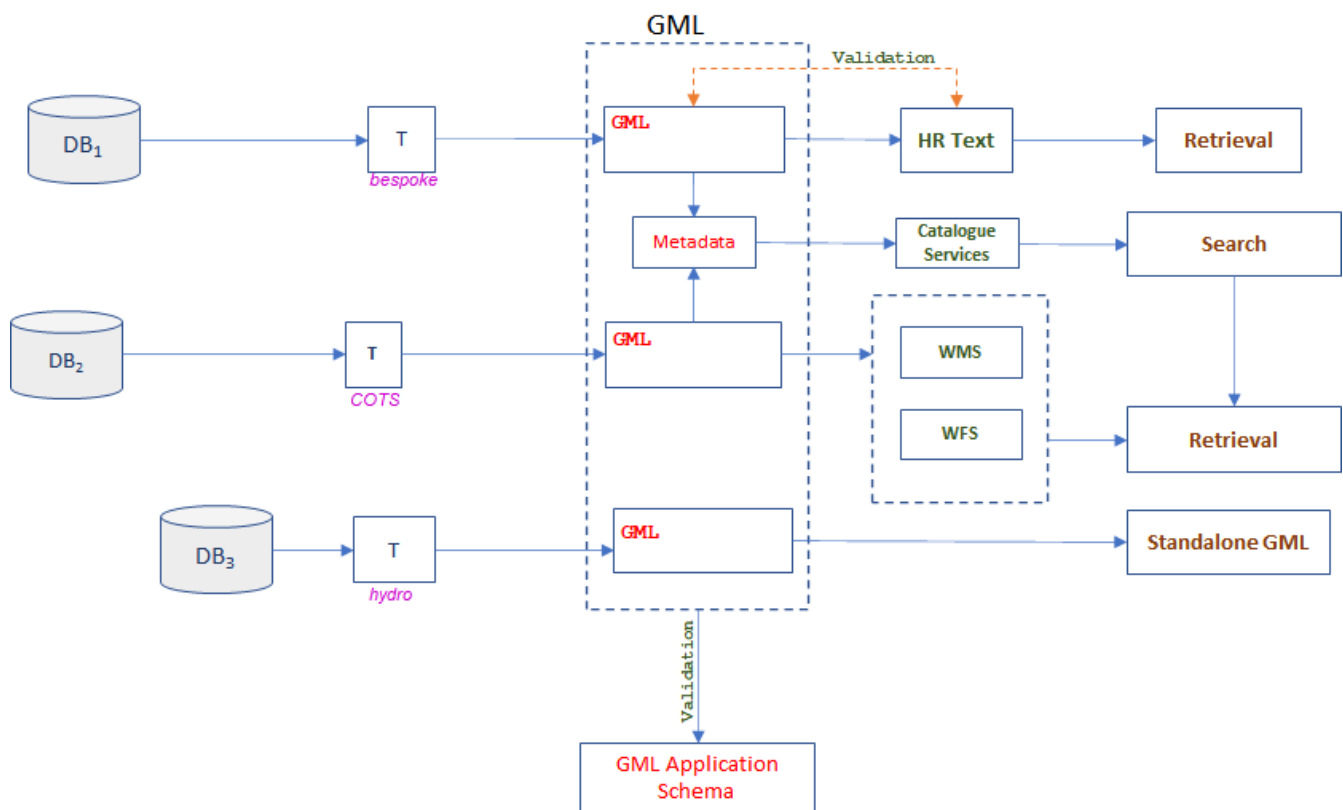
Figure 1. Data flow through the various components of the Pilot

Reading out the data flow from left to right, the diagram shows the following.

• Data, within the pilot this is MLB data in various "database" forms, conforming to internally adopted standards, models, and potentially expressing different aspects of definitions contained in the S-121 standard.

• Transformation (the "T" boxes) into GML conformant with the Pilot GML Application Schema via either COTS tools (from both the hydrographic domain and mainstream GIS) and bespoke software created during the Pilot. GML is validated, again using either bespoke or COTS packages against the schema and open standards.

• Onward use for:

  ◦ Direct use of GML via COTS and open source GIS packages;

  ◦ Population of CSW services from metadata built specifically for the certain datasets; and

  ◦ Promulgation via core OGC Web services, WFS and WMS.

• Production of Human Readable Text (including validation against source GML data) from the transformed GML.

• Validation of the HR Text against the original GML.

2.2. Selection of Key Findings.

The Pilot produced a number of outputs, a summary of which are presented here. The key

deliverables were schemas which allow the S-121 model contained in the IHO standard to be expressed and validated in a precise, technical format. Mechanisms for extending the domain of the core schema were developed, tested, and demonstrated and data supplied by sponsor organizations was transformed into conformant GML to test a range of onward uses.

2.2.1. GML Application Schema

- The IHO S-121 product specification has been published at version 1.0.0. This represents a significant step forward for its development and for the community driving it. The product specification included a draft S-100 feature catalogue which expresses the core features, attributes, relationships, and metadata in precise IHO terms. This feature catalogue was baselined at the same time as the product specification and forms the basis of the creation of the S-121 GML Application Schema within the Pilot.

- The construction of a GML Application Schema has been successful.

    ◦ The GML Application Schema has produced a viable encoding for data conforming to the S-121 model, and for related data by extension. The Schema has been baselined for both IHO S-121 content and for an initial extended vocabulary.

    ◦ The schema has used the IHO S-100 GML Profile, is conformant with that profile, and interoperable with the software tools and systems used by the participants within the project from both IHO specific applications and broader GIS tools, both proprietary/COTS and Open Source.

- The Application Schema developed required several design decisions to be made which could arguably have been more tightly specified within the S-100 GML Profile. Since S-100 defines a comprehensive XML feature catalogue structure it would be feasible to map that to a GML Application Schema in a much more specific way in order to take out the ambiguity of the schema development. Particular areas where the profile could benefit from this extra specificity have been noted in the report, particularly:

    ◦ Dataset aggregation and definition;

    ◦ Names of complex types;

    ◦ Dataset layout e.g., the order of geometry, information types, feature types;

    ◦ Conventions for enumeration value inclusion;

    ◦ Conventions for relationship definition; and

    ◦ Guidance on norms for multi-lingual support within the application schemas.

- Additionally, the GML Profile contained in Part 10b of S-100 may benefit from some modernization/update, particularly the definitions of DSID/DSSI content and identifiers. These, and other items, will be discussed with the S-100WG at a suitable upcoming meeting.

- The Schema has been thoroughly tested with sponsor data including complex examples such as the LADM structures, dual positional representation, and extensions into other domains. The data transformation technologies and their subsequent onward usage of data within OGC services demonstrate practically the utility of an "implementation first" approach. This approach focused on practical, rather than theoretical initiatives and working practices and concentrated on producing working prototypes rather than theoretical models and documents.

- S-121 is a complex standard which implements both S-100 and ISO 19152 structures and norms.

The Pilot demonstrated the various features of the S-121 model contained in the product specification as far as resources allowed and the GML Schema constructed represents the extents of that testing process. As S-121's implementations grow, the Pilot activities will form a valuable source of guidance and knowledge for those wishing to express and use MLB data for a variety of purposes.

2.2.2. ISO 19152 and LADM Implementation

- The S-121 standard includes discrete packages representing geospatial features, ISO 19115 metadata for their sources, and ISO 19152 structures for representing LADM structures attached to them. Participant data expressing combinations of all of these were constructed in the project and this provides an overview of the challenges and complexities with S-121. As a GML encoding of the data, the project has encapsulated the data supplied and shown example implementations, but these are just a start. The ability to encode these structures in a representative way requires encoding guidance. The IHO Data Classification and Encoding Guide for S-121 is currently at version 1.0.0 and as it matures it will be a valuable manual for populating S-121 data. The project has shown that sponsors have a great disparity in how their own data structure aligns with the definitions within the standard and the amount of transformation necessary to achieve truly conformant S-121 GML.

- Project activities focused within Phase 2 centered around exercising all the features of the GML Application Schema. In particular, the implementation of the LADM package entails a complex series of information types and relationships with them implementing the relevant ISO 19152 provisions. This complexity led to a thorough investigation of all aspects of implementation which can feed into recommended guidance for future implementers of IHO S-121.

- The ISO 19152 package within the S-121 product specification provides a broad conceptual framework for Rights, Responsibilities, and Restrictions which is, in the S-121 context, restricted to the domain of UNCLOS features. The project has developed a generic extension mechanism for the GML Schema which allows co-existence of features and relationships between them which will enable the LADM aspects to be extended to arbitrary domains such as Hydrocarbons and Fishing Limits. This mechanism will be promoted further post-project closure and the modular structure arrived at should be used as an input both to IHO and the developing ISO community.

2.2.3. Web Services

- Data transformation into GML enables its use within web services such as WFS, WMS, and for search and retrieval, via CSW along with structured metadata population and processing. These have all been demonstrated by project activities and, where possible, included in detail within this ER.

- These web services are able to process the complex GML produced and also deal with the S-121 specific elements, such as complex multi-lingual naming and textual position encoding.

- This process of opening data via OGC web services is the foundation for MLB data to sit alongside other geo-regulation datasets within current and future web services. OGC's evolution of web services into API based interfaces would be consistent with the approach taken by the Pilot.

2.2.4. Human Readable Text

- Production of a multi-lingual Human Readable Text rendering of GML Data has been achieved to a high level of detail. This development process has shown the viability of transformation from a sponsor-specific database into textual form which meets requirements from stakeholders and is a basis for taking forward with the S-121 community.

- Validation of the HR Text outputs against source data has been achieved to a high degree of detail as well. This contributes to the general area of validation alongside standard XML/GML validation by providing a reconciliation of produced HR text features with their original source.

- The HR text outputs can be easily manipulated into multiple other forms such as pdf and HTML as required.

2.3. Background

2.3.1. UNCLOS

The United Nations Convention on the Law of the Sea (UNCLOS), also called the Law of the Sea Convention or the Law of the Sea treaty, is the international agreement that resulted from the third United Nations Conference on the Law of the Sea (UNCLOS III), which took place between 1973 and 1982.

The Law of the Sea Convention defines the rights and responsibilities of nations with respect to their use of the world's oceans, establishing guidelines for businesses, the environment, and the management of marine natural resources. The Convention, concluded in 1982, replaced the quad-treaty 1958 Convention on the High Seas. UNCLOS came into force in 1994, a year after Guyana became the 60th nation to ratify the treaty. As of June 2016, 167 countries and the European Union are parties to the convention.

2.3.2. IHO and S-121 Background

Coastal States, under articles 16, 47, 75, 76, and 84 Convention, are required to deposit with the Secretary-General of the United Nations charts showing: straight baselines, including closing lines of mouths of rivers and bays, and archipelagic baselines; the outer limits, as well as lines of delimitation between States with adjacent or opposite coasts, of the territorial sea (including roadsteads, article 12); the contiguous zone; and the exclusive economic zone and the continental shelf. Alternatively, the lists of geographical coordinates of points, specifying the geodetic datum, may be substituted.

In its resolutions 49/28 of 6 December 1994 and 52/26 of 26 November 1997, the General Assembly requested the Secretary-General to establish appropriate facilities, as required by the Convention, for the deposit by States of maps, charts, and geographic coordinates concerning national maritime zones and establish a system for their recording and publicity and to develop and maintain [such] facilities for the deposit by States of charts and geographical coordinates concerning maritime zones, including lines of delimitation, and to give due publicity thereto, as required by article 16, paragraph 2, article 47, paragraph 9, article 75, paragraph 2, article 76, paragraph 9 and article 84, paragraph 2, of the Convention. The Division for Ocean Affairs and the Law of the Sea, Office of Legal Affairs of the United Nations is the unit of the Secretariat which performs these depositary

functions on behalf of the Secretary-General, as part of an integrated program on the law of the sea and ocean affairs, distinct from the usual depositary functions of the Secretary-General in respect to multilateral treaties.

Subsequently, in its resolution 59/24 of 17 November 2004, the General Assembly requested the Secretary-General to improve the existing geographic information system for the deposit by States of charts and geographical coordinates concerning maritime zones, including lines of delimitation in particular by implementing, in cooperation with relevant international organizations technical standards for the collection, storage and dissemination of the information deposited, in order to ensure compatibility among the Geographic Information System, electronic nautical charts, and other systems developed by these organizations. Recent General Assembly resolutions have noted ongoing efforts in this regard.

In 2015 the UN passed resolution A/RES/70 nominating a set of 17 Sustainable Development Goals (SDGs). Target C of SDG 14 encourages states to "Enhance the conservation and sustainable use of oceans and their resources by implementing international law as reflected in UNCLOS." The indicator 14.C.1 is "the number of countries making progress in ratifying, accepting and implementing through legal, policy and institutional frameworks, ocean-related instruments that implement international law, as reflected in the United Nation Convention on the Law of the Sea, for the conservation and sustainable use of the oceans and their resources." This work is required in order to assess indicator 14.2.1: "Proportion of national exclusive economic zones managed using ecosystem-based approaches."

In addition, the General Assembly, in its annual resolutions on Oceans and the law of the sea, calls upon States Parties to the Convention to fulfill their deposit obligations. Most recently, General Assembly resolution 71/257 of 23 December 2016 calls upon States Parties to the Convention that have not yet done so to deposit with the Secretary-General charts or lists of geographical coordinates, as provided for in the Convention, preferably using the generally accepted and most recent geodetic datums (para. 6).

To facilitate the implementation of the Secretary-General's depositary functions, coastal States are encouraged to deposit the following information, as a minimum.

- Geographic coordinates of points in decimal degrees on the straight baselines and archipelagic baselines in common global geodetic datum such as WGS 84, accompanied, as appropriate, by the relevant national legislation.

- Geographic coordinates of points in decimal degrees on the outer limits as well as lines of delimitation between States with adjacent or opposite coasts for the following maritime zones: territorial sea (including roadsteads); contiguous zone; exclusive economic zone and continental shelf, in common global geodetic datum such as WGS 84, accompanied, as appropriate, by the relevant national legislation. Ideally the points defining the outer limits should be close enough to each other to ensure that the line formed by connecting the points with geodesic lines accurately reflects the outer limit of the maritime zone(s). States Parties are also encouraged to identify in the deposit the points that are part of an international boundary. States Parties are further encouraged to accompany such deposits with the relevant national legislation and/or with references to relevant international treaties.

Accordingly, the Division for Ocean Affairs and the Law of the Sea, of the Office of Legal Affairs of the United Nations, the unit entrusted with carrying out these responsibilities on behalf of the

Secretary-General, approached the International Hydrographic Organization with a request to identifying appropriate technical standards. After consultations, the S-121 project team was formed by Member States of the International Hydrographic Organization.

S-121 is an open access method of providing digital representation of Maritime Limits and Boundaries (MLBs). Maritime limits and boundaries are constructs used to define maritime zones for nations around the world. With the United Nations Convention of the Law of the Sea (UNCLOS), they can form an international legal foundation of the marine domain. S-121 represents an essential extension of the International Hydrographic organization S-100 for the administration of the marine domain.

In December 2016, the International Hydrographic organization (IHO) distributed the initial document describing the standard at the United Nations headquarters in New York. This meeting began the international effort to define the core features of the S-121 standard.

There are two goals of the S-121 standard, first: to have an open, international, coordinate-based, representation of maritime boundaries. The second goal is to facilitate the deposit of maritime boundary claims with the Division for Ocean Affairs and the Law of the Sea (DOALOS) in a spatial coordinate-based format that fulfills legal requirements under UNCLOS.

2.3.3. Broader Uses of the S-121 Standard

Although S-121 is defined primarily within the narrow constraints of a subset of the UNCLOS Convention, the broader conceptual area of marine georegulation and marine spatial planning is an emerging area important to the marine geospatial community. Wikipedia defines Marine Spatial Planning (MSP) as "a process that brings together multiple users of the ocean – including energy, industry, government, conservation and recreation – to make informed and coordinated decisions about how to use marine resources sustainably. MSP generally uses maps to create a more comprehensive picture of a marine area – identifying where and how an ocean area is being used and what natural resources and habitat exist," although the definition of exactly what these uses are vary from region to region.

The areas of jurisdiction defined under UNCLOS, particularly the Territorial Sea Area, Contiguous Zone, and Exclusive Economic Zone, together with associated limits and baselines play an important role in marine georegulation and marine spatial planning as they define the geographic areas on which any legal regime is predicated. In contrast to UNCLOS, Marine Spatial Planning is a term which has many interpretations and definitions but its link with the UNCLOS-defined elements is present in all of them. The S-121 model, therefore, has a central part in these activities and the data used to express these concepts. The S-121 model also contains important interfaces to compatible structures which model ISO19152, the Land Administration Domain Model and a comprehensive source information repository on a per-feature basis which can enable the use of S-121 data in integrated marine cadastral applications conformant with international standards.

In order to increase interoperability with other domains of expertise, both in the georegulation field and marine spatial planning, phase 2 of the Pilot has considered the concept of "extending" the framework on which the S-121 model is realized in the GML Application Schema. This extension does not seek to modify or change the essential S-121 data but instead allow other domains to be co-located with its data in a single dataset. The mechanism for doing this has been evolved during Phase 2 of the project and is presented along with the other project outputs and using sponsor data

to illustrate the potential. The potential here is not limited purely to the marine domain.

The implementation of ISO19152 within S-121 and the extension mechanism formulated within the Phase 2 of the project does not require co-existence with marine domain data. Thus, terrestrial and land cadastral data can equally be represented alongside marine domain data in combined datasets.

2.3.4. ISO 19152

The ISO 19152 standard is a conceptual model which represents the domain of land administration, specifically rights, responsibilities and restrictions (RRR) affecting land and water. It has two overall goals:

- An extensible basis for the development of land (and, by implication, water) administration systems; and
- The enablement of communication between interested parties those structures contained within the model based on a shared vocabulary.

2.3.5. IHO, S-100 and S-121

The S-121 standard has been under development for some time and this section of the ER contains a broad overview of its origin, aims, and features relevant to the project. As documented in the previous section of this ER, the development of S-121 has been coordinated by an IHO project team under the S-100 working group (S-100 WG). IHO S-100, the framework standard used to define S-121 has been under development since the publication of its predecessor, IHO S-57 in 2001. IHO S-57, the current standard for the encoding of Electronic Navigational Charts under the SOLAS convention is a vector based standard developed specifically for the purpose of encoding charts for the purpose of safe navigation but S-100, conceived shortly afterwards represented a much bigger step forward.

S-100 aimed to overcome many of the perceived shortcomings of the newly released S-57 standard and was defined as a much broader standard where a framework of ISO-like structures was defined leaving the details of content and encoding to the authors of specific product specifications which would then sit alongside the main framework. S-100 therefore had the following goals:

- Production of a standard in close alignment with the ISO19100 framework;
- A framework standard which defines content through individual product specifications;
- A separation of data content from its representation in encodings;
- Fully machine-readable standards for both feature content and portrayal;
- The location within a registry located at the IHO, of features, their attributes and metadata; and
- The facility to update feature content and portrayal by end user systems.
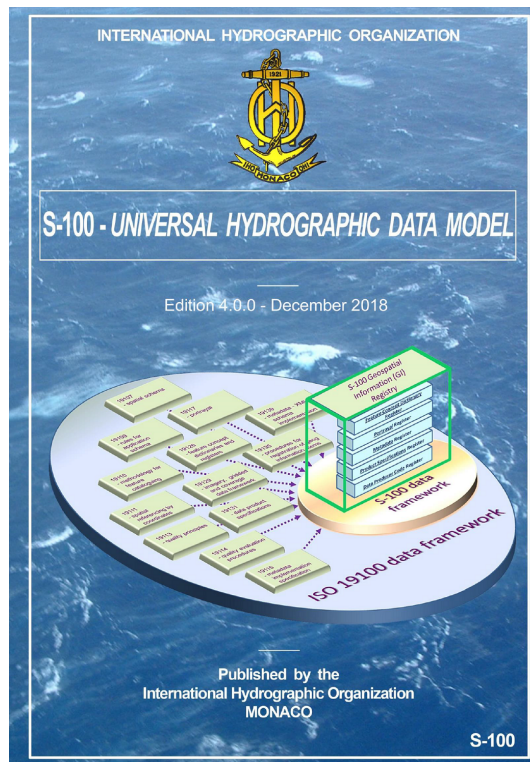
Figure 2. IHO S-100 Edition 4.0.0

S-100 is currently at edition 4.0.0 with edition 5.0.0 in preparation. As part of the development of S-100 two encodings for vector feature data were defined, one was ISO8211, a compact binary format used predominantly in the encoding of Vector ENC charts, and the other was the S-100 GML profile, Part 10b. The S-100 GML profile, a subset of GML (ISO 19136), was designed to support the encoding of simple vector datasets was developed by the UK around 2013 and incorporated into S-100 as Part 10b shortly thereafter. The profile has been used by various project teams within the IHOs Nautical Publications working group as well as other S-100 based product specification developments.

S-100 does not define data content but only provides a toolkit for its definition. Data content is defined within S-100 product specifications. These product specifications detail how data is defined, aggregated, and exchanged along with metadata such as coverage, CRS, geometry and encoding details. The detailed definitions of feature, attribute, and associations for an individual product specification are contained within its feature catalogue and also lodged in the IHOs geospatial registry; an ISO compliant registry where all S-100 products' definitions, concepts and details are kept and reconciled by a dedicated team.

Prior to the commencement of the project, the IHO S-121 features defined in the (then) current version of the product specification were added to the IHO geospatial registry and used to create a version 1.0.0 feature catalogue. This feature catalogue was then used by the MLB Pilot as input for the GML Application Schema. A set of proposed modifications to the S-121 feature catalogue will be output from the project and used to propose changes within the IHO geospatial registry. These are few in number but reflect a large amount of productionized testing during the project's lifespan.

It is important to note, then, that the development of IHO S-121 has continued during the project's progress and this is reflected in updates to the existing S-121 standard. In the main, however, there are few changes to the underlying S-121 model and feature content and the Pilot remains the most up to date reflection of the S-121 model constructed to date and certainly the only one with concrete reference implementations in software from multiple participating vendors.

2.3.6. OGC Standards, Architecture and the Challenges for the Marine Community

OGC's main interaction with the marine community is mainly through its established Marine Domain Working Group (MDWG), formed within the IHO/OGC memorandum of understanding (MOU). This group was formed to address the gaps in the OGC framework within the marine domain. For many years the IHO has run an MSDI community through its MSDIWG and the OGC MDWG works closely with the MSDIWG to cross-fertilize ideas and outline where opportunities exist to improve the ecosystem for the benefit of stakeholders.

The OGC MDWG has a focus on the S-100 framework and its broader integration into the OGC community. This process is likely to take some time and the MLB Pilot is a significant step in exploring such interfaces. IHO and OGC standards have many common elements and both derive largely from overarching ISO standards. The practicalities of their use alongside each other are not always well defined however and the project has sought to explore these practical steps as much as possible.

The OGC is currently revising, from scratch, part of its core architectural standards baseline. The publication in 2019 of the OGC Features API, (originally "WFS3") is a significant step forwards in this respect and signals OGC's intent to revise much of its standards base around content-neutral API based architectures. This will be supplemented in the near future by more standards covering common reporting standards (CRS) and common query languages and is likely to have a significant impact on how organizations like IHO engineer systems which seek to be interoperable within OGC standards. The move to an API based interconnecting system of standards clarifies the dividing line between content, its expression within a technical encapsulation (its encoding) and the transport of that data to its ultimate destination.
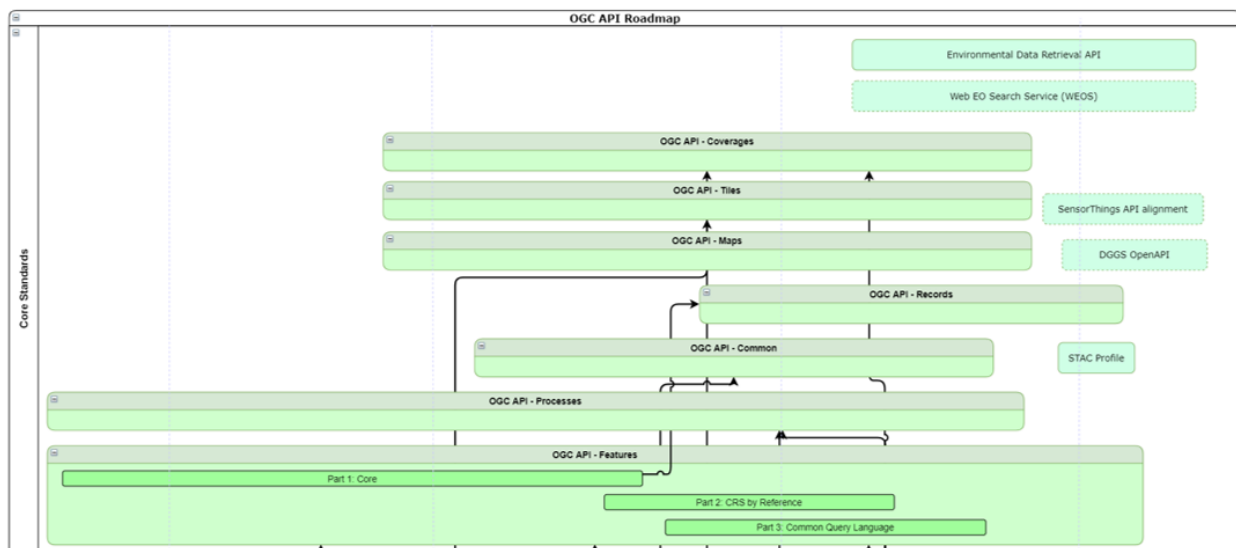
Figure 3. OGC Architectural Roadmap (from https://www.ogc.org/roadmap)

The IHO, similarly, is instigating a major drive towards implementation of IHO S-100. The standard, many years in the making, is already key to many activities and initiatives in the marine domain. Of particular note is the IMO's eNavigation initiative, for which S-100 forms the common maritime data structure (CMDS). The IHO is embarking on a push to get S-100 accepted as equivalent for carriage of charts and publications under the global SOLAS convention, a large undertaking and one which will embed its use in live vessel navigation for many years. As a dynamic standard

which relies on the creation of product specifications S-100 is wholly dependent on its ability to be interoperable with other standards frameworks and to remain current with those frameworks.

This project is intended to contribute positively to both the OGCs interface with the IHO and to assist the productionization of S-121 as part of the S-100 implementation roadmap.

2.4. Document Contributor Contact Points

All questions regarding this document should be directed to the editor or the contributors:

Contacts

Name	Organization	Role
Jonathan Pritchard	IIC Technologies	Editor
Dean Hintz	Safe Software	Contributing Author
Rob Thomas	OGC	Contributing Author

2.5. Foreword

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

Chapter 3. OGC MLB Pilot Context

This section describes the OGC Pilot in more detail. The section details the structure of the Pilot then describes the major use cases implemented by the participants during the Pilot's development.

3.1. Overall Objectives and Goals

The overall objective of the OGC MLB Pilot is to contribute to the implementation of the S-121 standard (as expressed in the Call for Participation in the MLB Pilot).

With the assistance and sponsorship from the major organizations Geoscience Australia, the Canadian Hydrographic Service and Natural Resources Canada, together with the United Kingdom Hydrographic Office, the pilot was designed to enable the implementation of the S-121 data model and architecture and to implement "operational prototypes to support the creation, management, integration, dissemination and onward use of official data for maritime baselines, limits, zones, and boundaries." Prototypes implemented as part of the prototype will demonstrate the ability to support:

- Country level publication, as a national obligation, of maritime baselines, limits and boundaries;
- Standards based interoperability between supplier, user and partner, within and across governments, public and commercial users; and
- Facilitating strategic awareness and operational decision making in the maritime environment supporting good governance and effective and efficient operation.

3.2. Project Structure

Ultimately the goal of the OGC MLB Pilot is to demonstrate a comprehensive, end-to-end chain for data relating to Maritime Limits and Boundaries unified under the S-121 standard. The project aims to show how data from sponsor organizations, arranged within a variety of forms specific to those organizations, can be encoded in GML and used to satisfy a variety of representative use cases both within and beyond the immediate use cases identified within the standard itself. A key aim of the Pilot was to put the S-121 data alongside data from other domains and to investigate the technical means to achieve such co-existence efficiently.

To achieve these goals, the project has been split into two phases. This Engineering Report details the outcomes and results for both phases of the project and treats them as a single output.

The structure of the OGC Pilot is depicted by the Architecture diagrams for phase 1 and phase 2 shown. Phase 1 of the project is centered around the construction and basic testing of a GML Schema representing S-121 data, in conformance with the IHO S-100 framework's GML profile and Phase 2 focuses on the use of the GML encoding to satisfy a number of further use cases, specifically:

- Transformation of sponsor data into S-121 conformant GML which exhaustively tests the main features of the schema;

- Display and Interrogation of GML data within COTS and Open Source GIS tools;

- Implementation of OGC web services for cataloging and distribution of S-121 data; and

- Extension of the Schema content to other domains.

In accordance with the summary here, this ER deals with the deliverables in a similar fashion. The development and testing of the GML Application Schema is documented first followed by detailed descriptions of each of the implementation use cases. The final sections of this report detail conclusions and outputs from the project focused on main external stakeholder groups and communities within this domain.

The two diagrams shown here represent the core architecture of the Pilot and how the data flow was originally conceived. The individual deliverables from the Pilot are documented in detail in this ER.
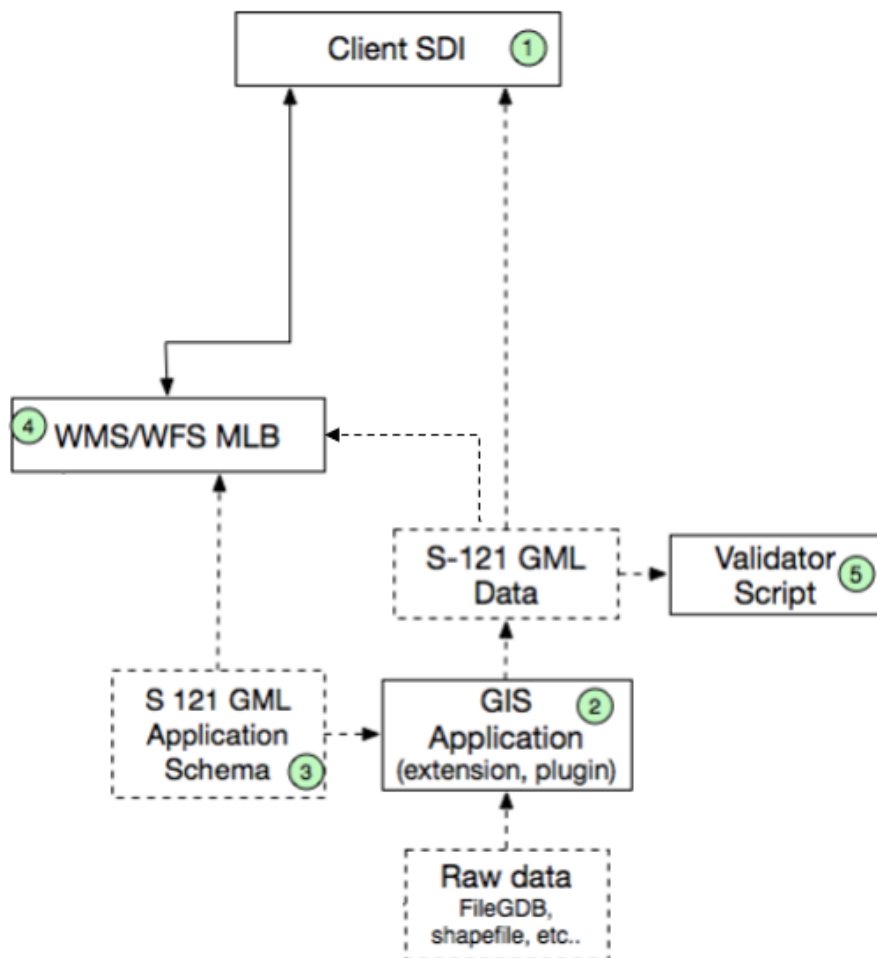
Figure 4. Project Architecture - Phase 1
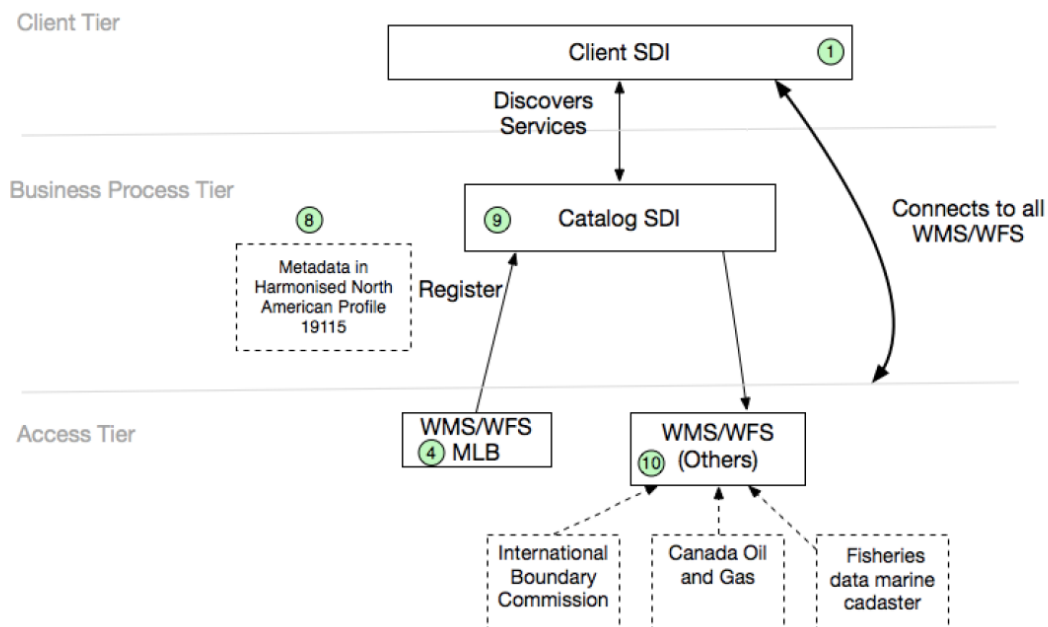
Figure 5. Project Architecture - Phase 2

The details of these goals have evolved and been refined as the project has progressed and the resources have been concentrated on areas where the most benefit can be obtained compatible with the overall aims of the project.

In parallel with the Pilot activities, the IHO project team has continued its work in baselining version 1.0.0 of the S-121 product specification. This came to its conclusion at the end of August 2019 with a final review of comments on the product specification and production of a finalized documentation set (including the model from which the feature catalogue and GML schema is derived).

It is not the intention of the OGC Pilot to remain 100% in step with developments within the IHO project team but as far as possible this has been done through cross-representation by members of the project team. The aim within the Pilot is very much to demonstrate the "art of the possible" in terms of interoperability, to show how data in the S-121 form can be exchanged and promulgated through web services for core S-121 use cases as well as broader use and as a foundation for many other marine activities. A major output of the project is a list of outputs to other groups, standards bodies and individual standards which may benefit from the experience of the participants in drawing the many threads concerned together.

In order to provide a better context for the content of this report the project outputs are preceded by sections detailing the core use cases and current end user perspectives with respect to S-121 data.

3.3. Core Use Cases - The User Perspective

This section describes current end user needs and describes issues within the current environment encountered by both users and intermediaries in the acquisition and processing of MLB data. S-121, as an IHO standard defines use cases as well but those presented here represent a broader set. In the OGC context the aim is to provide a universal encoding for S-121 modeled data which can then

be used to enable a range of different capabilities, some of which are generic such as the provision of standard OGC web services, and some specific to IHO S-121 - the main one being the output of "Human Readable Text" transformed from S-121 format GML.

In this section a more detailed description and analysis of the primary use cases addressed by the Pilot are described.

3.3.1. Legislative - Production of "Human Readable Text" Outputs

In contrast to broader usage of MLB data (primarily within the context of MSDI-type use cases), the user perspective is best described in relation to the main use cases as documented within the main S-121 Product Specification - these are the exchange of data between interested parties and the creation of datasets to support statutory deposit under the UNCLOS convention. Both use cases are dealt with extensively within the reference given but it should be recognized that they stand in stark contrast to the use cases described in the previous section. Exchange between parties can agree beforehand on data content, attribution and tend to be for well-defined uses such as preparation of documentation and inspection to support treaty negotiation and publication under very controlled conditions.

The subject of depository obligations under UNCLOS is, again, one which is well defined as a use case and although no mandated format exists (nor is ever likely to) the S-121 standard has been developed in response to direct requests from the UN General Assembly so the boundaries of the use cases are well known. There are numerous aspects to these complex position areas which are not explored in this engineering report further.

One of the most fundamental use cases, and the one most specific to IHO S-121 is the production of "Human Readable Text" output from S-121 GML data. this is best exemplified by the following excerpt from Article 16 of UNCLOS.

Article 16
Charts and lists of geographical coordinates

1. The baselines for measuring the breadth of the territorial sea determined in accordance with articles 7, 9 and 10, or the limits derived therefrom, and the lines of delimitation drawn in accordance with articles 12 and 15 shall be shown on charts of a scale or scales adequate for ascertaining their position. Alternatively, a list of geographical coordinates of points, specifying the geodetic datum, may be substituted.
2. The coastal State shall give due publicity to such charts or lists of geographical coordinates and shall deposit a copy of each such chart or list with the Secretary-General of the United Nations.

Figure 6. UNCLOS excerpt, Article 16

The production of a "list of geographical coordinates" satisfying the UNCLOS due publicity requirements is a key use case, these are currently deposited with the UN's Department for Ocean Affairs and the Law of the Sea (DOALOS). The form of such lists is not mandated, nor fixed in any way but the availability of an open standard which can support states in meeting the requirements of the convention is one of the key outputs of IHO S-121 itself.

In this case, the data content of IHO S-121 is transformed into a list of coordinates containing only those features required to satisfy the statutory obligations (whereas, of course S-121 itself defines a broad set of feature and information types) - this also includes a "round trip" requirement to take the list of coordinates and produce S-121 format data from it. The current edition of the S-121 v1.0.0

product specification contains an Annex detailing a textual encoding format which addresses this use case. During the development of IHO S-121, liaison with UN DOALOS was undertaken to arrive at a form of text which would satisfy their requirements for deposit and this is currently embedded within the S-121 Annex B. The key characteristic is its unambiguous interpretation by human inspection, the so called "Human Readable Text" encoding.

The OGC project more broadly redefines this use case to include production of multi-format lists of coordinates suitable for use within domestic legislation. The output of such lists of coordinates satisfies use cases for internal production for sponsor organizations and is a common requirement within the community. An example is shown of such a table from the Canadian domestic Territorial Sea Geographical Coordinates Order.

AREA 2

Southeast and East Newfoundland

	Column I	Column II	Column III	Column IV
	Locality	Latitude	Longitude	C.H.S. Chart
1	Lamaline Shag Rock	46°50'21"N.	55°49'30"W.	4016
2	Shag Rock	46°50'17"N.	55°44'51"W.	4016
3	Offer Island	46°51'21"N.	55°37'25"W.	4016
4	Ferryland Head	46°52'16"N.	55°23'04"W.	4016
5	Cape St. Marys	46°49'14"N.	54°11'54"W.	4016
6	S.W. Bull and Cow	46°46'34"N.	54°06'13"W.	4016
7	Shoal Point	46°36'50"N.	53°35'12"W.	4016
8	Cape Freels	46°36'42"N.	53°33'30"W.	4016
9	Mistaken Point	46°37'29"N.	53°09'48"W.	4016
10	Cripple Rock Point	46°38'29"N.	53°06'08"W.	4016
11	Big Johns Point	46°38'36"N.	53°05'51"W.	4016
12	Unnamed Peninsula	46°38'45"N.	53°05'32"W.	4016
13	Cape Race	46°39'30"N.	53°04'18"W.	4016
14	Cape Ballard	46°47'16"N.	52°56'52"W.	4016
15	Renews Rocks	46°52'39"N.	52°54'00"W.	4016
16	Bear Cove Point	46°56'27"N.	52°53'33"W.	4016
17	Hare's Ears	47°00'57"N.	52°51'13"W.	4016
18	Cape Broyle	47°03'47"N.	52°51'08"W.	4016
19	Great Island	47°10'57"N.	52°48'32"W.	4016
20	Green Island	47°14'10"N.	52°46'45"W.	4016
21	Bull Head	47°18'34"N.	52°44'51"W.	4016

Figure 7. List of Coordinates, Canada Territorial Sea Geographical Coordinates Order

This document, enacted in legislation, shows an example of the lists of coordinates required by this use case. Such lists, in contrast to the UNCLOS / DOALOS case do not need to be in a specific flat text format so other forms such as HTML, PDF and others can be implemented. The use case is to replicate the content of such lists from GML encoded S-121 data.

The aim of the project is to show how such transformations can be achieved and tangible results based on sponsor data.

3.3.2. Non-legislative Use Cases.

These use cases should be differentiated from the use cases of e.g., statutory deposit obligations under UNCLOS. This section considers some of these broader user needs as the scope of the Pilot is not restricted to the strict S-121 product specification use cases.

Broadly, customer needs are currently predicated on whether a position of interest or some other reference position is inside or outside a Maritime Limit or Boundary, so the user can understand what legislation or rules apply to their individual needs. As the majority of end users use a Geographical Information System (GIS), the simplest way to do this is to apply a spatial query as to whether they are inside or outside a polygon (surface) feature. Therefore the optimum structure for many MLB use cases data is polygons directly from source, by direct reference to source (whether by API or reference to authoritative data) or where the ability to create unambiguous polygons from the MLB source data using automated tools is possible.

Given that there are difficulties with contiguous geometry in some legislation definitive creation of a polygon may not be appropriate in all cases. The next best thing, in these cases would be a linear feature that the users could then manually assess. In all these cases the specification of "position" requires a correct interpretation and understanding of the difference between a geospatial position (as an intrinsic element of the data and plotted onscreen) and the legislative Textual position which in most cases is not *exactly* the same thing. The Human Readable text sections of this ER look at this in more detail and show how a dual representation of position can be used to construct datasets which carry, and express, this dual representation in an unambiguous way.

The most important elements for consumers of MLB datasets in these broader contexts are:

- Data Consistency of format, structure, and content;

- Documented data sources;

- Provenance of the data; and

- Data Interoperability, that is, the ability to mix MLB data with data from other sources, domains, and forms to enable geospatial processing, querying and analysis without the need to resort to manual processing.

It is these requirements which the activities of the Pilot therefore addresses. The establishment of a GML Schema which fully expresses the elements of the model and their relationships is key. It allows the model to be expressed unambiguously within a digital format which acts as a carrier (in our case, Geographic Markup Language (GML)). The complete encoding of MLB data enables it to be used to exchange between participants and between systems which can process its content. At this point OGC web services are enabled allowing a range of pre-defined functionality to be brought into play for end users.

The final piece in the jigsaw puzzle for these non-legislative use cases is the interoperability of the MLB data with other non-MLB datasets. This is partly achieved using GML which enjoys a certain amount of interoperability with other datasets and partly by the more formalized extension mechanism developed in Phase 2 of the Pilot. This allows for extensions of the feature model in S-121 and the ISO 19152 structures, textual position representation, and feature Naming to domains other than UNCLOS features. This extension mechanisms is not a data "format" as such, more a way of re-using S-121 data within combined GML Datasets forming a tight integration with the other

domains thus expressed.

3.4. Specific Challenges in Working with MLB Data

This section describes some of the challenges unique to the domain of MLB. These have been authored by Pilot participants and are intended to provide context and supporting background information for the Pilot's activities. Many participants in the Pilot work with MLB datasets routinely and these challenges reflect current experience across a wide range of samples.

3.4.1. Data Sources

Of fundamental importance with any spatial dataset is knowing where it has come from, who created it, the coordinate reference system, and how up-to-date it is. Without this information, end users may use the data incorrectly and make incorrect decisions.

Generally, this type of information is supplied as metadata that accompanies the datasets, however there can be use in putting some of this information in attributes on the datasets. With MLB datasets this can be a particular advantage as the provenance of data can be legally complex with no room for ambiguity. The dual representation of position is an example where the geospatial position is provided "for illustrative purposes only" whereas the textual position forms the normative, legal reference to the source.

3.4.2. Data Structure and Format

Currently, many producers of MLB data create datasets reflective of their own internal requirements and capabilities. These formats vary from GIS standard formats, e.g., KML, to proprietary formats such as Esri File Geodatabase or Shapefile. Consistency of format between different producers makes it easier to manage the processing of datasets and the bespoke nature of their supply often results in changes to format or content/structure without warning requiring reprocessing and testing for exchange or broader use between providers.

Other measures which can assist with more streamlined distribution and supply chains are:

- Consistent file naming conventions (this could be factored into the S-121 product specification); and

- The extension of consistent content and structure to web services where adopted by suppliers. This is a major challenge and an opportunity for the OGC community in providing a way of expressing standards such as IHO S-121 as web services without loss of clarity or provenance.

3.4.3. Data Attribution and Relationships

MLB dataset attribution can vary considerably between different providers and also between individual datasets and over time. Impacts are similar to the previous issue with inconsistencies requiring manual processing and rework within any ongoing delivery chain to end users.

As with all spatial datasets, it is essential to provide suitable metadata to enable users to establish whether data is appropriate for their use. This should as a minimum include currency and source information. In the case of MLB, it would be most useful to attribute the normative legislation

references and, where possible, coordinate reference systems.

S-121 goes a long way in providing structures to do this but requires consistent interpretation and encoding by multiple providers. A defined encoding guide and comprehensive, normative validation tests for attribution of MLB datasets would make handling, processing, and use much simpler. This would also benefit the end users, as they can then make assessments on the usability of the data, with appropriate provenance.

This is challenging, however, because the models as implemented are currently uniquely determined by the originating data producer. It should be noted that whatever the schema and encoding advice which emerges, the importance of consistent attribution should not be underestimated in its benefits for the end user.

3.4.4. Interoperability with Other Domains and Data.

Because MLB data is defined in legislation, it must match the legislation as written. However, for the datasets to be usable in a broader sense by organizations, there is a need for more guidance to be published on how it relates to other spatial datasets from the national infrastructure and, if possible, how end users should resolve such issues, e.g., where lines do not intersect with other defined limits or coastlines. For instance, a published Continental Shelf Limit may comprise just the outer limits themselves with no indication of what the limits connect to (which may be published in other legislative datasets) so a user who needs to establish whether they are inside or outside the Continental Shelf Limit would find it difficult unless the MLB data was supplemented with other datasets and concrete definitions.

This then becomes an issue of geometric consistency of that provider's entire portfolio. While S-121 is comprehensive in terms of its scope (i.e., it includes locations, limits, boundaries, and zones) there will ultimately be a point of interface to a non MLB dataset it becomes an issue of consistency per provider to ensure these datasets are, then, consistent.

If this issue is left to end users or intermediaries to address by, for example, closing lines arbitrarily or using varying coastlines as the landward boundary then this will undoubtedly result in confusion and inconsistency. Currently, users frequently arbitrarily decide a coastline boundary, taking into account:

- Tidal level, MHW, MHWS, 'Extent of the Realm' (in UK) etc.;
- Nautical charts, land mapping agency, boundary commission; and
- Selection of an appropriate point of intersection between the MLB and the chosen landward boundary.

3.4.5. Other Challenges

There are generic, technical challenges with any implementation of MLB data within a technical standard and its promulgation via a set of web services. Issues of completeness of implementation by COTS and open source providers, validation testing, and misinterpretation are always issues and may have a far greater impact with MLB data (one of the reasons why the human readable text use case is so well documented is to eliminate any notion of technical ambiguity around datasets).

Taking these issues into account though (and there may well be mitigating actions which can be

taken over and above the establishment of a global standard for data content and exchange), the potential benefits of data exchange, automated processing, and use can offer enormous benefits to infrastructure and integrated cadastral systems as well as enabling comprehensive distribution systems for broader use.

The legislative interface makes working with such data challenging at the best of times. There are often, within a domain, multiple uses for such data and all organizations may have different methodologies for preparation, maintenance, and promulgation. Related to this difficulty is that some states simply build their own capabilities in isolation of all others - as is their prerogative, but this then makes exchange and the adoption of a common language even more difficult.

The UNCLOS convention currently acts as a normative guide only - it is definitive but does not constitute a "standard" in any sense. S-121, therefore, does not "redefine" any elements of the convention, merely referring to their normative definitions within it. The standard remains a simulacrum though and can not replace or enhance the convention or its interpretation by states parties. The detail of attribution, encoding, relationships, and realization into a "feature model" and a consistent approach is where a lot of the long-term efficiency of the standard will be realized and that is achieved by consensus amongst implementers.

3.5. Summary - Project Goals

Currently S-121 is the only data standard designed with the express purpose of describing UNCLOS data. ISO 19152 forms the foundation of many cadastral data structures on land and is a valuable tool for many administrations, but no globally accepted data standard meets the needs of all users and use cases described, nor unites such cadastral structures with marine zones.

The MLB project is not intending to define such a data standard, IHO has this task within its activities under the S121 Project Team but the OGC pilot has the aim of exploring the technical options and architecture for data representation using the S-121 model, demonstrating such encodings and researching deeper interoperability questions. It is hoped that feedback from this project (through the IHO MOU) will strengthen the existing standard and encourage wider take-up amongst data producers.

Chapter 4. Source Data

This section documents the source for many of the Pilot's activities, the sponsor data used to guide and demonstrate the deliverables. A selection of the data sources have been documented in detail here along with any characteristics which make them particularly suitable for the Pilot initiative. This could be, for instance, the categories of S-121 features they include or certain characteristics of the data source which exercise elements of the implemented schema.

4.1. United Kingdom Hydrographic Office (UKHO) Data

UKHO provided published statutory instruments and treaties together with the spatial data defined in the legal documents (data examples were straight baselines and limits for territorial seas and other treaty-established maritime zones). The instruments and data are contemporary and applied in current management and users' systems as well as illustrating a State's maritime limits from the definition in the legal document.

The UKHO datasets supplied to the pilot project served as representative data for testing the basic S-121 features and information types. UKHO data has no ISO 19152 structures built into it so none of the LADM package from the model required building.

Early iterations of the GML Schema used UKHO data transformed from shapefiles to build arrays of Baseline Points and limit data conformant with the schema (shown below).

The Granville bay dataset was also used to generate test data for the GML schema extensions as in the following example.

Excerpt from UKHO GML data built in Phase 1

```
<ns3:Source ns1:id="src1">
    <ns3:responsibleParty>
        <ns3:responsiblePartyorganizationName>United Kingdom Hydrographic
            Office</ns3:responsiblePartyorganizationName>
    </ns3:responsibleParty>
    <ns3:sourceDocumentName>The Territorial Sea (Baselines) Order
        2014</ns3:sourceDocumentName>
    <ns3:sourceRegistryNumber>2014 No. 1353</ns3:sourceRegistryNumber>
    <ns3:sourceAvailabilityStatus>Document Available</ns3:sourceAvailabilityStatus>
    <ns3:administrativeSourceType>Domestic Legislative Instrument</ns3:administrativeSourceType>
    <ns3:label>The Territorial Sea (Baselines) Order 2014</ns3:label>
    <ns3:sourceSubmissionDate>2019-01-01</ns3:sourceSubmissionDate>
    <ns3:sourceRecordationDate>2019-01-01</ns3:sourceRecordationDate>
</ns3:Source>

<ns3:AdditionalSpatialInformation ns1:id="ad1">
    <ns3:locationReference>
        <ns3:textLat>58Â° 37'.642 </ns3:textLat>
        <ns3:textLon>5Â° 00'.295</ns3:textLon>
        <ns3:referenceSystem>World Geodetic System 1984 Datum (WGS84)</ns3:referenceSystem>
    </ns3:locationReference>
</ns3:AdditionalSpatialInformation>

<ns3:BaselinePoint ns1:id="b1">
    <ns1:boundedBy xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:nil="true"/>
    <ns2:featureObjectIdentifier>
        <ns2:agency>GB</ns2:agency>
        <ns2:featureIdentificationNumber>2</ns2:featureIdentificationNumber>
        <ns2:featureIdentificationSubdivision>1</ns2:featureIdentificationSubdivision>
    </ns2:featureObjectIdentifier>
    <ns3:label>1</ns3:label>
    <ns3:featureName>
        <ns3:Language>en</ns3:Language>
        <ns3:Name>Cape Wrath</ns3:Name>
    </ns3:featureName>
    <ns3:source ns4:href="#src1"/>
    <ns3:refersTo ns4:href="#ad1"/>
    <ns3:geometry>
        <ns1:Point srsName="urn:ogc:def:crs:OGC:1.3:CRS84" ns1:id="b1g">
            <ns1:pos>-5.0049166666666665 58.62736666666667</ns1:pos>
        </ns1:Point>
    </ns3:geometry>
</ns3:BaselinePoint>
```

4.2. S-121 Canadian Maritime Limits and Boundaries Sample Data.

This section contains details of one of the main datasets used to test the entire feature set of the GML Application Schema, the Canadian MLB Sample data. This was by no means the only data used within the project, data from all sponsors has been used to test and evaluate the Schema and will be demonstrated at upcoming MLB meetings and presentations. The Canadian data was a good dataset to focus on, however, because it constitutes interlocking examples of all the Schema elements,

namely:

- Geospatial S-121 features representing UNCLOS concepts;

- Information types representing LADM ISO 19152 structures;

- Data structures representing Sources; and

- Party and Party group relationships.

4.2.1. Data Format

The format of the data was ESRI File Geodatabase. This comprised several geospatial features and non-geospatial tables joined via named bidirectional relations.

4.2.2. Vision

Along with the IHO's creators of the S-100 Universal Hydrographic Data Model, Fisheries and Oceans Canada (DFO) and the Canadian Hydrographic Service (CHS) carry a vision of integrated and interoperable marine data available from its authoritative source, delivered digitally and structured to international standards. DFO is laying the groundwork to support delivery of marine spatial data through its Marine Spatial Data Infrastructure (MSDI), based on IHO's MSDI Working Group work and guidelines.

DFO/CHS see the Maritime Limits and Boundaries of Canada as a crucial foundational layer to support this infrastructure and applications. S-121 itself, by integrating S-100 data structures with elements of the ISO 19152 Land Administration Domain Model will then become a vehicle able to deliver MLB data while retaining their legislative content. This transcends their traditional publication as part of navigational chart products. DFO/CHS see a vision of one authoritative MLB data source with many uses to better support decision makers. Uses for MLB data will include the submission of Canadian limits and boundary data to the United Nations Division for Ocean Affairs and the Law of the Sea (UN DOALOS).

4.2.3. Format and Origins

The ESRI Geodatabase (GDB) format was chosen for several reasons. One of these was its ability to demonstrate the relational nature of the S-121 conceptual model. Another reason was the access to ESRI software. Canadian MLB data was also already managed in a Geodatabase which made it easy to create sample subsets of real MLB data.

4.2.4. Structure

The components of the GDB dataset includes feature classes, object classes (tables), various attributes, relationship classes, a dataset, and some coded domains. The three feature classes represent base types of geographic feature units described in the S-121 UML model, namely Locations, Limits, and Zones. Actual MLB features, such as Baselines or Territorial Sea Limits are subtypes of these. The subtypes are reflected as attributes only for simplicity. They are not coded as feature subtypes since all features of the same base type use the same attributes. Tables are used to represent all other units of the model.

There is a separate table for each unit found in the GDB: the Administrative, Source, and Party

groups in the model. The relational structure of the model is realized by the relationship classes that connect all of the tables and feature classes together. By selecting an instance of data or an individual feature in the GDB, one can navigate through that features extended attribution found in other tables. It is also possible to select groups of features or attributes that are related. An Administrative zone, for example, may have many sub-zones, multiple component limits, and many locations. Some limits may form a part of more than one zone. As a result, there is a complex aggregation that requires the use of a many to many cardinality in most of the GDBs relationship classes.

4.2.5. Content

At the start of Phase 2 of the OGC pilot project, it was decided to create a new joint Canadian example dataset to test the standard's extensibility and align it with the latest stable implementation schema produced from Phase I. This dataset would include Maritime Limits and Boundary Data (Fisheries and Oceans Canada, Canadian Hydrographic Service), Oil & Gas data (Natural Resources Canada, Surveyor General Branch), and a hypothetical Marine Conservation Area that would prototype such an area in the Arctic (Natural Resources Canada, GeoConnections).

The Southeast and East Newfoundland area of the Grand Banks of Newfoundland was chosen for the Canadian Example for the following reasons:

- It excluded areas of international boundaries;
- It represented a table of coordinates assigned to a geographic area, Southeast and East Newfoundland, appearing in three schedules of the Territorial Sea Geographical Coordinates Order;
- It was complex enough to provide a good representation of a legislative order in council's data;
- Fulfilling the requirements of a complex example would allow to generalize to a simple example;
- It allowed the demonstration how governance, source, BAUnits, Additional Spatial Information (textual version of geographic position) and relationships from S-121 could be used to make the data much more useful and interconnected;
- It allowed the data to be structured so as to produce a human readable legislative text output in both official languages that was representative of the level of complexity of an existing legislative document; and
- Canada's Surveyor General Branch had offshore Oil & Gas geospatial information to contribute to the dataset in this area.

4.2.6. History

Various versions of MLB sample data were prepared throughout the course of the OGC Pilot Project. The schema of the first versions created for Phase I matched what was drawn up in the S-121 product specification UML diagrams. This data was useful to explain to others how the S-121 model worked.

As Phase I of the pilot project ended, it had become apparent that it was difficult for participants in the project to connect the UML diagram or the GDB to features, associations, and extendable or

complex attributes in the S-100 world. This was largely due to the fact the XML/GML language used in S-100 is incredibly powerful, its structure is very different from how data is described and structured in a relational database. Another reason is that S-100 utilizes feature catalogs and defines registries containing common conceptual definitions that are shared between various other S-100 standards. As a result, parts of S-121 needed to be renamed to avoid unnecessary overlap with other standards.

Normally there would have been a comprehensive Data Classification and Encoding Guide (DCEG) created by the standards development team that described the individual components of the model in element by element, in detail and their encoding norms. The DCEG would have been used to define a precise feature catalogue, and thus an application schema. The Feature Catalogue describes the model in S-100 format using the XML language using the IHO specific schema. The GML Application Schema, also written in XML, utilizes an XSD format to describe the elements and attributes used in the final S-100 GML file. The OGC participants were, therefore managing these three elements at the same time. The latest version of the Canadian sample data was created specifically to solve this problem. By the end of Phase I, a workable early version of the GML application schema was created. By and large this was built by walking the participants through the model through simple concept diagrams.

It was decided to create a new GDB from the ground up based on the working implementation schema and to test how the data could be encoded. This required an iterative process where adding pieces to the GDB forced changes to the schema and corrections in the schema forced changes in the GDB. The sample data went through iterative changes throughout the course of the project. This work could not be completed before a stable implementation schema was produced.

The final publicly available version of the CHS Canadian Extended Continental Shelf GDB file: S-121_CanadianExample_20200526.gdb.zip.

> **Disclaimer**
>
> Imagery displaying part of the coast of Canada in this report is for demonstration purposes only. It does not necessarily represent Canada's maritime boundaries, zones or the outer limits of its continental shelf, nor does it necessarily represent actual areas, zones or regimes established by Canada for resource exploration or exploitation, or for the protection or preservation of the marine environment.

4.2.7. Observations on Data Transformation

Although multiple language support was added to the original XML feature catalogue, its representation in the model diverged from the provisions within S-100 and within other S-100 product specifications. Canada has a requirement to encode data in both official languages of English and French. It is expected that support for national languages will also be required as well as English at the international level. There are many ways to support multiple languages in a source GDB. This includes duplicating fields, splitting text within a field, separate tables containing alternate language attributes, as well as creating separate coded domain lists.

S-100 did not have a fully comprehensive support for multiple languages, therefore pushing for this to be included in the schema would now requires a proposal for changes to S-100. Some

observations on support for multiple languages are to be included in the outputs to the S-100WG

The mapping between a relational database and S-100's rich structure of simple and complex attributes is not completely straightforward and little guidance exists in the standards which prescribes how such a transformation should take place.

Complex features and complex attributes are not easy to implement in the relational database structure of the GDB. The functionality and efficiency of S-100 is expressed through its XML feature catalogue. CHS found it somewhat difficult initially to communicate a vision of complex features in the final standard when these were not easily apparent from the sample data created.

It was hoped to take advantage of S-100 shared geometry through references from features to base geometric objects. This would have allowed, for example, an edit to a geometry of one feature automatically changing other features and a sharing of interconnectivity through the rest of the S-121 model as described.

It was decided for the sample GDB was to only use relationships to reflect the interconnectivity and an implied shared geometry. This was much easier to illustrate during presentations of the model to show referencing of those same nodes in their geometry. The GDB, however, uses concepts of a topological network or parcel fabric networks for managing shared geometry. Neither of these options worked for illustrating an S-121 compatible sample dataset. Although topology is something used in the source GDB to manage data, the geometry itself is still duplicated between separate feature classes and the functionality is heavily focused towards edition and maintenance of the geographic features. This does not facilitate the level 3a topology within the product specification. The difficulty of full topology within the S-100 GML profile is also part of the outputs to S-100WG

Chapter 5. The GML Application Schema

This section documents the creation of the GML Application Schema, designed to provide an encoding of S-121 data for exchange between parties.

5.1. Introduction - IHO S-100 - The Universal Hydrographic Data Model.

The development of electronic charts over the last 25 years is synonymous with IHO S-57, published in 2000. IHO S-100 has been under development since that time and is currently at edition 4.0.0 (although edition 5.0.0 is currently being drafted).

IHO S-57 defines many entities and important publications:

- The main Standard which describes the language and structures available to implementers;

- The overall specification for ENC (defined in Appendix B.1) drawn from a superset of allowable structures (defined in the main document of S-57);

- An encoding (i.e. the format of the data for exchange with the end user via hard media – this is ISO 8211, a self-describing 2-dimensional binary format) along with the method of aggregation of features into datasets embedded on "hard media;" and

- The feature and attribute catalogues detailing the data model of the ENC.

IHO S-100 took this assortment of structures and content and used it to define a much broader standard where, as with S-57, the framework was defined, but left the definition of actual content to the implementer of individual product specifications. The content itself is defined within a machine readable XML document called the feature catalogue which reflects features defined in a geospatial registry (normally the IHO registry located at the IHO in Monaco – although there is no reason why the registry cannot be one located elsewhere).

Portrayal, similarly is defined in terms of a machine readable catalogue with individual product specifications using whatever they require.

The goals of IHO S-100 are the following.

- To produce a data standard in closer alignment with the ISO191xx framework.

- To be neutral in terms of actual data itself leaving definition of content to the creation of individual product specifications.

- For documentation and content to be fully machine readable, paving the way for "plug and play" end user systems which can interpret arbitrary product specifications without the need to reconfiguration or manufacturer intervention.

- To separate, logically and completely, data content from its portrayal and encoding.

- To allow the update and maintenance of product content and portrayal by the end user.

IHO S-100, therefore defines a number of product specifications (developed by different communities coordinated by individual "domain controllers") and their encodings. S-100, as stated,

has offers the ISO 8211 encoding as well as GML through the S-100 GML profile (S-100 Part 10b). S-121 is a product specification implementing the S-100 framework. Currently within the S-121 standard no GML defined encoding exists.

5.2. The Relationship to the Feature Catalogue.

The GML Application Schema is intimately related to the IHO S-121 Feature Catalogue. This XML document, defined with the S-121 product specification defines the features, attributes, metadata, relationships, and geometry primitives required for the encoding of S-121 data. Entities within the S-121 Feature Catalogue are defined in the IHO geospatial registry.

The essential "content" of the GML Schema, therefore exists as an implementation of the register of features defined by the feature catalogue.

A short extract from the S-121 Feature Catalogue is shown in the following diagram. In it the structure can be seen of feature naming, the binding of features to specific attributes and the name and type of attributes. The Feature Catalogue description is contained within [1] in its entirety and the S-121 feature catalogue forms the major input source for the creation of the GML Application Schema.

```
<S100FC:S100_FC_FeatureType>
          <S100FC:name>Boundary</S100FC:name>
          <S100FC:definition>Boundary [UNCLOS ref: ]</S100FC:definition>
          <S100FC:code>Boundary</S100FC:code>
          <S100FC:remarks>definition as per UNCLOS </S100FC:remarks>
          <S100FC:alias>BNDARY</S100FC:alias>
          <S100FC:attributeBinding sequential="false">
              <S100FC:multiplicity>
                  <S100Base:lower>0</S100Base:lower>
                  <S100Base:upper xsi:nil="false" infinite="false">1</S100Base:upper>
              </S100FC:multiplicity>
              <S100FC:attribute ref="context"/>
          </S100FC:attributeBinding>
          <S100FC:attributeBinding sequential="false">
              <S100FC:multiplicity>
                  <S100Base:lower>0</S100Base:lower>
                  <S100Base:upper xsi:nil="false" infinite="false">1</S100Base:upper>
              </S100FC:multiplicity>
              <S100FC:attribute ref="fixedDateRange"/>
          </S100FC:attributeBinding>
          <S100FC:attributeBinding sequential="false">
              <S100FC:multiplicity>
                  <S100Base:lower>0</S100Base:lower>
                  <S100Base:upper xsi:nil="false" infinite="false">1</S100Base:upper>
              </S100FC:multiplicity>
              <S100FC:attribute ref="label"/>
          </S100FC:attributeBinding>
          <S100FC:attributeBinding sequential="false">
              <S100FC:multiplicity>
                  <S100Base:lower>0</S100Base:lower>
                  <S100Base:upper xsi:nil="false" infinite="false">1</S100Base:upper>
              </S100FC:multiplicity>
              <S100FC:permittedValues>
                  <S100FC:value>1</S100FC:value>
                  <S100FC:value>2</S100FC:value>
                  <S100FC:value>3</S100FC:value>
              </S100FC:permittedValues>
              <S100FC:attribute ref="releasabilityType"/>
          </S100FC:attributeBinding>
          <S100FC:attributeBinding sequential="false">
              <S100FC:multiplicity>
                  <S100Base:lower>0</S100Base:lower>
                  <S100Base:upper xsi:nil="false" infinite="false">1</S100Base:upper>
              </S100FC:multiplicity>
              <S100FC:attribute ref="interpolationType"/>
          </S100FC:attributeBinding>
          <S100FC:featureUseType>geographic</S100FC:featureUseType>
          <S100FC:permittedPrimitives>curve</S100FC:permittedPrimitives>
      </S100FC:S100_FC_FeatureType>
```

5.3. Background to the GML Application Schema for S-121

The GML Application Schema for IHO S-121 was the foundation of Phase 1 of the project and further developed in its Phase 2 to enable the additional functionality described by the architecture. Being focused on interoperability the pilot requires a "common language" in which to describe the data supplied by the sponsors and then used for the interoperability testing carried out by the project.

The Application Schema defines a structure which determines how a GML encoding of the S-121 data can be constructed. GML Application Schemas are a well documented and well-defined XML structure by themselves and the architectural decision was taken early in the project to use the GML vocabulary defined under Part 10b of IHO S-100. This GML profile, added to S-100 around 2013 defines a set of GML Structures used specifically for the modeling of S-100 product specifications and, being a (from an S-100 point of view) reasonably simple vector product, S-121 should lend itself well to such treatment.

The development of the GML application schema was executed very much as per an OGC project though with the emphasis, jointly, on implementation, iterative development, and interoperability through defined and bespoke software components. In doing this a GML schema more well suited to implementation by project stakeholders was arrived at along with outputs suitable for feedback back to IHO working groups.

So, in defining the application schema there were unique challenges by virtue of the S-100-centric approach taken, and also by virtue of its domain as an expression of features defined under the UNCLOS convention. This places some unique constraints on certain aspects of the data which require consideration when the application schema is developed and which are documented in the next section.

Once defined, the application schema provides a way of structuring and encoding S-121 data which meets the requirements of the model and S-121 feature catalogue - data encoded in this form was used extensively in Phase 1 of the Pilot to show basic functionality of S-121 data transformation and creation and then in Phase 2 for more sophisticated encoding and interoperability testing.

5.4. Implementation of GML Application Schema

The project, both in Phase 1 and Phase 2 spent a great deal of time and resources adapting the schema to certain aspects of the data structures in the sponsor data, and also to the conceptual schema defined within S-121 (and norms within the S-100 framework Part 10b). This harmonization and multiple considerations meant the schema continued to evolve even up to the final week of the project. The complex nature of the product specification, particularly its implementation of ISO19152 structures, ISO19115 metadata and UNCLOS-defined S-121 geospatial features required attention from many project participants in order to achieve its goals.

Ultimately the outputs of the project still reflect a work in progress and a first iteration of the GML schema is a prime output. It reflects many conversations/discussions and design decisions, but its iterative nature made it challenging for the project to build on the schema in terms of web services implementations and comprehensive datasets which test all elements of the schema. But resources

are always limited in these respects. The pilot provides a foundation on which to build more comprehensive implementations within the community and presentations to the OGC and IHO communities post-project will build on the foundation.

5.5. Structure of Schema

The GML schema defined is, in S-100 terms, a fairly standard one conforming to the structure and content of Part10b of S-100 as well as the Rules for GML Schema defined in ISO 19136. It has a modular structure expressed in the following ways.

1. It defines a hierarchy of feature types based on the structures within the UML model from the IHO product specification and reflected in the current edition of the feature catalogue. As the model has a hierarchical structure, so too does the GML Schema.

2. It groups features together by type and places them all within a single feature collection root element "Dataset."

3. It uses the S-100 GML profile structure and includes standard references utilizing the Part 10b GML profile elements only.

As part of the project a map of the GML structure expressed in S-100 style UML has been constructed. This is shown in the following diagram. The aim is to show the essential structure of the GML Schema and how it implements the various elements of the S-121 model. This diagram therefore shows, in blue, the Abstract types developed for the geospatial features together with their S-100 simple and complex attributes, as well as the realisations of the information types used for the ISO 19115 metadata and source as well as the ISO 19152 structures for LADM rights, responsibilities and restrictions content.
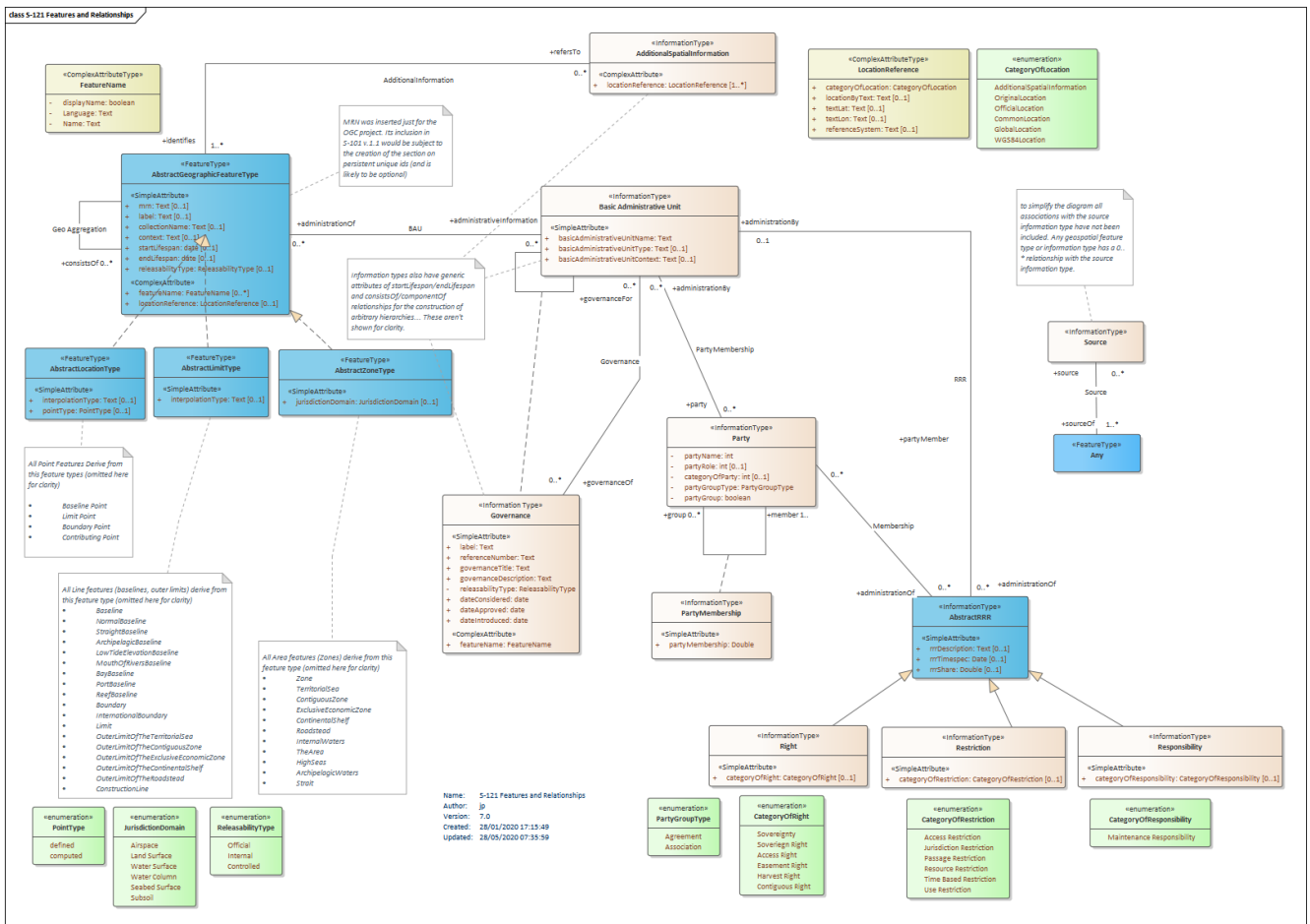
Figure 8. S-121 GML Application Schema structure.

The Schema defines, GML elements for the following:

- Metadata

- Geospatial features

- Feature attributes

- Information types

- Geometry representation

- Associations between entities in the schema

These are dealt with in detail in the following subsections.

5.6. Metadata

Metadata elements comprise the standard dataset metadata elements which are part of the GML Profile under Part 10b. This currently is the "dataset identification field" and the "dataset information field." These elements are part of the GML profile and are also not derived from Abstract GML (which can make them opaque to some GML readers and therefore difficult to parse). The field contents are also heavily modeled on the original DSID/DSSI fields used in the ISO 8211 encoding within S-100. A longer discussion on this issue is presented later in the report along with some recommendations for the S-100WG to take on board. S-100 also defines a separate metadata implementation based heavily around ISO 19115 and documented within S-100 Part 4a. The

purpose of the S-100 metadata structure described in Part 4a is to implement a profile of ISO 19115 for the two purposes:

- The implementation of a standardized metadata profile to facilitate interoperability for the purposes of data sharing, discovery and querying; and

- The aggregation of datasets (themselves aggregations of geospatial features and information types) into "exchange sets." These exchange sets are an integral part of S-100.

Metadata should be interoperable with that defined in other domains. The metadata content deliverables within the Pilot have centered on content discussions in relation to the Canadian Government community profile of ISO 19115:2003, commonly referred to as "HNAP" and their implementation within OGC web services.

5.7. Geospatial Features

Geospatial features are modeled as per Part 10b with a single named GML element per feature type defined in the feature catalogue and matching the definitions within the feature catalogue (which were, in turn, derived from the UML model prior to the commencement of the OGC project). Due to the structured makeup of the UML model it is possible to derive most of the geospatial features from a single type with a common set of attributes. These are tabulated below.

The geographic feature type within the GML Schema is itself derived from the S-100 AbstractFeatureType defined by Part 10b.

From this type, three subtypes are derived for geospatial point, curve and surface features with slightly different attribution based on the features in the S-121 model and feature catalogue. It is not the geometric primitives which differentiate these groups of features but attributes specific to those features so three subtypes are derived "locationType," "limitType," and "zoneType" reflecting the nomenclature in the model.

Notable common attributes in the generic feature type are as follows.

- MRN - this has been added into the generic geospatial feature type. There is much discussion in the IHO and S-121 community about the inclusion of Marine Resource Names (MNRs) in S-100 product specification but the current guidance in the standard is a strong recommendation for their inclusion in product specifications. S-121 is currently debating their inclusion but for the Pilot is was decided to include them given their existence in one of the sponsor datasets. In the schema they are optional (currently).

- startLifespan/endLifespan - enables versioning of features by date.

- featureName - a complex attribute which is unbounded, it holds a language attribute and a name defined in that language. This enables multi-lingual naming of features. This is another area where considerable discussion was held in the project and where alternative approaches could be utilized as the requirement for data to be multi-lingual extends beyond the immediate "naming" property. this is also discussed later in this section.

- locationReference/refersTo (reference) - allows textual representation of positions alongside (or by reference). Detailed later in this section.
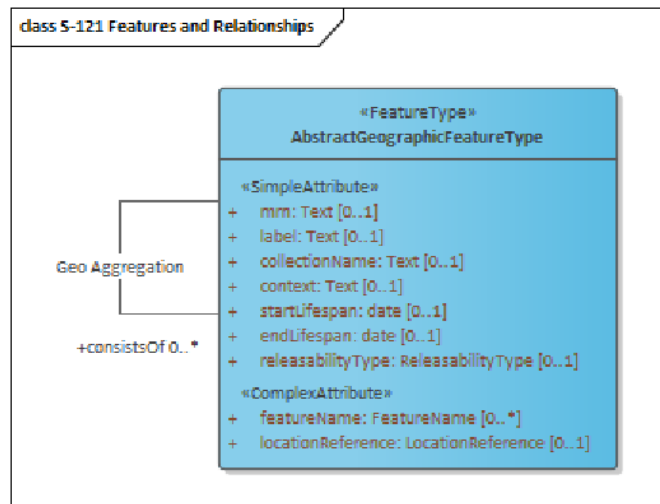
Figure 9. Abstract Feature Type - geospatial features base types.

Geometry, under the GML profile is an attribute of the features and can be inline or referenced. A geospatial S-121 feature can only be one of 3 2D geometry types: point, line, polygon, or multiples of any of these. One GML feature cannot have multiple different geometries (such as both a point and a polygon). These properties are also encoded in the S-121 XML Feature Catalogue.

Each of the individual feature types within the model (the actual UNCLOS entities themselves) are, then, one of the geographic feature types. As a minimum, the geospatial feature types express the extent of the product specification's coverage of the UNCLOS convention.

5.8. Attributes

Attributes of features (whether geospatial or information types) are implemented either as simple xml types (xsi:string or date etc.) or complex types where necessary.

The implementation of the UML model within the S-121 product specification requires a number of complex types, particularly for the information types relating to source, governance and the ISO 19152 (LADM). Within the GML application schema these are modeled as defined within the S-121 feature catalogue and cross-checked against the UML model.

During Phase 1 only rudimentary use was made in any of the data relating to these entities however so little cross-checking of the fields against the model and feature catalogue was initially done. This also highlights the difficulty of ingesting and displaying to an end user complex GML data (and associations between features and information types) - discussed in Use of WFS-SF and List Attribute Flattening.

5.9. Information Types

Information Types are implemented as features derived from S100:AbstractInformationType. This type is defined within the S-100 GML profile and consists of AbstractGML coupled with generic feature associations (implemented by GML reference) as defined in Part 10b.

A local base S-121 information type is defined and all information types within the feature catalogue then modeled as individual GML elements with simple/complex features as defined in the feature catalogue. The Information Types defined are for recording of source information, for

governance and for implementation of the ISO19152 LADM structures articulated within the product specification. Little formal guidance exists for how these attributes are encoded although some of the sponsor datasets within the project offered example instances for implementation, most notably the Canadian GDB example. This aspect of S-121 implementation is also dealt with in Encoding of LADM RRR Data.

5.10. Geometry

Broadly geometry is defined in the application schema within an element called "geometry" - this element can either include a reference to a geometry element defined elsewhere in the schema or it can define its own geometry implicitly.

Each of the three geospatial types defined (for the groups of point, curve, and surface features) have a different geometry definition specifying S100PointPropertyType, S100CurvePropertyType, and S100SurfacePropertyType. Geometry primitives. The Application Schema only uses the S-100 GML Profile geometry primitives to represent the content of the feature catalogue XML.

This allows a GML file to embed either inline geometry (where the actual geospatial information is embedded in the feature definition itself), or geometry by reference where a xlink:href xml attribute points to a geometry element elsewhere in the GML file. This dual nature of the geometry elements within the GML schema (and data) has some implications which are explored further in Additional Spatial Information and within the Executive Summary.

5.11. Relationships and Associations

The initial version of the feature catalogue has few relationships within it defined. A large part of the Pilot has centered around those relationships. Relationships form the structures of the information types in the FC (and Schema) for the ISO 19152 structures in the LADM package of S-121 and it was necessary to reproduce those structures within the schema to adequately describe the LADM elements in their entirety.

A lot of testing has been done, mainly based on CHS's structures (which are extensive). After much discussion generic relationships to generate hierarchies (consistsOf/componentOf) were inserted into the two major abstract types (geospatial and information types). This allows arbitrary hierarchies to be developed. This, although unconventional in IHO product specification terms is a good addition to S-121 given how specific states may be in terms of their implementations.

Other significant relationships deal with:

- Basic Administrative Units (BAU)<→Geospatial features;
- Features<→Source (where there is still likely to be testing following the Pilot);
- Party <→ Group/membership;
- BAU<→Governance; and
- AdditionalSpatialInformation relationships.

S-100 has a rich structure of associations and the feature catalogue mechanism (S-100 Part 5) defines ways of associating features together whether geospatial features or information types. S-

100 Part 10b defines two methods by which associations can be implemented, either by generic association references (included in the S-100 profile S100AbstractFeatureType and S100AbstractInformationType) or by name where the role name of the association is given to the name of an element within the feature which then contains a reference to the associated feature. The decision was taken to use the role name approach in order to differentiate relationships more easily in the GML. These are reflected in the overall diagram describing the schema.

S-100 also contains, within the Abstract Feature types, inverse associations which can be used to implement the opposite end of an association. The S-100 feature catalogue mechanism allows the author of a product specification to be very specific in which associations are allowed between different feature and information types but the generic associations embedded in the GML Profile Abstract types lose this specificity to individual feature classes so the choice to adopt the named elements for associations means the associations embedded within the model (e.g., everything can have a source association but only geographic features link to a BAU information type) can be preserved.

For this reason, the associations are all named within the application schema. For example all geographic features have a "source" element which contains a GML reference to a source element defined elsewhere in the GML file.

Relationship names are shown in the overall diagram representing the Application Schema. The decision was taken to avoid names where the relationship names clash with the type name of the element at either end of the relationship to avoid issues with some software implementations.

It should be noted that a considerable amount of effort was spent establishing relationships from the underlying UML model and using the S-100 mechanisms to define them. This will be implemented in the S-121 feature catalogue after the project to complete its next iteration. Representative relationships are crucial for development of semantic web specifications. This is an important trend in the development of searchable, semantically rich information. While S-100 product specifications are not RDF per se the comprehensive relationships defined during the pilot could certainly be transformed for such purposes and used to. for example, feed an inferencing engine.

5.12. Aggregation of Features into Collections.

A dataset elements "Dataset" is defined with a metadata section containing a sequence of S-100 DSSI and DSID elements. these are a standard part of the S-100 profile (although not a normative part of the profile). Indeed, the structure of the collection feature was the subject of much discussion and the current specification reflects a best view in terms of interoperability with the various clients.

The structure, shown in the following image shows the (dataset) structure adopted:

- A dataset metadata section containing the S-100 Profile DSID and DSSI elements;
- A collection of S-100 geometry (these are standalone geometries which can be referenced by features); and
- A "members" element which contains a sequence of features from a group formed of all possible S-121 geospatial feature and information types.
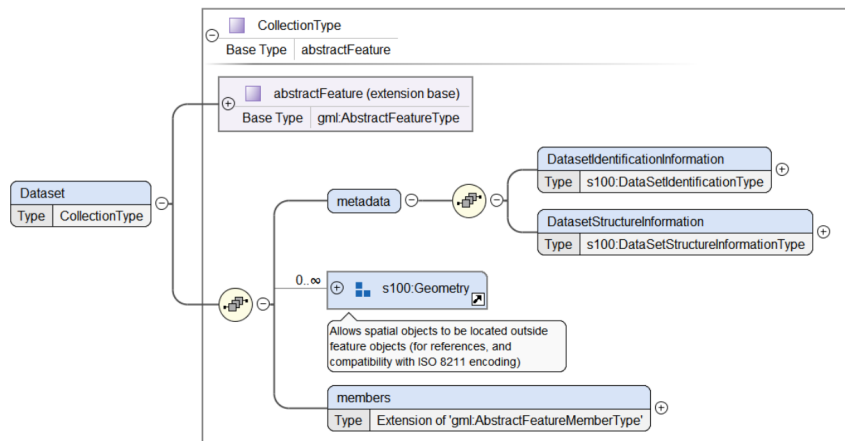
Figure 10. GML Schema dataset structure

The decision was made to aggregate the features within a single "members" element, rather than splitting them into "member" and "imember" representing feature and information types (which is common in other S-100 GML Schemas). This reduces the overhead of including a wrapper "type" element around every single feature but also means that there is no implicit ordering of features within the dataset. This may cause interoperability issues with other S-100 base GML schemas and probably requires some discussion with the S-100 community - there is currently no mandated method for describing GML feature collections but the member/imember concept has become a common scheme. A submission to the S-100WG will seek to clarify this point and establish a method (or number of normative alternatives) for aggregating well-defined features into datasets.

5.13. Implementation Topics

The project implemented the GML profile for S-100 in a particular way and there are a few challenges in how that is done. Some of the design decisions taken are documented in this section.

Separately some of the outputs of the process are documented in terms of IHO inputs (mainly to Part 10b of S-100 to better focus future efforts) and partly whether there are inputs for e.g., vendors or implementers in the open source community who would benefit from the experience within the pilot.

5.13.1. Additional Spatial Information

The ability to represent locations textually is a major cornerstone of S-121 and this underwent several iterations in the design of the GML Schema. A final design emerged after much discussion around the merits of information types vs inline attribution, multiplicities and the exact representation required. The final design aspects are as follows.

- All geospatial features need to represent their positions both geospatially (through the geometry element) and textually.

- The textual representation needs to be broad and non-specific. However, a distinction can be made between textual descriptions of position (latitude and longitude) and "location" without coordinates.

- A complex attribute locationReference is defined which contains textual latitude, longitude attributes. It also contains a generic locationByText attribute and another for documenting

reference systems.

- Each geospatial feature can include, inline 0 or more locationReference attributes to denote any number of textual positions.

- Sharing of textual representation may be a desired feature for some implementing producers. Therefore, an information type called "AdditionalSpatialInformation" is included which can be associated with a set of geospatial features (via their "refersTo" association). These information types can also contain an arbitrary number of locationReference attributes.

This model of "additionalSpatialInformation" seems to balance all requirements for inline ease of specification as well as sharing via information types.

5.13.2. Feature Naming

As documented within this ER in several places the ability of the schema to support multi-lingual "names" via the complex attribute feature name is an integral part of its design. This feature is key to the requirements of one of the major sponsors of the Pilot and has been extensively discussed and implemented in detail for two prime use cases.

Multi-lingual names are implemented by establishing a complex attribute "featureName" which is composed of a language indicator and name attribute, both textual. These complex attributes are unbounded in the schema and allow for multiple, parallel languages to be established for any feature name. This structure derives from established attribution already within the S-100 geospatial registry at the IHO and used in several IHO product specifications with similar encoding.

As is documented in Outstanding Issues and Ongoing Work Items with the GML Schema and Input to IHO S-100 Working Group the shortcoming of this is where many attributes potentially form "outputs" of datasets and require translation. Descriptions, column headings, and other elements within the model are potentially written out to end users - this is seen extensively in the Human Readable Text use case, documented later. The extent of the fields output in the HR Text makes establishing multiple multilingual attributes somewhat unwieldy. This requires further discussion within the S-100WG and, potentially the S-121PT as the GML Schema is drafted for inclusion within the standard as an annex.

A related issue is whether multi-lingual support should also need to extend to the naming of the internals of schemas and content, such as GML types and UML diagrams and encodings. This was discussed during the pilot. Where administrations have statutory obligations to support multiple languages it can be a particular issue, particularly as many standards are written with their technical components implicitlyin the English language.

5.13.3. Marine Resource Names (MRN)

These have been included in the GML Application Schema as simple textual attributes. Their introduction is not meant as an endorsement in general for the S-121 product specification but has allowed the Pilot the opportunity to produce populated data for demonstration to stakeholders. As stated in the Input to IHO S-100 Working Group there are broader issues with identifiers that require consideration with the S-121 and IHO community, and this will be an ongoing topic within the S-121 project team after the pilot has concluded. There are few "technical" issues with MRNs as they represent simple text-based data but their introduction as a universal persistent identifier for

vector data requires a more thorough treatment across the S-100 and S-121 stakeholders.

5.13.4. Metadata and S-100 Part 10b

The current dataset metadata content is essentially taken from the header of the ISO 8211 (part 10a) encoding in [1] and much of it is probably not relevant to the GML encoding.

There is a contrast between the dataset metadata, however, and the ISO 19115 metadata included in the schema. If an S-100 exchange set is being used for delivery, then discovery and ISO 19115 type content can be included in an accompanying XML file in the exchange set (as per S-100). However, when the GML is delivered on its own such as when delivered via WFS it might be good to have the metadata embedded in the a single file.

The DatasetIdentificationInformation field (taken from ISO 8211 header DSID) does not seem to be derived from a GML type. It probably should be an extension of the gml:AbstractMetadataPropertyType. This seems like an oversight in S-100 and will be passed on to the S-100 WG.

The DatasetStructureInformation (DSSI) field is also taken from the ISO 8211 encoding (presumably to establish an equivalence between the encodings) and not a GML type.
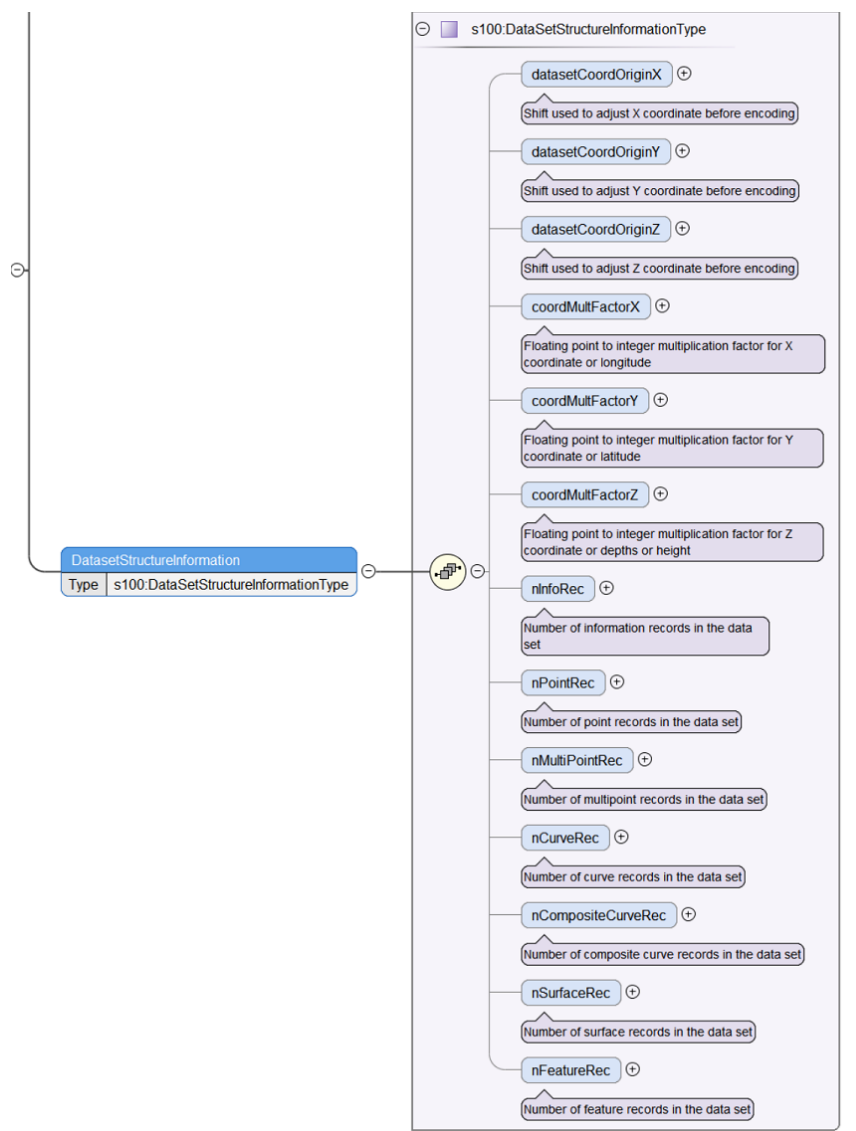
Figure 11. DSSI field structure

DSSI does not seem to hold much value in a GML encoding. This field has elements for origin and coordinate multiplication factors used when converting the coordinates to/from integers. In a GML context these make little sense and the GML profile contains no equivalent structure so in interoperability terms use of these multiplicative factors would render a GML file unreadable.

The rest of the DSSI element are just totals of the numbers of each type of records in the file. These counters were originally intended to help the ECDIS know how much space to allocate for internal structures and to validate the counts. The counts could still be useful but probably not mandatory and their value in GML datasets should be examined with stakeholders. Totals could form a useful check of dataset consistency but this should be balanced against, e.g., the data integrity mechanisms present in S-100 Part 15.

In general metadata in terms of the GML profile in Part 10b probably needs a revamp - using DSID/DSSI fields is mot really a good structure and reflects the aged status of the GML profile. Some of the work currently going on in the S-100 community may well address this, and certainly the state of metadata as far as vanilla S-100 is concerned is far more advanced than is currently the case within the GML profile. This is an issue for the S-100 GML profile to address.

5.13.5. Feature and Information Type Mixing

In the schema model, the information types are mixed in with the feature types in the same group of abstract objects. Normally in the S-100 order the information types would be listed before the geometries and features which could be referencing them.

It was noted during the project that some implementing systems ignore the information types and just read features with geospatial content. It was also noted that within vanilla GIS systems it is difficult to inspect the relationships between the information types and the geospatial features with which they have relationships. In general, GIS systems display only features which have a geospatial representation and information type data is rendered as static relational tables. The issue here is the use of complex GML to represent this data as there is no prescribed way of displaying complex GML structures which is adopted across multiple GIS providers.

This maybe points to the need for bespoke viewing/editing and import interfaces for manual data inspection (for use cases of exchange between interested parties) and also editing. Some participants provide solutions which are capable of natively working with Complex GML and S-100-type structures and this should be acknowledged as a current issue across the industry - GML implementations can be variable due to the sheer size of the standard and yet the Simple Features profile, with its lack of support for Complex Attribution is therefore incompatible with S-100.

Further testing is required within Phase 2 of the project where, undoubtedly, there will be an enhancement of the models examined. It should be established whether plugins and underlying drivers are able to render data in an intuitive way for the end user in a generic fashion and to what extent GML and Complex GML are supported by the open standards community in general. Plugins like GMLAS are a good start to this effort.

It is also worth noting at this stage that for full model exchange the existing S-100 ISO 8211 format can handle the information types as shared entities.

5.13.6. Dataset Ordering

The comments on feature ordering are valid. The general approach in S-100 (which stems from S-57 ultimately) is that anything should be defined before it can be used. When datasets have references, links or separate geometries, all of these things should be defined before they are used. This comes from a time where early ECDIS would read in datasets as streams of data but probably has applicability now. Certainly, in GML there is no such requirement, nor in other formats too, e.g., in HTML you can refer by href to things which are later in the file. This characteristic of S-57 (and S-100/S-101) has some difficulties in implementing if (as there are in S-121) there exist associations between information types.

Broadly, it is easy to enforce ordering by defining dataset order as:

- Metadata

- Information types

- Geometry

- Geographic feature types

But even this broad order requires e.g., that features are ordered so that the associations are properly definable and when you allow associations within categories (i.e., so information types can have associations with other information types they should be ordered appropriately).

In short, it's not easy to order these features and even harder in a GML schema to ensure something is defined before it is used but a mitigation could be that a specific validation test on a dataset is defined which enforces this requirement. This is for discussion within the project and also, potentially for the S-100 WG to look at in the context of Part 10b (and other encodings).

5.13.7. Geometry Representation within the GML Schema

This section describes the approach to geometry adopted within the Application Schema. This subject caused much debate and experimentation in the schema development and many interactions and different methodologies. It was also highlighted as one of the areas where the current S-100 GML profile could be improved in terms of its guidance and specific requirements of Application Schema developers.

The Geometry part of the application schema describes how the geospatial aspects of the data are dealt with. This should be differentiated from the (optional) textual representation of data which may be embedded in the locationByText attribute of the additionalSpatialInformation information type (and which may, or may not be, the same information).

In the context of the application schema these elements are the geometric primitive points, curves, and surfaces (this ER uses the S-100 context of "Surface" throughout, essentially a 2-D closed polygon) which may be shared and referenced from features. The requirement from the S-121 product specification is that the geometry of the dataset conforms to Level 3a topology as defined in IHO S-100 Part 7. This is problematic as Level 3a requires a fairly rich topology model requiring edges to have externally referenced start and end points, directed curves, and validation for self-intersection and other consistency guidelines.

This places a large responsibility on producers to validate datasets and produce consistent data but it is in place for good reason. It ensures that when point location, outer limits, and zones are coincident, that the geospatial positions are forced to be equivalent and cannot be erroneously produced. Under the default ISO 8211 encoding of S_100, this is achieved by the development of a topological model where locations by reference are embedded within the encoding itself. The GML profile of S-100 Part 10b contains no such provisions, excluding the GML topology model. It is unknown the extent to which GML topology is included in participant implementations of GML. Certainly, some participants (e.g., Safe Software) reported their software as capable of reading GML with geometry by reference but the extent of implementation by e.g., QGIS and GDAL is less certain (support for references is present within simple GML but complex GML through schemas is incomplete).

The project architecture was predicated on following the S-100 GML profile and therefore this was followed, and an approach made, which was "topological" as far as possible within the constraints of the S-100 GML profile. This still leaves architectural choices within the Application Schema, however, which are documented here.

The basic choice to be made is whether to include Geometry inline or by reference within datasets. Inline geometry is included in each feature and by reference is where geometry is defined elsewhere in the dataset and referred to by implementing features. Many of the test datasets used inline geometry (which would place a responsibility on implementing system to resolve GML references to tie individual geospatial features to their geometry - this was implemented by one of the participant systems but a comprehensive survey is not available at this time) and test datasets comprising referenced geometry were constructed during the project.

The final result of the geometry specification, at the conclusion of the project, was the following.

- To include geometry within each geospatial feature as a specific element <geometry> which contains **either** a reference to the GML element containing the geometry component for that feature, or the geometry itself, **but not both** (this would have to be enforced by validation by the producer).

- If both are included in a dataset, then by GML definition the inline version of the geometry is considered to be a "cached" copy and the referenced version to be normative/definitive. This is well defined within GML and should be reinforced with data producers.

- Although full topology support is not included in the standard GML profile, this can be addressed through validation testing and a more comprehensive set of validation tests in that respect should be output to the relevant groups defining later versions of the product specification.

Duplication of geometry - as described above the use of duplication of content inline could improve the generic use of the GML files however it does present a concern for validation and for importing into a system that can keep these as shared entities which is desirable for maintenance and consistency.

In cases where data is duplicated inline and read into a system that supports the full structure and sharing of content (information types and geometry) the reading system could look for duplicates during import but it would make the process more robust if these objects were carrying an identifier which could be used to indicate they are the same entity. Also having identifiers would

help in validation of content that has been duplicated inline.

5.13.8. Conceptual Schema Construction

It would be theoretically possible to design a transformation from S-100 XML Feature Catalogue to GML Schema. This would make explicit the conversions and equivalences which are currently hardwired into the Schema. Both documents do effectively the same thing, the GML Schema introduces "extra" elements though, through types and through its inclusion of the broader GML Schema from its source.

It would make sense to define the mapping from FC to GML Schema directly and leave it as a FC design process. The hierarchy of inheritance is lost but this is dealt with in the application schema in the product specification and does not need to be a separate entity.

5.14. Outstanding Issues and Ongoing Work Items with the GML Schema

- Should a generic feature (with categoryOf attribute) be used instead of individual features (this is an issue for the S-121PT)? This was raised several times by several participants. It is a long standing subject of discussion within the S-121 project team as well and has always been resolved in favor of discrete features representing the core UNCLOS concepts. Interoperability questions should be taken into account though as well as ease of transformation from sponsor databases to S-121 GML.

- Many-to-many relationships between the Basic Administrative Unit and geographic features to provide aggregation into generally described administrative areas while still allowing separate administration for each specific feature. These many-to-many relationships have been added to the draft v4.1 of the GML schema for onward discussion with the S-121 Project Team as the GML Schema is proposed as an encoding to the product specification.

- No DCEG-type guidance exists currently guiding implementers on definitions of features/attributes and relationships nor of established norms implementing ISO19152 which would help its translation into the marine domain.

- Minor amendments to feature/attribution (case in sub-attributes), also naming of some relationships.

- MRN is not supported officially in S-121 (yet). When the schema is proposed to S-121 (for v.121) MRN, suitably qualified, should be considered as part of that proposal. The planned submission of the GML Schema as a proposed annex to the existing standard will make this a concrete proposal and allow the issue to be resolved.

- "A better way of doing multi-lingual." Features which represent "textual output" within a product specification all need to, potentially, include an arbitrary number of language translations. S-100's established mechanism is to use complex attribution adding language alongside the text but this is only implemented in complex attribute featureName currently. A more holistic approach should probably be arrived at and also used to influence S-100's evolution in this aspect as well. It may be possible to reuse the S-124 localtext attribute to achieve this end. The sponsors and Pilot team participants should take this on both in the context of S-121 and within the S-100WG to address the broader issue.

- A more formalized approach to documenting how the relationships are implemented - the schema is very broad and allows a lot of freedom. It is worth documenting in more detail how some sponsors use the attribution and (optional) packages. The generic addition of componentOf/consistsOf relationships may well be controversial but was seen as necessary in order to allow implementers considerable flexibility in creating arbitrary hierarchies of features, whether geospatial or non-geospatial in nature. The adaptability of the LADM structures to the marine domain has been tested extensively within the project but it is likely that other data producers may need more flexibility in its application which would be supported by the generic relationships.

- More exhaustive testing of other elements of the schema. Source relationships have not been exhaustively tested.

- A way of constructing the schema in a modular fashion needs to be arrived at. This would allow the full S-121 schema and the extension schema to be derived from a common base of elements rather than duplicated content. More on this topic is included in the ER in the section discussing extensions. Establishing a more modular structure for schema creation would enable this dual-fuel approach.

- S-121 geometry defined as gml PropertyType directly is a different pattern than the expected S-100 PropertyTypes. This does conflict with S-100 Part 10b geometry specification but opens up interoperability with the open source community considerably. This is a workaround to gain compatibility with the GDAL library and should be fixed within that community. Some support in that area has been received already and this is an ongoing item. Discussion with the S-100 group could define better a more prescriptive methodology.

- Associations//Relationships:

  ◦ Some association definitions such as between Party and PartyMember are still largely untested. The structures have been defined but not implemented other than in a simple context for one of the sponsors and not tested to the same degree as the LADM package or the geospatial features. This could use some more testing with different sponsor data.

  ◦ In the xsd, there are broad componentOf and consistsOf associations which are allowed between almost any two features. With some experience with actual data, these can be made more prescriptive with encoding guidance about how to use these associations.

- Attribute names in lower camel case and feature names in upper camel case need to be made consistent with S-100:

  ◦ In S121 xsd featureName subattributes 'Language' should be 'language' and 'Name' should be 'name'.

  ◦ AdditionalSpatialInformation vs additionalSpatialInformation.

- There are some outstanding unused type names in the schema which should be updated. infoFeatureName is one for consideration.

5.15. Extensions of the GML Application Schema

One of the requirements within Phase 2 was the requirement to extend the GML Application Schema. This was not a requirement to extend the schema per se, but rather to use it to represent an extended vocabulary of feature types and data. The S-121 schema represents a comprehensive

language of features, attributes and relationships specific to the MLB domain but many broader use cases require (for interoperability) other domains to be tightly integrated with such foundational definitions. That is, the ability to refer by reference to S-121 features without changing their essential content, presenting a harmonized view alongside the definitive S-121 picture.

The requirement, therefore, is to use the schema to express other domains of data (the examples given at the outset of the pilot were for other administrative areas and zones relating to exploration and hydrocarbons).

The longer term view of such extensions is to enable the creation of data to support "georegulation." Although no single definition of georegulation exists, it is a concept popular in many states and administrations and one where there is significant interest in development and standardization. Fundamentally, it concerns the management and regulation of a delimited space, be it marine or terrestrial, in three dimensions (as well as with respect to time) and is a primarily legislative concept. There is a need to support any georegulation effort with authoritative standards for the data on which it is predicated and also to provide an underlying model which is defined both for land and sea equally. Georegulation, therefore, naturally leads to a concept of "extension" of data for the reasons outlined above, "harmonized with but independent of" the data representing the legal basis of such legislation. The georegulation concept, and standards to support it is an area currently under development and a key one for MSDI interests and Marine Spatial Planning.

5.15.1. Implementation

In order to maximise the use of such extensions the requirement was implemented by means of a "generic" extension mechanism. This allows a basic aggregation of feature types from different domains grouped in their individual domains within the same GML dataset, identified and partitioned via namespaces.

The Example data below shows how core s121 data can be mixed with other domains without changing it. The extension mechanism works as follows.

The S-121 content is taken from the main S-121 GML Schema and only the feature content and information types are preserved (along with any relationships also defined). This is embedded in the file s121.xsd.

- A dataset element representing a "generic" feature collection is defined within file dataset.xsd.

- "Extensions" are modeled independently. Two are presented by way of example, Canada ONG (and other) data types ext.xsd and fishing limits (based on a prototype from UK) fsh.xsd.

- A single integrated schema file georeg.xsd includes the dataset.xsd, the extensions, and defines a single group called "features," which contains all the features to be aggregated, imported from the other imported application schemas: "s121", "ext" and "fsh" in the pilot's case. Note that any arbitrary selection of feature types is possible, allowing a customized subset of component feature types should it be required. Using XML include for the dataset elements rather than importing (used for each subdomain) allowed preservation of namespaces, crucial to the requirement of the extension mechanism.

- This top level xsd also defines the dataset element "Dataset," which defines dataset metadata and the enclosing "members" element, which contains the individual feature data. Each

extension's namespace defines the domain of each of its members and keeps them partitioned from one another. References provide means to link features together within the same dataset - additionally, information types can be shared between features without enabling modification of the referenced features

This allows a combined dataset to be put together as shown in the following diagram (the components of the schema are defined in the documentation for the schema itself).
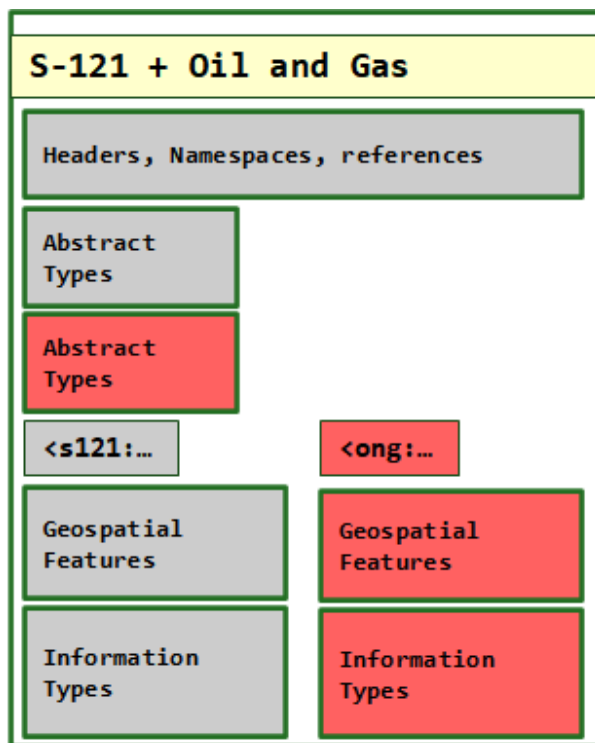

Figure 12. Schema Extension Combined Dataset

This mechanism allows GML Data such as an S-121 feature below:

```
<s121:BaselinePoint gml:id="Location_56">
        <s121:label/>
        <s121:collectionName>CA</s121:collectionName>
        <s121:featureName>
            <s121:displayName>false</s121:displayName>
            <s121:Language>en</s121:Language>
            <s121:Name>Cape Fox</s121:Name>
        </s121:featureName>
        <s121:context>Area 2</s121:context>
        <s121:releasabilityType>Controlled</s121:releasabilityType>
        <s121:interpolationType>densification</s121:interpolationType>
        <s121:pointType>defined</s121:pointType>
        <s121:geometry>
            <gml:Point gml:id="Location_56-0" srsName="urn:ogc:def:crs:OGC:1.3:CRS84"
                srsDimension="2">
                <gml:pos>-52.88972222 47.86138889</gml:pos>
            </gml:Point>
        </s121:geometry>
    </s121:BaselinePoint>
```

to be combined with ONG elements such as

```
        <ong:ongGridPoint gml:id="gp1">
            <ong:collectionName>ONG</ong:collectionName>
            <ong:featureName>
                <ong:language>en</ong:language>
                <ong:Name>ong Grid point 1...</ong:Name>
            </ong:featureName>
            <ong:geometry>
                <gml:Point gml:id="Location-ONG1" srsName="urn:ogc:def:crs:OGC:1.3:CRS84"
                    srsDimension="2">
                    <gml:pos>-52.99138889 48.58888889</gml:pos>
                </gml:Point>
            </ong:geometry>
        </ong:ongGridPoint>
```

Within a single dataset, i.e.

```
<?xml version="1.0" encoding="UTF-8"?>
<georeg:Dataset xmlns:gml="http://www.opengis.net/gml/3.2"
    xmlns:s100_profile="http://www.iho.int/S-100/profile/s100_gmlProfile"
    xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:gmlce="http://www.opengis.net/gml/3.3/ce"
    xmlns:s100="http://www.iho.int/s100gml/1.0" xmlns:s121="http://www.iho.int/S-121"
    xmlns:ong="http://nrcan.gc.ca/ong" xmlns:fsh="http://iho.int/fsh"
    xmlns:georeg="http://www.iho.int/georeg" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.iho.int/georeg georeg.xsd" gml:id="dataset04">

    <georeg:metadata>
        <georeg:DatasetIdentificationInformation>
            <s100:encodingSpecification/>
            <s100:encodingSpecificationEdition/>
            <s100:productIdentifier/>
            <s100:productEdition/>
            <s100:applicationProfile/>
            <s100:datasetFileIdentifier/>
            <s100:datasetTitle/>
            <s100:datasetReferenceDate>2020-01-01</s100:datasetReferenceDate>
            <s100:datasetLanguage>en</s100:datasetLanguage>
            <s100:datasetTopicCategory>oceans</s100:datasetTopicCategory>
        </georeg:DatasetIdentificationInformation>
        <georeg:DatasetStructureInformation/>
    </georeg:metadata>
    <georeg:members>
        <s121:BaselinePoint gml:id="Location_69">
            <s121:label/>
            <s121:collectionName>CA</s121:collectionName>
            <s121:featureName>
                <s121:displayName>false</s121:displayName>
                <s121:Language>en</s121:Language>
                <s121:Name>Whalesback</s121:Name>
            </s121:featureName>
            <s121:context>Area 2</s121:context>
            <s121:releasabilityType>Controlled</s121:releasabilityType>
```

```
            <s121:interpolationType>densification</s121:interpolationType>
            <s121:pointType>defined</s121:pointType>
            <s121:geometry>
                <gml:Point gml:id="Location_69-0" srsName="urn:ogc:def:crs:OGC:1.3:CRS84"
                    srsDimension="2">
                    <gml:pos>-52.99138889 48.58888889</gml:pos>
                </gml:Point>
            </s121:geometry>
        </s121:BaselinePoint>

        <ong:ongGridPoint gml:id="gp1">
            <ong:collectionName>ONG</ong:collectionName>
            <ong:featureName>
                <ong:language>en</ong:language>
                <ong:Name>ong Grid point 1...</ong:Name>
            </ong:featureName>
            <ong:geometry>
                <gml:Point gml:id="Location-ONG1" srsName="urn:ogc:def:crs:OGC:1.3:CRS84"
                    srsDimension="2">
                    <gml:pos>-52.99138889 48.58888889</gml:pos>
                </gml:Point>
            </ong:geometry>
        </ong:ongGridPoint>
    </georeg:members>
 </georeg:Dataset>
```

This dataset is valid according to GML rules against the schemas which define it, namely (taken from the example):

- georeg.xsd ("GeoRegulation") the top level defined XSD which defines complete set of features constituting the extensions (so, the content exposed from each of the component parts). It also includes:

- dataset.xsd, defines the feature Collection type "Dataset;"

- s121.xsd, containing only the feature content for the S-121 standard;

- ext.xsd containing feature content for hydrocarbons; and

- fsh.xsd, feature containing an elementary model for fishing limits.

Although a prototype mechanism designed to implement the requirement within the OGC Pilot, this simple mechanism of partitioning domains via separate schema files and namespaces has a lot of potential for broader uses of data and in the georegulation and MSDI domains. Potential illustrative use cases are as follows

- Modeling of regulatory regimes which refer to S-121 features in the same dataset. So, for instance fishing limits can include sections of S-121 EEZ outer limits by reference without the need to (potentially inaccurately) include such features in separate S-121 datasets.

- Preservation of the all-important links to legislative source, multi-lingual naming, and textual representation of position within combined datasets.

- Independence of modeling and schemas between domains. It is possible for domains to independently model their own features without them colliding in the combined GML. The import mechanism keeps the features separate from each other.

- The addition of any number of domains to data representing broader cadastral management. The mechanism specified only requires individual features to be specified. The feature collection can be arbitrarily extended to cover each domain so a dataset representing every category of data within a cadastral system can be assembled from its component parts.

5.15.2. Example Data

The project received data representing two use cases where S-121 data in a dataset was "extended."

- The CHS GDB dataset which contained data for ONG grid areas, Sections and Units, ONG administrative areas, wells, and grid points in addition to other regulatory areas, CBMP Arctic Marine areas and AMSAIIC Significant Areas. A schema (ext.xsd) is defined in the project which models the attributes contained in this domain. The assigned namespace is "ong" but this could be changed. The data was pre-defined and used a feature content identical to that of some of the s-121 features, therefore no new modeling was required.

- UKHO supplied data for fishing limits, the sources for them, and their extents. Additionally, auxiliary examples of limit data were provided to illustrate the likely requirements, supported by discussion with other sponsors on established norms in this important area.

Some example datasets were produced during the course of the Pilot but a single large scale extended dataset has not been produced due to resource constraints and the focus on refining the core S-121 schema for the LADM implementation. The project outputs consist of the georeg.xsd schema which contains the definitive S-121 content as baselined in the GML Application Schema, and the extension data for the CHS GDB features. A full transformation will be produced for future discussion and demonstration through OGC events.

5.15.3. Consumption of Extended Data

As the extended data is consistent GML the consumption of it by OGC conversant clients is unaffected. The two clients tested were Safe Software's FME and the QGIS plugin for S-121 consumption, S121Loader. The only issue encountered when processing extended data was related to the open source GDAL tools (used as the foundation for many of QGIS's data import functionality).

In order to parse data from the extended georeg schema directly, GDAL requires specification of all component schemas. The knock-on effect of this is that the default QGIS plugin for complex GML (the GMLAS plugin) does not import data directly. The S121Loader plugin however, contains a set of schema file links which enable the combined dataset to be imported and shown within QGIS along with all the relevant attribution and details as shown in the following image.
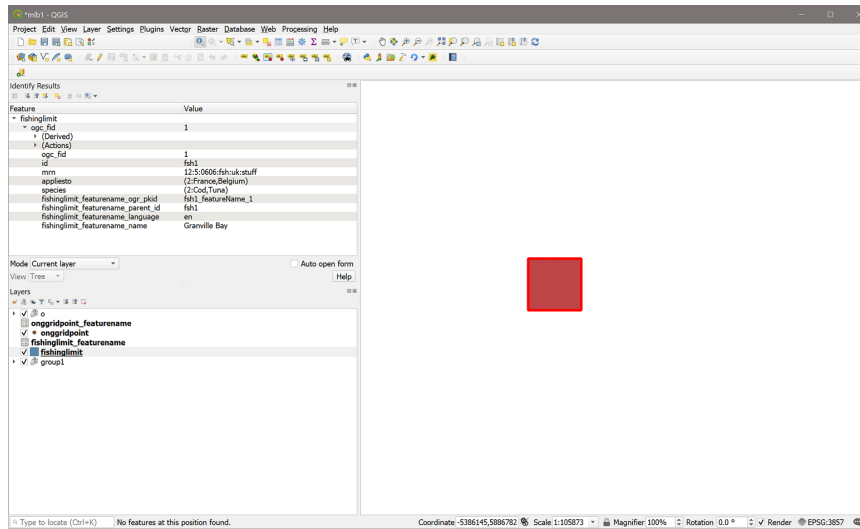
Figure 13. Relevant Attribution and Details Shown in QGIS Application

5.15.4. Further Development

The areas for future development of this capability are identified as follows.

- Complete examples with the Canadian dataset. This is a good example as the features are already modeled and share much of their characteristics and requirements with the existing S-121 data extensively processed during the project. This will be implemented for the upcoming OGC demonstrations along with its consumption in various clients.

- Modeling and examples for fishing limits domain data. An example is output in the project "fsh.xsd" (shown in the simple example dataset and its import in the previous section). Fishing limits are a significant use case for the future within the georegulation domain and have their own set of individual challenges. Examples from this domain will also be presented to the upcoming OGC meetings by way of demonstrating the project's outputs.

- Outputs to the S-100WG will demonstrate the conceptual advantages of having multiple data types within a single GML dataset and some of the benefits of that, particularly in the context of S-121. It is hoped that, ultimately, as part of a more general revision of Part 10b that the feature collection element "Dataset" could be defined as a standalone xsd within the S-100 profile and could then be included separately. S-100 itself would then facilitate the construction of multi-themed GML datasets. This ability to put modular schemas together could then be part of the broader S-100 framework. A future S-100 meeting will consider such developments alongside the other observations on the S-100 profile identified in the Pilot.

- A communication with the GDAL development community will be ongoing, to examine how to resolve the "multiple schema definitions" issue within the GMLAS driver for complex GML. If this can be resolved in a future edition of GDAL then the existing GMLAS plugin would be a usable solution for many implementers of the standard and the extensions.

Chapter 6. Data Creation - Conversion of Sponsor Data to GML

6.1. Introduction

This section of the ER looks in detail at how various clients, processes, and participants worked to convert sponsor data in its original form to S-121 conformant GML. While many participants worked with the S-121 GML, and a number of client GIS tools were tested in this context, the primary GIS application used to generate and edit S121 GML was FME from Safe Software. The contents of this section should be contrasted with sections on reading/ingesting GML. This section concentrates primarily on the ability to generate S-121 conformant GML and the use made of associated tools by the project participants.

During the pilot the focus on data creation really stepped up in Phase 2. The GML Schema reflects the content of the IHO Feature Catalogue and product specification and, towards the end of Phase 1 of the project it became apparent that many elements of the GML Schema had not been exercised fully. So, the Phase 2 focus was readjusted to concentrate on an iterative process for developing GML data which tested the major components of the Schema (and underlying specification).

6.2. Challenges

The challenge in making GML data in this phase of the project is partly the transformation from the sponsor database structure which can range from something completely bespoke to something that implements the features and structure of the S-121 model closely. The GML schema contains complex attribution (most notably featureName which implements multi-lingual naming) and locationReference, used for the multiple representation of positions as described in legislative text. So the challenge in making GML is being able to represent the features, their attribution (all according to the intervening S-100 profile), and the relationships between them according to those stipulated in the schema.

This section examines the data transformation process from the point of view of Commercial Off-The-Shelf (COTS), open source, and bespoke software packages.

6.3. COTS - Safe Software, FME

6.3.1. Requirement Description

The primary requirements for the MLB Pilot GIS Application were as follows: *Open, create, and edit S-121 data including all of the direct attributes and information objects (Rights, Restrictions, Responsibilities, Parties, and Governance) and associated S-100 defined metadata. The implementation must support all of the capabilities of the standard, including display, editing, export, import, and conversion of the data between formats. The GIS Applications will read raw data and convert it to S-121 in GML, based on the GML Application Schema.*

6.3.2. GIS Application Description: FME

Safe Software's FME was the primary GIS application used to generate S121 GML from the sponsor source databases using the S121 GML application schema developed within d3-GML-Application-Schema. FME is a model-based data integration and automation platform. FME Desktop includes Workbench which can be used to author translations between any of the 400+ data formats and web services that FME supports. FME has strong support for both open standards, industry and proprietary data formats and as such can serve as a bridge between open standards and leading GIS or geospatial application vendor formats.

The most comprehensive translation conducted in Phase 2 was from the CHS extended continental shelf geodatabase to S121.gml using the S121.xsd version 4.0 GML application schema. The CHS data required an implementation of multiple relationships and joins between parsed features which was challenging to integrate with the feature creation itself. Additionally, verifying such structures is difficult when the web of relationships themselves represent complex structures of ownership, source and responsibility. This remains an open problem in terms of data creation and perhaps an opportunity for bespoke software to fill a gap.

FME also was used to perform the full range of GIS functions listed in the requirements above, including display, edit, import, export, conversion and validation.

6.3.3. Requirement Summary

- S121 compliant GML sufficient to drive the GML to Human Readable text generation compliant with DOALOS requirements.
- S121 GML that captures all the relevant content available from the sponsor database for the selected test area – in this case Area 2 off the east coast of Newfoundland and Labrador.
- S121 GML produced must validate against the S121 application schema supplied – in this case S121.xsd version 4.0.
- Support downstream applications such as GML clients and WFS services and their clients.

6.3.4. Inputs

CHS geodatabases:

- Canadian Extended Continental Shelf: S-121_CanadianExample_20200526.gdb.zip. Additional information is available online here: https://knowledge.safe.com/articles/112578/ogc-maritime-limits-and-boundaries-s-121-gml-pilot.html.

6.3.5. Process Steps

FME uses a model oriented spatial ETL transformation process to convert data from one format and structure to another. The source database and destination GML data models are entirely schema driven so the work is largely oriented around transforming and mapping the data structures from the source to match the requirements of the destination. Once completed this transformation model, also known as an FME workspace, can be run using set of parameters via a user interface, command line or via a service.

This approach allows for both rapid prototyping and automation without the need for any coding. It was also conducive to an agile, continual development approach which was a requirement in this project since the schemas involved were changing right up until the last week of the pilot.

The main process steps in the transformation CHS_Geodb_to_S121gml_v1.1.fmw workflow are as follows:

1. Capture run-time parameters

2. Read source geodatabase tables and relations

3. Modify values by default language choice

4. Use relations to perform joins and embed related keys

5. For one to many relations, embed relation arrays into each feature's external xlink:href reference lists

6. Set required GML feature properties

7. Set required GML geometry properties

8. Perform required schema mappings

9. Apply language defaults to monolingual fields such as titles and descriptions

10. Perform feature type mapping

11. Generate metadata

12. Generate S-121 GML

13. Validate output S-121 GML against S121.xsd

An image of the resultant workspace baselined at the conclusion of the project is shown below.
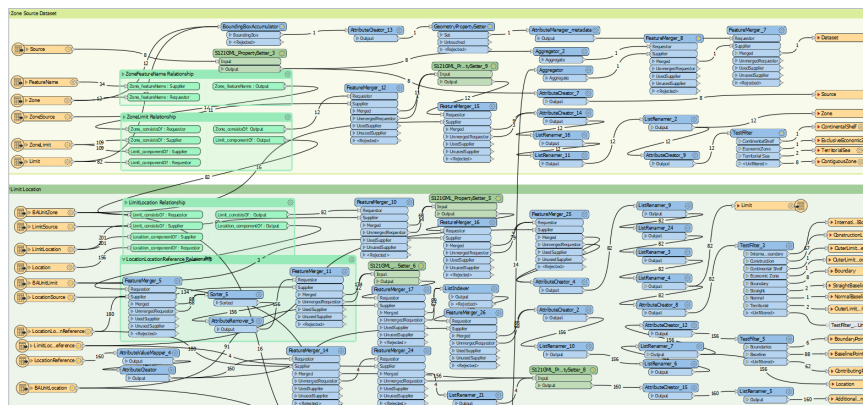
Figure 14. FME data conversion workspace (all elements)

GML Writer parameters critical to production of complaint results –> published parameters are highlighted in purple
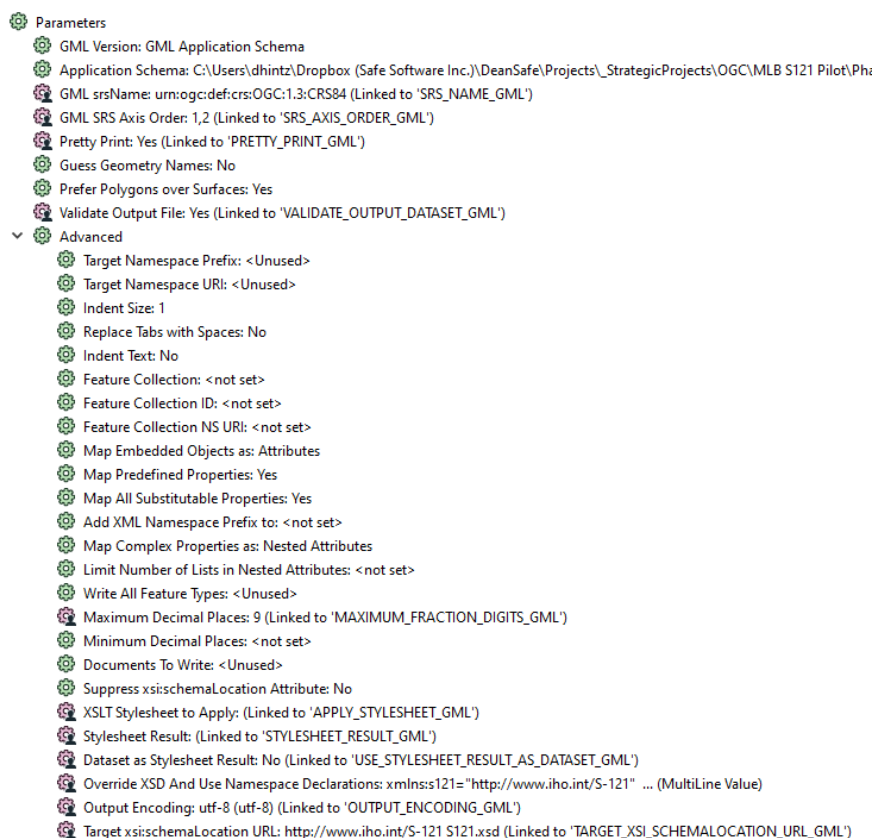
Figure 15. FME Workspace GML writer parameters

6.3.6. Outputs

CHS S121 GML results can be accessed at the project repository.

The latest versions posted are:

- CHS S-121 Canadian Sample GML English - CHS_S121_v1.2_en.gml; and
- CHS S-121 Canadian Sample GML French - CHS_S121_v1.2_fr.gml.

The script used to perform the translation is called an FME workspace. The FME workspaces used to generate the above S121.gml are available.

The most recent FME workspace (CHS_Geodb_to_S121gml_v1.1.fmw) and schema (S121.xsd) used to generate S121.gml is 36829-chs-geodb-to-s121gml-v11-fmw.zip.

6.3.7. Challenges

There were several difficulties related to transforming data from a relational database to S121 GML. First of all the data models are very different. The source database is composed of a number of flat tables that use external relation tables to inter-relate them. The GML has a nested XML structure where many of the related objects are embedded within parent objects.

The S100 family of schemas uses AbstractInfoType for many of the non-geometry feature types, rather than AbstractFeatureType. Many tools and GIS applications designed to work with GML tend to use AbstractFeatureType exclusively. Safe Software needed to modify the GML reader / writer to support AbstractInfoType and AbstractGML in order to support the S-121 GML.

The form of the GML generated needed to consider downstream applications. For example, while still valid S121 GML, there are conventions around axis order and schemaLocation that can make it easier for third party GML client applications to read S121 GML. Also, the type of CRS used needed to be adjusted. By default, FME reproduced the EPSG CRS as read from the database. Overriding this with the OGC CRS definition "urn:ogc:def:crs:OGC:1.3:CRS84" made it easier for clients to recognize and also preserved the correct axis order.

Once the basic schema structure was fully supported, the next main phase of translation development centered around the actual translation and reproduction of the source geodatabase as S121 GML. The translation sought to preserve as much as possible all data from the database that was relevant to S121. To provide a practical benchmark of minimal functionality, a basic goal was to reproduce the geodatabase as GML sufficiently to support the generation of Human Readable Text for Area 2.

As it happened, the geometries were relatively trivial to migrate. What became a much greater challenge was migrating the relationships from the geodatabase relationship classes to the xlink references embedded in most features. S121 constitutes a rather complex topology of relationships from Location to Limit to Zone to BasicAdministrativeUnit to Governance and back again. It was a significant challenge to reproduce all these relationships as derived from the source geodatabase. These relationships also had to be defined in such a way that the various client tools were able to discover and resolve them. For example, W3C requires the '#' prefix for xlink:hrefs in order to read them. Initially this prefix was missing. Correcting this allowed clients such as CARIS to better resolve the links.

Finally, one of the final challenges was implementing multi-lingual support, for example French and English. Gaps in multilingual support both in the schemas and in the source geodatabase structure and content meant that there were changes to both right up until the last week of the pilot. Because of this some changes were required in the model itself to edit the data during translation in order to expedite corrections to the data sufficient to produce French language Human Readable text.

Other challenges and how they were mitigated are explored in more detail elsewhere within this Engineering Report in Onward Uses for Generated Data, Client SDI, relating to S-121 clients and Validation of Data.

6.3.8. Results

In the end the FME GIS Application was able to generate both English and French language S121 compliant GML datasets with content sufficient to support Human Readable DOALOS text for the CHS Continental Shelf Area2 test dataset. The S121 GML output was also used to drive the other dependent MLB Pilot components such as the CSW, WFS and WMS services.

6.3.9. Georeg Extension data

An important additional goal of Phase 2 of the MLB pilot was to test the development of S-121 GML extensions. In particular, additional data was added to the CHS geodatabase related to fisheries zones and oil and gas exploration zones. One objective was to allow for the selection of some S-121 feature types and combine this with some feature types from the extension schemas.

In order to support this, a modular approach was taken so that the dataset, S121 elements and Georeg elements were all defined in separate schemas with separate namespaces. This introduced some complexity for FME's namespace handling which was addressed in some of the software modifications listed previously.

Ultimately it was possible to demonstrate a set of modular schemas that could support the generation of a combination of S121 and georeg elements within a georeg:dataset all using constructs conformant with S100 types.

Georeg extension GML showing combination of S121 and georeg elements

```xml
<georeg:dataset xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gts="http://www.isotc211.org/2005/gts" xmlns:gmlce="http://www.opengis.net/gml/3.3/ce"
xmlns:gsr="http://www.isotc211.org/2005/gsr" xmlns:s121="http://www.iho.int/S-121"
xmlns:s100_profile="http://www.iho.int/S-100/profile/s100_gmlProfile" xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:georeg="http://www.iho.int/georeg" xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:s100="http://www.iho.int/s100gml/1.0"
xmlns:com="http://www.iho.int/S-10Xcommon" xmlns:ong="http://nrcan.gc.ca/ong"
xmlns:gss="http://www.isotc211.org/2005/gss" gml:id="dataset04" xsi:schemaLocation="http://www.iho.int/georeg
georeg.xsd">
<georeg:members>
<s121:BaselinePoint gml:id="Location_83">
   <s121:collectionName>CA</s121:collectionName>
   <s121:featureName>
     <s121:displayName>false</s121:displayName>
     <s121:language>en</s121:language>
     <s121:Name>Red Island 4</s121:Name>
   </s121:featureName>
   <s121:context>Area 2</s121:context>
   <s121:startLifespan>2019-01-01</s121:startLifespan>
   <s121:endLifespan>2019-12-31</s121:endLifespan>
   <s121:releasabilityType>Controlled</s121:releasabilityType>
   <s121:interpolationType>densification</s121:interpolationType>
   <s121:pointType>defined</s121:pointType>
   <s121:geometry>
     <gml:Point gml:id="Location_83-0" srsName="urn:ogc:def:crs:OGC:1.3:CRS84" srsDimension="2">
       <gml:pos>-55.516111111 50.721666667</gml:pos>
     </gml:Point>
</s121:geometry>
   </s121:BaselinePoint>
...
<ong:ongSection gml:id="Zone_11">
   <ong:collectionName>CA</ong:collectionName>
   <ong:featureName>
     <ong:language>en</ong:language>
     <ong:Name>46-40 N 48-15 W 40</ong:Name>
   </ong:featureName>
   <ong:startLifespan>2019-01-01</ong:startLifespan>
   <ong:endLifespan>2019-12-31</ong:endLifespan>
   <ong:releasabilityType>Controlled</ong:releasabilityType>
   <ong:geometry>
   <s100:Polygon gml:id="Zone_11-0">
     <gml:exterior>
       <gml:LinearRing>
               <gml:posList>...</gml:posList>
       </gml:LinearRing>
     </gml:exterior>
   </s100:Polygon>
   </ong:geometry>
</ong:ongSection>
```
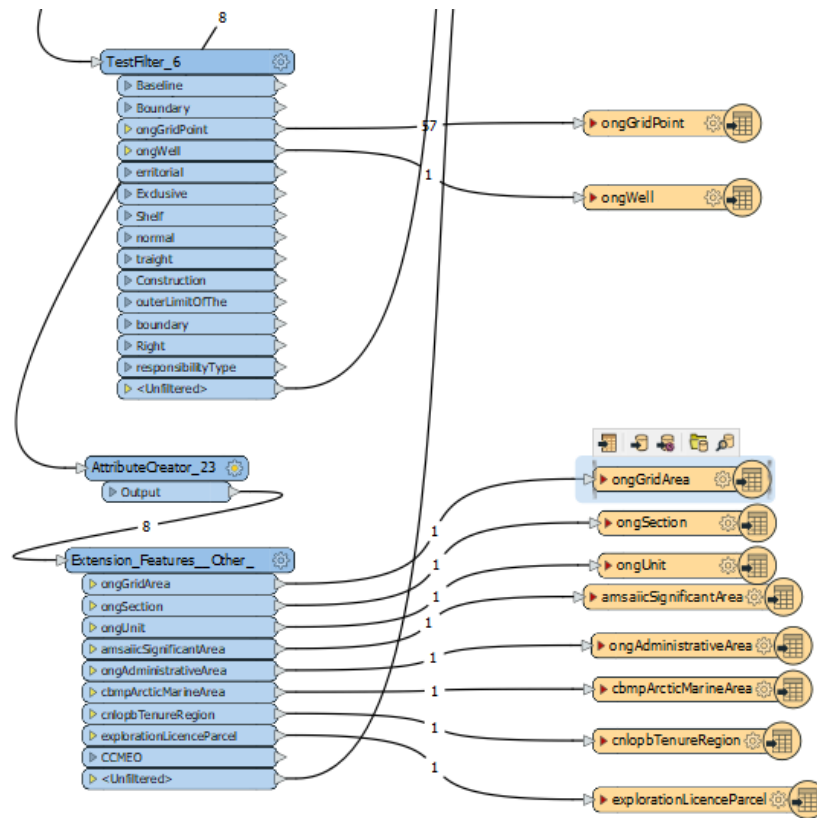
Figure 16. Selection of Georeg elements for writing to CHS_S121_v0.71.xml and CHS_S121_v0.73.xml
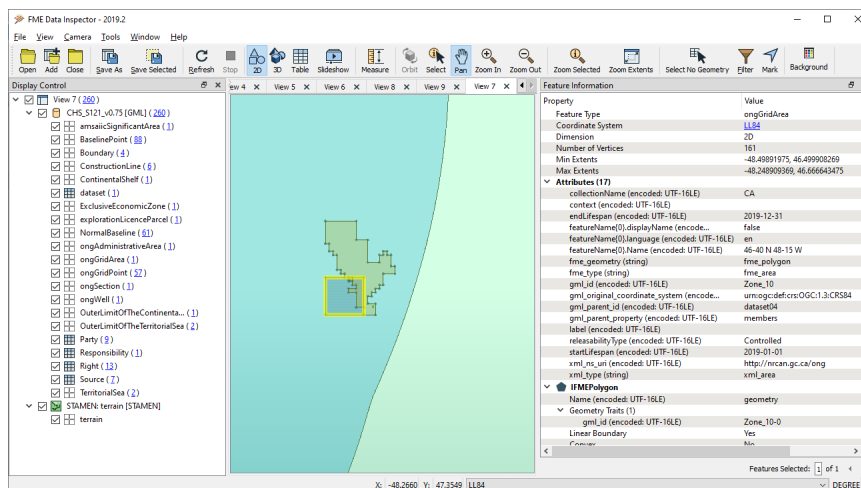
Figure 17. Georeg extension combined with S121 elements – ongGridArea selected [Note]

In Extensions of the GML Application Schema the issues and results of extension development and testing are described in greater detail. Because the core S-121 schemas and source geodatabase underwent changes up to the last week of the pilot, it will be left to follow-on work to further test and explore the possibilities of S-121 and S-100 extensions.

One useful lesson learned was the modularization of the S100 dataset definition as a separate schema. In this way the parent schema, in this case georeg.xsd, can include the dataset schema and thereby create a dataset container using the same namespace as the parent schema. This is how dataset:georeg is generated in this case. On the other hand the ext.xsd, fsh.xsd, and s121.xsd schemas are all imported in order to preserve their respective namespaces. This modular approach was developed because the S100 schemas were not prescriptive enough to give specific guidance in terms of how to define datasets for S100 family profiles such as extensions, a topic highlighted in the outputs to the S-100WG.

The georeg datasets are available at the project repository.

The FME georeg generation and testing workspaces can be found here:

- https://knowledge.safe.com/articles/112578/ogc-maritime-limits-and-boundaries-s-121-gml-pilot.html

6.3.10. Integration of Data Creation with HR Text Processing

As a quick integration test, it was decided to test the XSLT created by Geomatys for human readable text generation to see if it could be invoked directly from the FME workspace that was used to generate the S-121 GML.

FME's XML and GML writers have an option that allows them to invoke an XSLT script to provide for post processing of the output XML. Normally this is used to reformat the GML or XML to fit a specific schema or application. However, in this case the goal is to use the XSLT to reformat the S-121 GML into human readable text. Note that the XSLT parser employed by FME is limited to XSLT version 1.0.

The XSLT script selected for this test was: Canadian Legislative Output EN.xsl. The GML writer parameters for XSLT and HTML output can be seen in Figure 11 below.
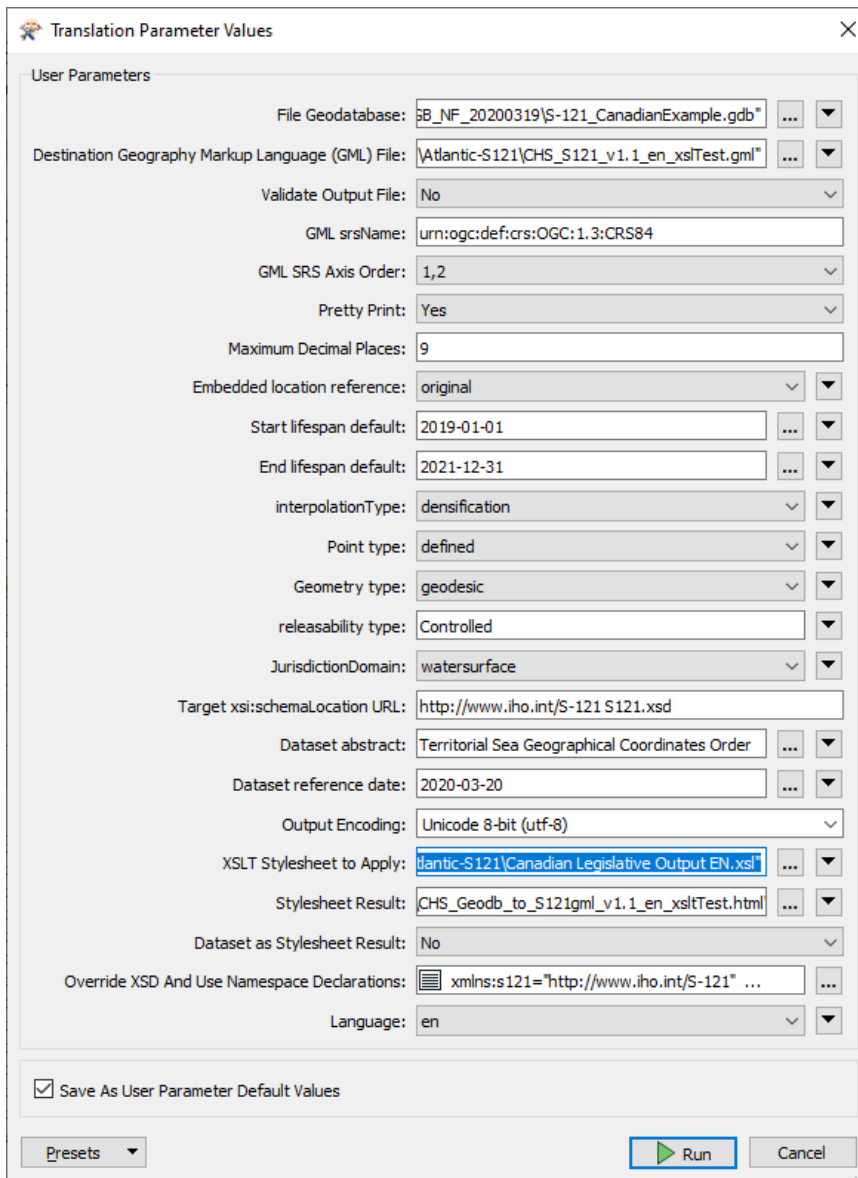
Figure 18. CHS_Geodb_to_S121gml workspace configured to invoke XSLT script to generate HR HTML from S121 GML.

The results were impressive. The output is indistinguishable from what was produced by Geomatys. This demonstrates that the human readable text generation capability developed by Geomatys can be readily integrated with FME's S-121 GML generation so that data can be transformed directly from the sponsor's source geodatabase all the way to HR Text without any manual intervention.

Figure 19. Human readable text generated directly from CHS source geodatabase using an integrated chain of FME geodb to S-121 GML workspace combined with Geomatys' HR XSLT script.

An integrated FME workspace and XSLT demo for S121 GML and HR Text can be found here:

- https://knowledge.safe.com/articles/112578/ogc-maritime-limits-and-boundaries-s-121-gml-pilot.html

6.3.11. QGIS GIS import

The S121Loader plugin developed by IIC provides an import of S-121 data in Complex GML for use within QGIS. The plugin is based on the GMLAS plugin already available within QGIS but enables import of the GML Extension schema as well as the native S-121 data.

The plugin GUI is shown below.
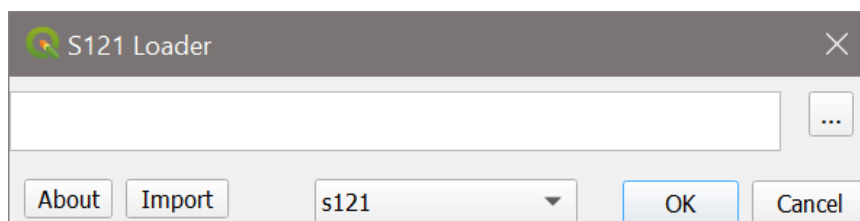
Figure 20. QGIS S121 Loader plugin

Currently configured to load both full S-121 and Extended "georeg" S-121, the plugin is based on the existing GMLAS plugin available for QGIS v3.4.5 and above. The methodology loads the complex GML in its native form into QGIS rendering complex attributes and information types as non-spatial tables. QGIS is able to form joins against those tables forming an integrated GUI for display and editing of the data content.

The plugin is capable of outputting GML data in addition to its import and viewing but this is restricted to table editing and edits of data already in place. New feature addition is difficult particularly where related to joins (inserted by the information types) and metadata editing. The potential certainly exists (due to the GMLAS driver) to grow the plugin via an open source effort into a bespoke S-121 GML client but this would require development resources.

6.3.12. Bespoke GML Creation

6.3.12.1. S-121 GML API Library

The S-121 JAR file provides an API to the S-121 GML format for Java based application programs and enables verification of GML format as well as some elementary dataset metadata editing capabilities. This JAR file is auto-generated from the S-121 XSD and allows application developers to develop interfaces to datasets in the standard using the JAXB interface to the GML.

This allows simple GUIs for read/edit and creation of GML to be constructed, e.g

```
    CollectionType aM = getAllFeatures(openfilename);

    public static CollectionType getAllFeatures(String fn)
    {
        try
        {
            //unmarshal back to object
            JAXBContext ctx = JAXBContext.newInstance(ObjectFactory.class, net.opengis.gml._3.ObjectFactory.class,
_int.iho.s_121.ObjectFactory.class);
            JAXBElement<CollectionType> aM = (JAXBElement<CollectionType>) ctx.createUnmarshaller().unmarshal(new
FileReader(new File(fn)));
            mdl.getI().setaM(aM);
            return aM.getValue();
        }
        catch (Exception e)
        {
            e.printStackTrace();
        }
        return null;
    }
```

This enables the building of GUI elements for bespoke data editing and processing, e.g...
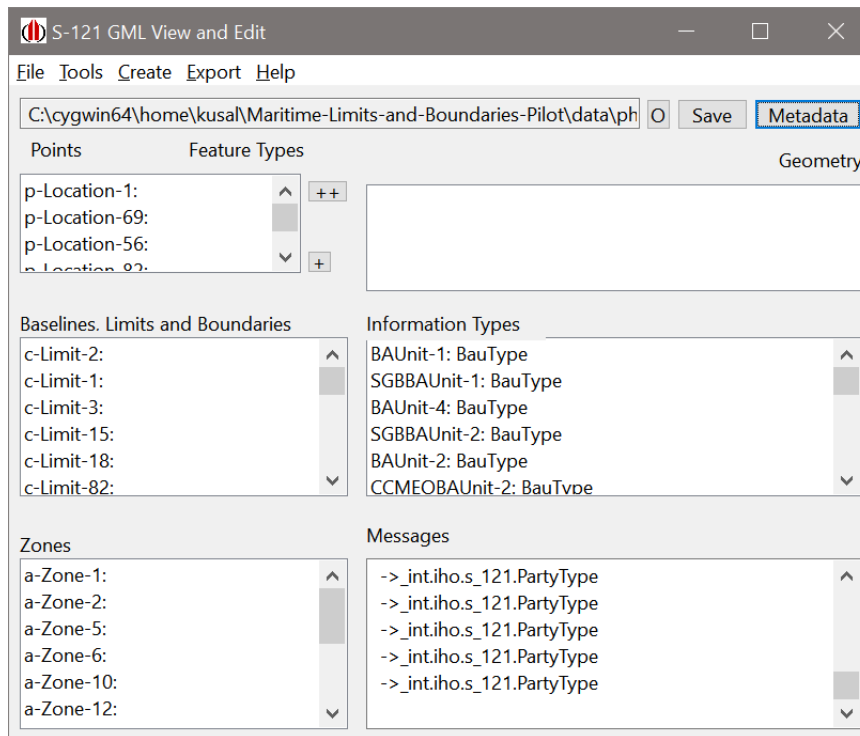
Figure 21. Bespoke Editor for GML data

During the pilot, during Phase 1 conformant GML was produced using the API generated but the primary GIS application used for Phase 1 and Phase 2 remained FME from Safe Software.

6.4. Validation of Data

6.4.1. GML Validation

Validation of created GML is an essential final step in the process of data creation. Two examples of validation are presented within this section of the ER, those implemented by Geomatys' bespoke tooling and use of Safe Software's FME toolkit for validation. The question of validation is a complex one. Validation functions at many levels of detail and against many rulesets. The two categories of validation presented within this section deal with

- Validation of S-121 GML against the normative GML and XML standards; and
- Validation of S-121 content post-transformation into the alternate Textual Encoding.

These are dealt with in detail within this section. It is worth noting, however, that S-121 itself provides several basic validation tests of content, not explicitly tested for here but which could be added to validation scripts in the future. A comprehensive and structured set of validation processes would be straightforward to build on top of the scripts/tools detailed here and could be the subject of a future development. S-121 could also enhance its coverage of validation tests outside basic formatting/structuring of content and into areas such as topological completeness, more complex encoding guidance, relationship validation, and others.

As S-121 matures in its adoption and rollout amongst data producers it is likely that validation requirements will increase. The IHO community has a well developed and comprehensive validation ecosystem for electronic charts which runs to many hundreds of tests carried out systematically by multiple stakeholders prior to ENCs being released for use within commercial

navigation systems. It may well be that S-121 could absorb some of the approaches from IHO S-101/S-57 in this respect.

6.4.2. FME Validation of Created GML Data

FME's GML writer has a validator option that verifies the output GML against the selected application schema, in this case, S121.xsd (the full S-121 GML Application schema). This results in a call to a third party XML validator provided by the open source Apache Xerces library. This ensures that the validation is conducted according to an independently and externally verifiable standard.

XML/GML validation can also be invoked using the XMLValidator transformer from within an FME workspace configured to read the GML file and application schema being tested against. In either case, any validation failures are detailed in the translation log. Or, validation can be called by the GML writer simply by setting the "Validate GML Dataset File" option to Yes - as in the following figures.
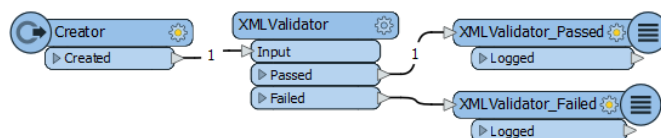
Figure 22. Validation in FME via embedded workspace entities
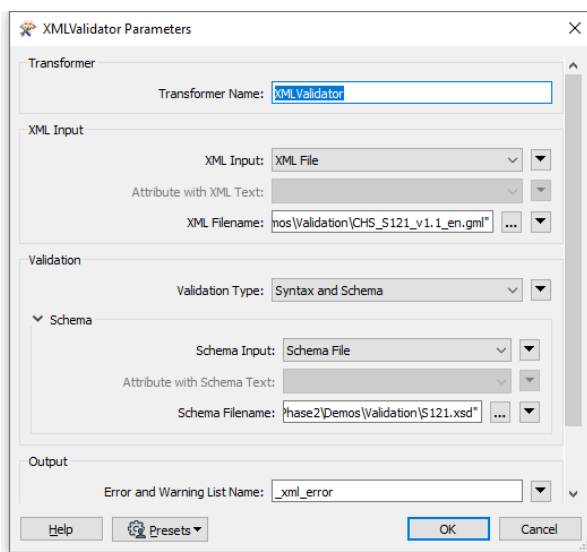
Figure 23. Validation in FME via runtime options

6.4.3. Geomatys Validation tools

Geomatys produced during the project a validation tool bespoke to S-121 which delivers a detailed range of validation services for created GML conformant with the S-121 schema. These tools are included in the project repository and run standalone.

```
This tool works using three configurations blocks.
[global parameters] [tool name] [tool parameters]
-ui         : Display graphical interface (if Geotoolkit JavaFX module is available).
-list -l    : Display list of available tools.
-help -h    : Display help for a tool, must be followed by the tool name.
-silent -s  : Silently execute tool (will not show the result).
Tool name   : can be authority.name or name alone if unique.
Tool params : can be found using -help for a given tool.

Examples
java -jar S121-tools.jar -list
java -jar S121-tools.jar -help validation
java -jar S121-tools.jar -help s121.validation
java -jar S121-tools.jar -ui
java -jar S121-tools.jar validation -file /path/to/file/s121.gml
java -jar S121-tools.jar validation -file "/path/with/ s p a c e s /to/file/s121.gml"
java -jar S121-tools.jar script -file "/path/with/ s p a c e s /to/file/script.lua"
java -jar S121-tools.jar asciidoc -file /path/to/file/s121.gml
java -jar S121-tools.jar xslt -file /path/to/file/s121.gml -xsl /path/to/file/transform.xsl
```
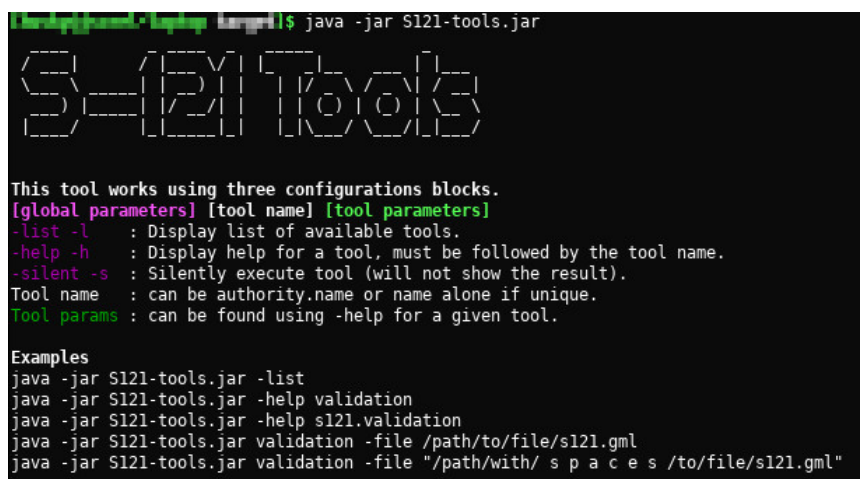
An example is shown in the following image.

Figure 24. Geomatys Validation Tool

In addition to the validation of GML data, Geomatys created bespoke data transformation tools for transformation of sponsor data to S-121 GML using Lua scripting. Lua is a scripting engine used extensively within S-100 for control of portrayal. These tools all could be adapted in the future as bespoke engines for transformation of database based repositories of S-121 data to conformant GML.

6.4.4. HR Text Validation

The primary requirements for the optional MLB Pilot HR Text Validator were as follows.

The HR Text Validator should be able to assess HRText for problems and inconsistencies when compared to the S121-GML. Since the HRText is always a subset of the associated S-121 GML, it is important to consider the GML as the baseline and then report on deficiencies found within the HR Text when comparing it to the S-121 GML. The HR Text Validator should generate a tabular report

listing problem elements and type of error or problem encountered.

There are two categories of validator presented within this ER representing the efforts of the participants within the project. Initial validation against the GML application schema was implemented by both Geomatys and Safe Software. An additional model of validation was also developed using FME's workbench by Safe Software in order to verify produced Human Readable Text against the source GML to assess differences. This "validation" process is designed to ensure that data has been converted to human readable text "correctly" by reading both the Text and GML S121 representations and comparing the features with their equivalents. Ideally, the HR Text Validator should be able to perform the following:

- consume S-121 GML data following the S-121 GML Application Schema;

- consume S-121 HR Text created according to the S-121 Annex B rules; and

- demonstrate comparison of features from both to assess commonalities and differences.

Safe Software developed the experimental S-121 HR Validator by constructing an FME data transformation model to read the S-121 HR Text, the S-121 GML and perform comparisons.

The basic approach for the HR Text is as follows;

1. Read S-121 GML using the S121.xsd application schema.

2. Read HR Text using text parsing methods.

3. Match HR Text elements with S-121 GML elements using identifier, name, latitude, longitude, and point geometries.

4. Generate a list of all elements within HR Text that are matching the S-121 GML.

5. Generate a list of all elements within HR Text that are not in the S-121 GML.

6. Generate a list of all elements within HR Text whose locations are outside a specific tolerance when compared with S-121 GML Locations.

7. Generate a list of all S-121 elements that are directly related to the known HR-Text elements but are not part of the HR-Text.

8. Generate a tabular report summarizing key differences and S-121 rule violations and where relevant, also display results in an inset map.

6.4.5. Initial Testing and Results: AU HR Text

The first step for building the HRTextValidator was reading the HRText. An early prototype of this was built to read 10-_Example_Output_29_Nov18.txt since at the time no CHS HR Text was available. The results of these early tests are shown below.
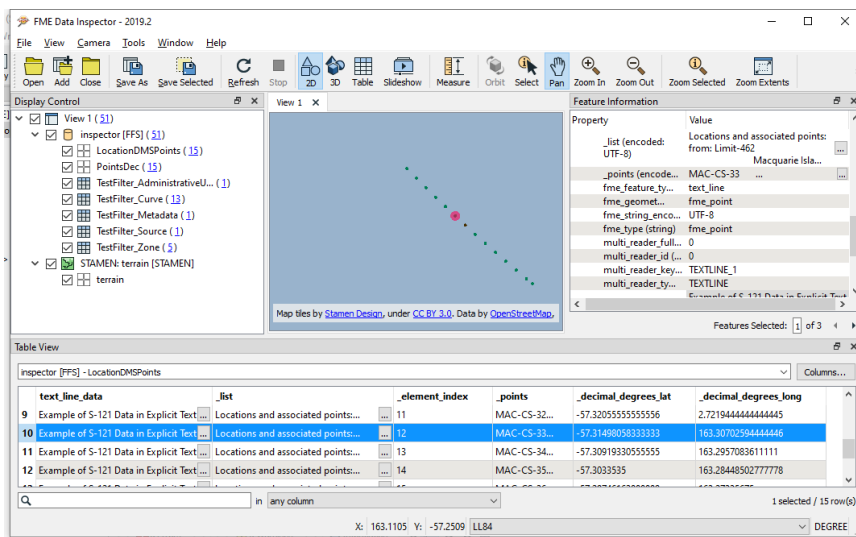
Figure 25. Example HR Text reading 10-_Example_Output_29_Nov18.txt
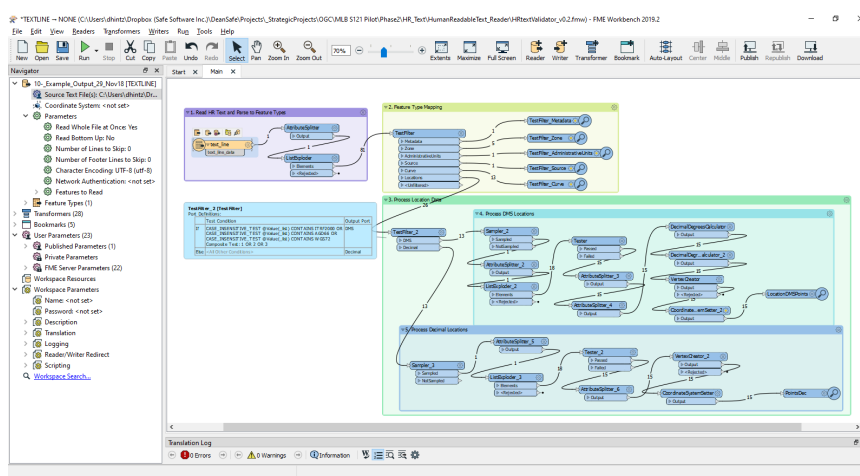
Figure 26. Initial HR Reader / Validator test FME workspace: HRtextValidator_v0.2.fmw

Initial results provided some promising progress in terms of reading the Human Readable (HR) Text. Given the consistent nature of the S121 HR Text and the limited number of feature types involved, it should not be particularly difficult to read the Locations, Limits and Zone information associated with a particular HR Text dataset. Figure 1 shows a line of Location points read from HR Text, displayed SE of Australia near Macquarie Island. Reading the S121 GML is already supported via the GML reader so no additional configuration is required.

However, once both the datasets are read, the scope for potential validation appears to be somewhat limited. It should be relatively straightforward to compare the Location points and names against the S-121 GML. The Limits and Zones can be checked by name but not by geometry since the Limit and Zone geometries are not stored in the HR Text form.

The real challenge lies in verifying the various relationships fundamental to the S-121 GML data structure. These relations are used to generate the HR Text. However, there is no facility in the HR Text to store most of these relations. The exception to this is the different document schedule sections. The schedule groupings could perhaps be used to derive which Zone or BAUnit some of the Locations belong to. However, it is very doubtful that all the relations could be derived in this way. Because of this an HRTextValidator is likely to have limited usefulness. It may be useful for detecting violations of the S121 HR Text rules and provide a way to verify that all the Locations in the HRText are valid. But beyond that, it is too early to say if more comprehensive testing would be

practical.

6.4.6. CHS S-121 HR Text Validation and Testing

Initial testing was performed on Australian HR Text because earlier in the Pilot no Canadian HR Text was available. Once CHS HR Text was available, the FME transform model was recalibrated to read the output of the HR Text XSLT Script which transformed the CHS Extended Continental Shelf S-121 GML to HR Text. Due to the constraints on time, resources, and other priorities, only detailed comparisons on Location were performed. Other feature types from the HR Text were parsed and read such as Metadata, Dataset, Limit, Zone, and Governance. However, only Locations were analyzed in detail.
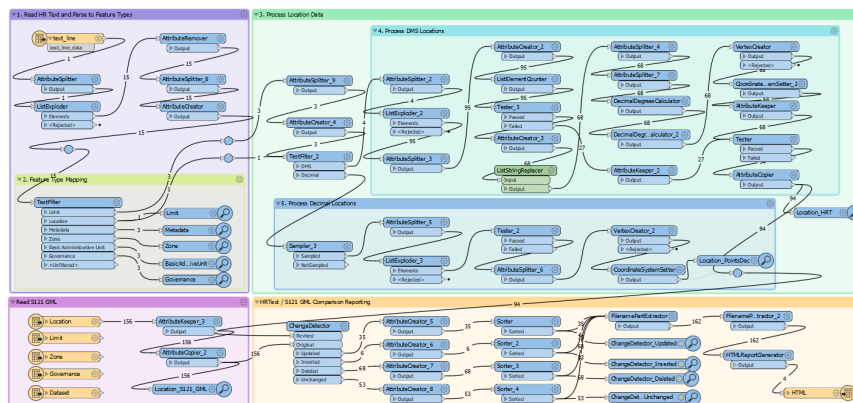
Figure 27. HRtext S121 GML Validator: S121_HRtextValidator_v1.0.fmw

The above workspace, and its supporting input files and sample output HTML reports are available in the attachment 'S121_HRtextValidator_v1.0.zip' at https://knowledge.safe.com/articles/112578/ogc-maritime-limits-and-boundaries-s-121-gml-pilot.html

While FME has many text reading, parsing and manipulation functions, raw text parsing is in principle more challenging than parsing structured text such as XML or JSON. Therefore a number of steps are involved to read the text file, split it up into text blocks by feature type, and then split by carriage returns to break up into individual rows. The text reading and feature type parsing occurs at the top left of the workspace above in the section with the TestFilter with outputs to Location, Limit, Metadata, Zone, etc.

Once this was accomplished, additional parsing was required to separate Location elements into separate features. Differences were then detected between DMS coordinates and decimal degrees and handled accordingly. The DMS coordinates were converted to decimal degrees before they could finally be applied to point geometry generation. The largest section of the FME workspace above - the green/blue bookmarked area - involves this text parsing and geometry extraction process.

It should be noted that the precise rules for HR Text generation as defined in the S-121 HR Text Annex (http://www.s-121.com/w/index.php/Documents#S-121_Standard_Components) are essential, not only for HR Text production, but also to HR Text consumption and conversion back into GIS features.

The bottom of the workspace is where the S121 GML is read and compared with the HRText Locations. The central transformer used to do this comparison is the ChangeDetector. This tool

compares HR Text Location features and S121 GML Location features using common gml_id key and assesses common attributes and point geometries for differences. A tolerance was provided to compensate for the fact that DMS values are only accurate to one second and are inherently less accurate than decimal degree points exported from a GIS. Sample ChangeDetector parameters are shown below.
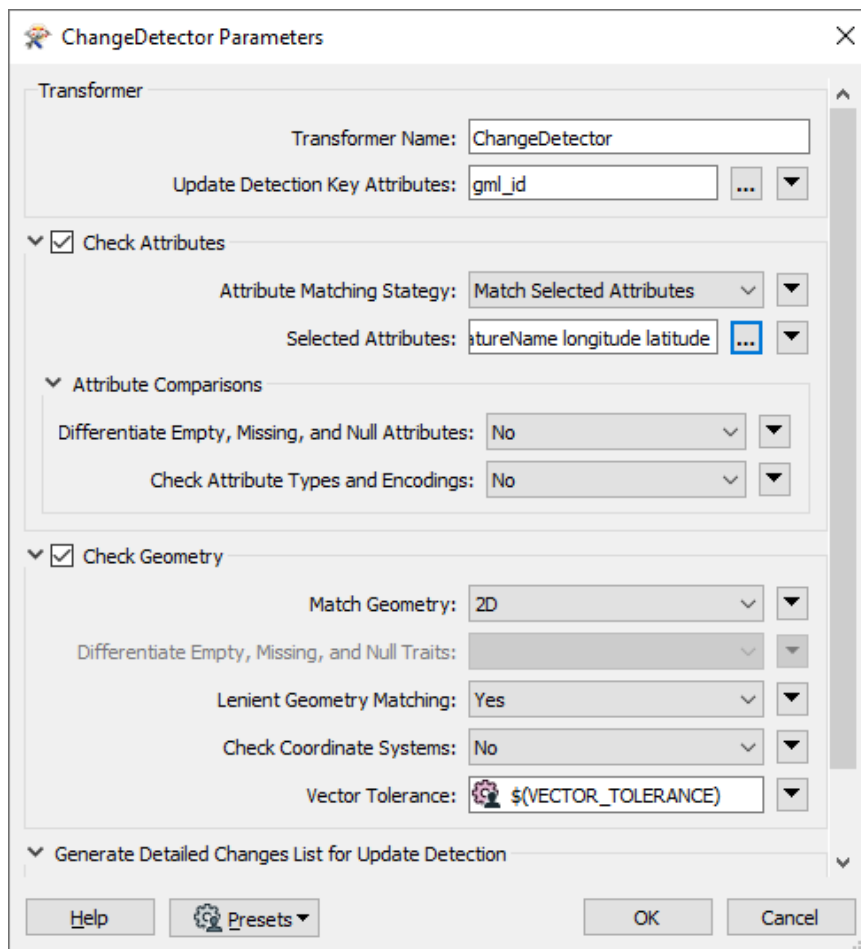

Figure 28. ChangeDetector configuration for HRText validation against S-121 features.

The ChangeDetector reads 2 inputs: Original and Revised. It then produces 4 outputs depending on the results of comparing the inputs with each other: Updated, Inserted, Deleted, and Unchanged. Each of the ChangeDetector outputs were assessed as follows:

- Updated: Location Differences HR vs GML;

- Inserted: HR Text Locations not in GML;

- Deleted: Unused GML Locations; and

- Unchanged: HR Text Locations Matching GML.

After this, the results are assembled into an HTML report using the HTMLReportGenerator. One report page is generated per result type. The following screen shots show some example outputs for matching, inserted and updated comparison results.
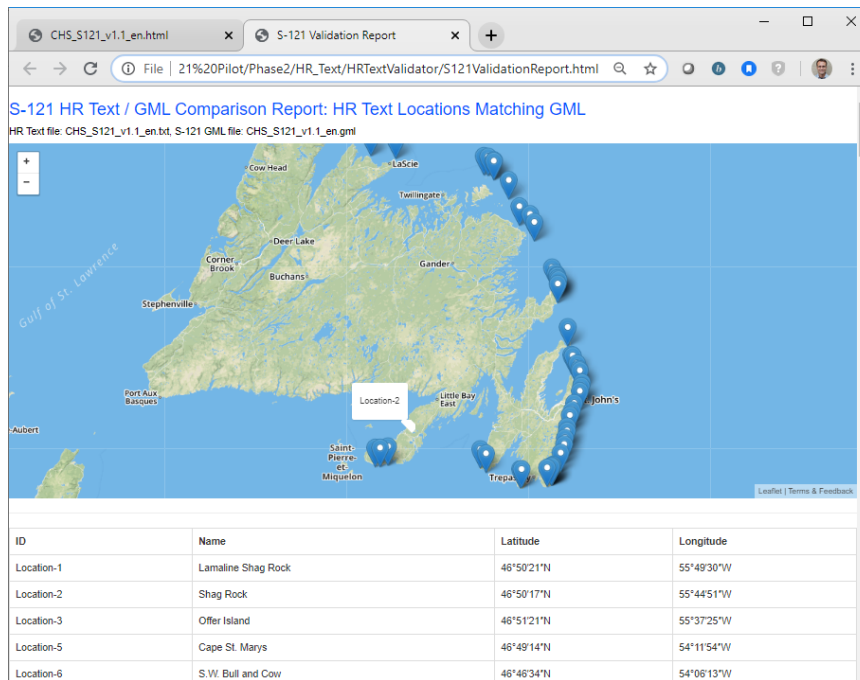
Figure 29. Validation report showing matching Locations [Note]

The above table shows the results of comparing HRText and GML Locations. This shows a high level of confidence that the results indeed match since both the ids, names and locations are found to agree.
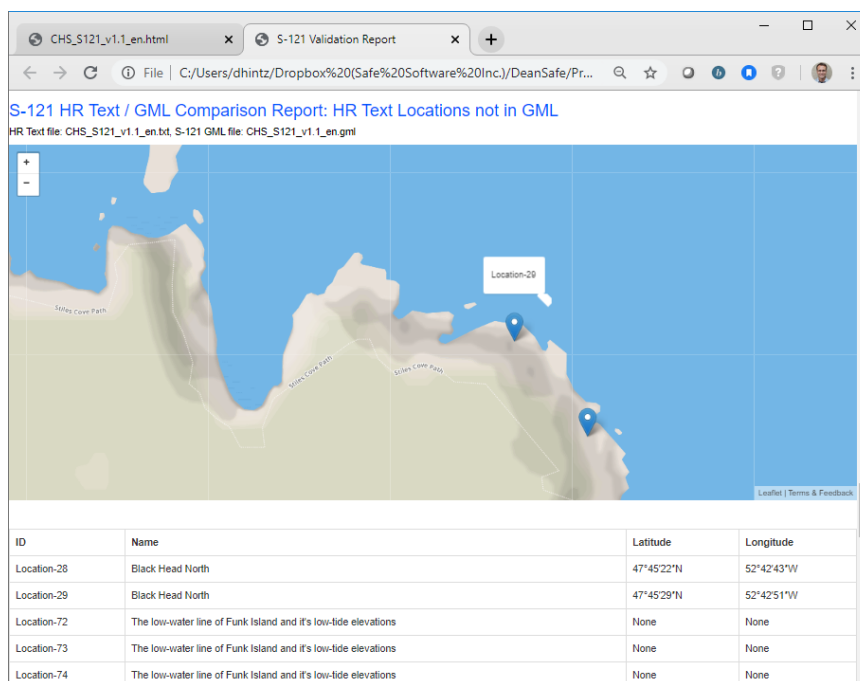
Figure 30. Validation report showing HR Locations not in GML [Note]

The above result shows some Location elements that occur in the HR Text but apparently are not found in the GML dataset. In this case, the locations appear to coincide, but the names are different. The HRText for Location-28 is read as 'Black Head North' whereas the Location-28 in the GML dataset has the name 'Black Head North 1'.
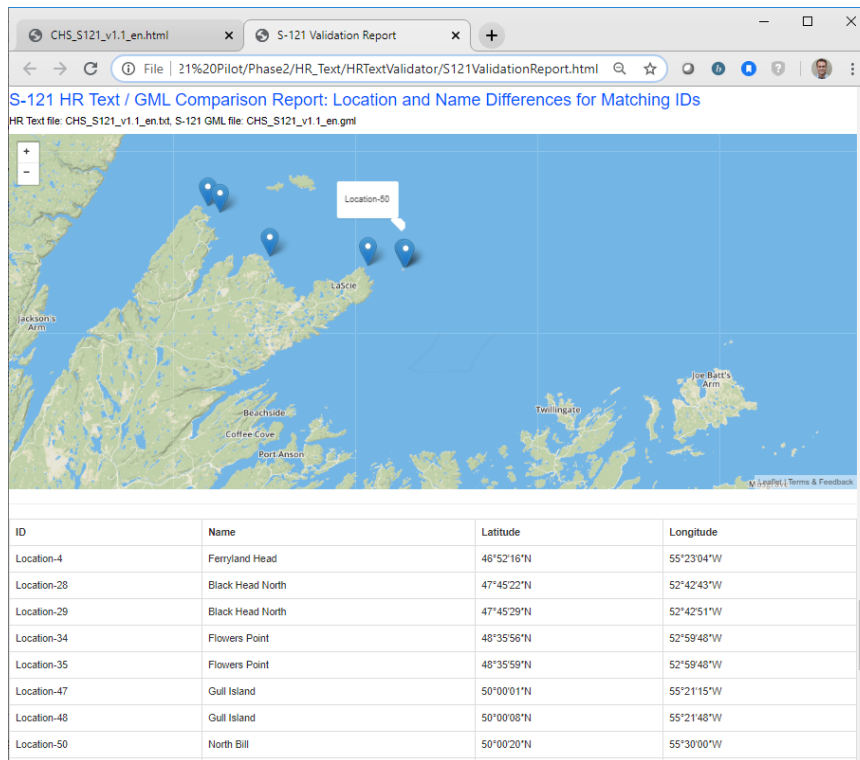
Figure 31. Validation report showing results for matching HRT/GML Locations with attribute differences, or geometry positions exceeding tolerance distance [Note]

Some results produced a match by ID but mismatch by location that exceeded the tolerance. In this case the tolerance was set to 0.01 degrees. An error this large is significant and should be greater than what would be expected from a DMS to decimal degree conversion. In the case of Location-4 above:

HR Text value is: Location-50 North Bill 50°00'20"N 55°30'00"W The GML value is: Location-50, Unnamed Point, 50.005555556, -55.528333333

So, it appears that the ID produced a match but both the name and the location were different enough to produce a significant location and name value difference.

In the case below, the Data Inspector was used to measure the distances between HR Text and GML Location coordinates.
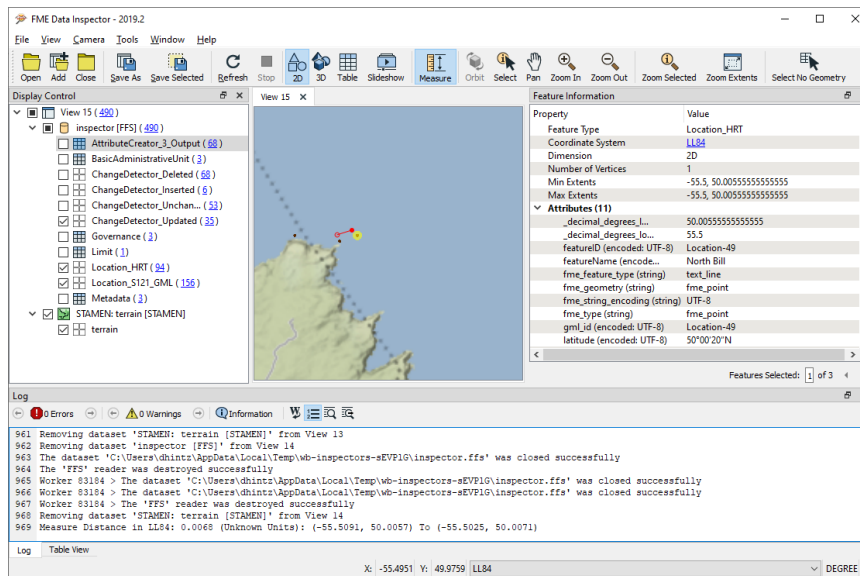
Figure 32. Comparison results directed to FME Data Inspector for detailed review. [Note]

To enhance the validator, it would likely be beneficial to perform multiple comparisons in sequence rather than everything at once. One test might only compare property values and another could compare only geometries. This would allow the production of more detailed reports to better characterize the type of differences between Location features from different datasets.

6.4.7. HRTextValidator: Sources of Error, Outstanding Issues and Conclusions

There were a number of project delays that were difficult to avoid given the need for S-121 and Georeg schema development throughout the course of the Pilot. In addition, there were implementation complexities around relationship representation and multi-lingual / English / French language support that resulted in changes to the source CHS geodatabase into the last week of the Pilot. The D2 FME GIS Application / translator was central to providing S121 GML for the other components such as Human Readable Text generation and the various OGC web services.

Because of this, priority was given to the GIS application development rather than the other optional components such as HR Text Validation. As such, there was not sufficient time in the project to develop a comprehensive HR Text Validator. Instead, a minimal validator was built to test comparisons based on Location features alone. Still, this produced results sufficient to demonstrate both that the HR Text can be read back into a GIS and that it is possible to automate at least a partial HR Text validation against the S121 GML dataset from which it was derived.

Note that the business rules and comparisons associated with validation necessarily need to be refined further to produce more accurate results sufficient to verify that a given set of HR Text is indeed equivalent to its corresponding S121 GML dataset. As mentioned above, a series of targeted ChangeDetectors would allow for more specific reporting so that issues with attribute comparisons could be distinguished from geometry comparisons.

Closer inspection and tuning of a variety of sample test runs would help ensure that there are fewer false positives when it comes to reporting differences. The addition of white space filtering and DMS position rounding might help with this as well. Finally, addition of handling for Limits, Zones, and BAUnits and their relationships would produce a more comprehensive validation report. Still, as noted in the introduction, there are limits to how many relationships can be derived

from the HR Text alone.

It will necessarily be left to follow-on work to further test and explore the possibilities related to HRText Validation, perhaps in conjunction with the development of Pilot demos for a Marine DWG presentation at an upcoming OGC TC. It would be interesting to conduct additional tests to see what other feature types can be read, and what additional comparisons and relationships with the equivalent S121 GML can be assessed.

6.5. Metadata for Catalogue Services

A Phase 2 component of the Pilot was concerned with the onward use of GML S-121 and extension) data within Catalogue Services for the Web (CS-W) services. These services are used extensively in many OGC conformant SDIs and there was therefore a need, as part of the data creation process, to create metadata to support such services. The stated standard within the requirement was the the Canadian Government community profile of ISO 19115:2003 (HNAP) profile.

A great deal of expertise within both participants and sponsors generated metadata for a selection of the datasets involved and these were used in preparation of the CS-W services. An additional enabler was a structured metadata editor used by CHS internally which was demonstrated and used by participants. A discussion of the issues in respect of HNAP vs ISO 19115 is contained in the section on Metadata

Chapter 7. Onward Uses for Generated Data

This section focuses on how generated GML data is able to be used to feed and drive web services of various types within the OGC ecosystem.

7.1. Web Services

7.1.1. Requirement Description

Servers that provide S-121 Data conforming to OGC Web Feature Service (WFS) and Web Map Service (WMS). The data should follow the agreed S-121 GML Application Schema.

7.1.2. Implementation

Two providers of OGC WFS/WMS services are documented within this ER and present on the GitHub repository. These are listed below. The promulgation of GML data through standards compliant web services should be straightforward as the S-121 GML defined by the application schema is standard, uses the S-100 GML profile and contains only simple geometry primitives and standardized datums (aside, of course, from the textual positions within the locationReference attributes).

7.1.3. Use of WFS-SF and List Attribute Flattening

The GML profile and schema are both defined using complex GML and processing and distribution. Many discussions within the pilot highlighted the benefits for some stakeholders of "flattened" schemas which would allow many clients and servers to use Simple Features capabilities to distribute this data. This is referenced in the ER outputs at the end of this report for possible future development.

The WFS that is supported and can be served out by ArcGIS Server and other WFS server and client implementations uses GML Simple Features. WFS-SF services support simple features and do not support extended geodatabase functionality such as relationship classes, joins, networks, and rules, therefore the relational attribute tables within the schema and implemented in some of the source datasets are not supported in these implementations. The workflow for publishing GML data to a WFS using ArcGIS (ArcMap or ArcGIS Pro) must include importing or writing the data to a geodatabase first.

The user can use the Quick Import tool from the Data Interoperability toolbox or launch an FME workbench from the ribbon. List attributes from the GML file are not compatible with geodatabases, so are transformed in order to write the values to feature class attribute fields. The easiest way to expose the list attribute values in an attribute field is to concatenate them in a single field. Alternatively, each value in the list can be written to a separate indexed field in the feature class attribute table, although this introduces additional complication in the table schema.

If list attribute values need to be exposed in a feature class attribute table and the concatenation method is chosen, then the user must follow the following steps.

- Launch FME workbench from the ArcGIS Pro ribbon.

- In the workbench add a reader for the GML file using the provided application schema, then also add a writer for a previously created empty geodatabase and the schema from the GML.

- The feature types from the GML and feature classes in the geodatabase should align.

- Review the list attributes in the GML and add new corresponding fields to every feature class which will require them.

- Configure a ListConcatenator transformer for each list attribute and the new feature class fields.

- Once all list attributes are configured with a transformer the process can be run and the resulting feature classes can be reviewed in ArcGIS Pro or ArcMap.

Other transformers can be used to expose the list values differently. Once the data is published to a WFS, it will use the feature class data structure and whichever list attribute transformation was used.

7.1.3.1. Geomatys End-points

UKHO Jersey Data

- WFS version 2.0.0 and 1.1.0 (4 layers) : https://mlb.examind.com/examind/WS/wfs/UKHO

- WMS version 1.0.1 and 1.3.0 (4 layers): https://mlb.examind.com/examind/WS/wms/UKHO

UKHO Data Granville

- WFS version 2.0.0 and 1.1.0 (4 layers) : https://mlb.examind.com/examind/WS/wfs/UKHO-Granville

- WMS version 1.0.1 and 1.3.0 (4 layers) : https://mlb.examind.com/examind/WS/wms/UKHOGranville Source data : UK Granville Dataset

Canada Extended Continental Shelf (CHS_S121_v1.04.xml)

- WMS version 1.0.1 and 1.3.0 Canadian Extended Continental Shelf (14 Layers) : https://mlb.examind.com/examind/WS/wms/Canadian_Extended_Continental_Shelf

- WFS version 2.0.0 and 1.1.0 Canadian Extended Continental Shelf (13 Layers) : https://mlb.examind.com/examind/WS/wfs/Canadian_Extended_Continental_Shelf

7.1.3.2. Esri Canada End-Points

Canadian Data (CHS_S121_v1.0.xml) 17 layers: Baseline Point, Boundary Point, Contributing Point, Location, Boundary, Construction Line, Limit, Normal Baseline, Out Limit of the Continental Shelf, Outer Limit of the Exclusive Economic Zone, Outer Limit of the Territorial Sea, Straight Baseline, Contiguous Zone, Continental Shelf, Exclusive Economic Zone, Territorial Sea, Zone.

9 Standalone tables: Basic Administrative Unit, Governance, Party, Responsibility, Right, Source, Dataset, Additional Spatial Information, Party Member.

- WFS Canadian Extended Continental Shelf: http://wilsondemo.eastus.cloudapp.azure.com/arcgis/services/CanadaExtendedContinentalShelf_June2020/MapServer/WFSServer?request=GetCapabilities&service=WFS

- WMS Canadian Extended Continental Shelf: http://wilsondemo.eastus.cloudapp.azure.com/arcgis/services/CanadaExtendedContinentalShelf_June2020/MapServer/WMSServer?request=GetCapabilities&service=WMS

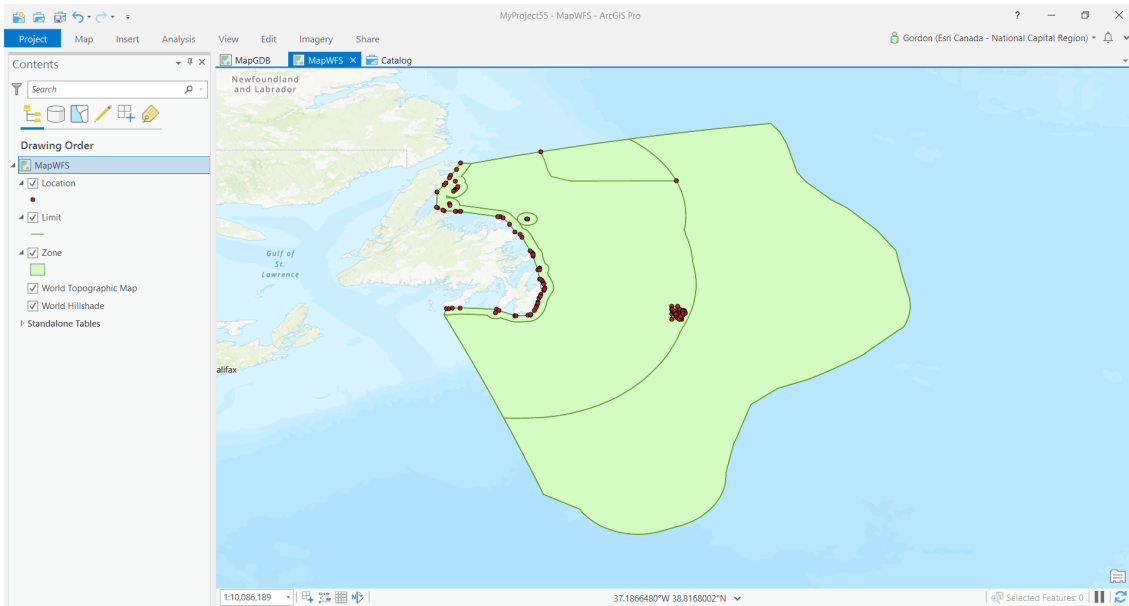The following image shows a WFS rendering of S-121 GML data within a GIS

Figure 33. WFS rendering of S-121 GML data within a GIS [Note]

showing the GML received from the WFS endpoint. As a key use case the GML itself from the request is reproduced in part below. This shows two of the features and the header of the request.

```
<wfs:FeatureCollection xmlns:wfs="http://www.opengis.net/wfs"
                       xmlns:gml="http://www.opengis.net/gml"
                       xmlns:gco="http://www.isotc211.org/2005/gco"
                       xmlns:gmd="http://www.isotc211.org/2005/gmd"
                       xmlns:gmx="http://www.isotc211.org/2005/gmx"
                       xmlns:srv1="http://www.isotc211.org/2005/srv"
                       xmlns:xlink="http://www.w3.org/1999/xlink"
                       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                       xmlns:ns0="http://www.iho.int/S-121"
                       xmlns:ns1="http://www.iho.int/s100gml/1.0"
                       gml:id="collection-1" timeStamp="2020-03-25T10:48:59"
                       xsi:schemaLocation="http://www.iho.int/S-121

https://mlb.examind.com/examind/WS/wfs/JuanDeFuca?request=DescribeFeatureType&version=1.1.0&service=WFS&namespace=xmlns(n
s35=http://www.iho.int/S-121)&typename=ns35:BaselinePoint http://www.opengis.net/wfs
http://schemas.opengis.net/wfs/1.1.0/wfs.xsd http://www.opengis.net/gml
http://schemas.opengis.net/gml/3.1.1/base/gml.xsd" numberOfFeatures="66">

<gml:featureMember>
     <wfs:FeatureCollection gml:id="http://www.iho.int/S-121:BaselinePoint" timeStamp="2020-03-25T10:48:59"
xsi:schemaLocation="http://www.iho.int/S-121
https://mlb.examind.com/examind/WS/wfs/JuanDeFuca?request=DescribeFeatureType&version=1.1.0&service=WFS&namespace=xmlns(n
s35=http://www.iho.int/S-121)&typename=ns35:BaselinePoint http://www.opengis.net/wfs
http://schemas.opengis.net/wfs/1.1.0/wfs.xsd http://www.opengis.net/gml
http://schemas.opengis.net/gml/3.1.1/base/gml.xsd" numberOfFeatures="19">
     <gml:boundedBy>
        <gml:Envelope srsName="urn:ogc:def:crs:OGC:1.3:CRS84">
            <gml:lowerCorner>-124.73672499999998 48.15043055600006</gml:lowerCorner>
            <gml:upperCorner>-122.88477777799994 48.78621944400004</gml:upperCorner>
        </gml:Envelope>
     </gml:boundedBy>
     <gml:featureMember>
        <ns0:BaselinePoint gml:id="Features_Point_1138">
            <ns0:label>Andrew Bay Reference Mark</ns0:label>
            <ns0:pointType>defined</ns0:pointType>
            <ns0:geometry>
                <gml:Point srsName="urn:ogc:def:crs:OGC:1.3:CRS84">
                    <gml:pos>-123.16874722199998 48.551513889000034</gml:pos>
                </gml:Point>
            </ns0:geometry>
        </ns0:BaselinePoint>
     </gml:featureMember>
     <gml:featureMember>
        <ns0:BaselinePoint gml:id="Features_Point_1139">
            <ns0:label>Beechy Head Reference Mark</ns0:label>
            <ns0:pointType>defined</ns0:pointType>
            <ns0:geometry>
                <gml:Point srsName="urn:ogc:def:crs:OGC:1.3:CRS84">
                    <gml:pos>-123.65298888899997 48.31420000000003</gml:pos>
                </gml:Point>
            </ns0:geometry>
        </ns0:BaselinePoint>
     </gml:featureMember>

<--! [only 2 features shown for brevity] -->
```

A more complex example from the Canadian Continental Shelf is shown in the following GML content.

```
<gml:featureMember>
    <ns0:BaselinePoint gml:id="p-Location-62">
        <ns0:label/>
        <ns0:collectionName>CA</ns0:collectionName>
        <ns0:featureName>
        <ns0:displayName>false</ns0:displayName>
            <ns0:Language>en</ns0:Language>
            <ns0:Name>St. Mary□s Cays 1</ns0:Name>
        </ns0:featureName>
        <ns0:context>Area 2</ns0:context>
        <ns0:startLifespan>2019-11-26Z</ns0:startLifespan>
        <ns0:endLifespan>2021-12-31Z</ns0:endLifespan>
        <ns0:releasabilityType>Controlled</ns0:releasabilityType>
        <ns0:source xlink:href="#Source-1"/>
        <ns0:administration xlink:href="#BAUnit-7"/>
        <ns0:refersTo xlink:href="#Reference-60"/>
        <ns0:refersTo xlink:href="#Reference-133"/>
        <ns0:locationReference>
            <ns0:categoryOfLocation>OriginalLocation</ns0:categoryOfLocation>
            <ns0:locationByText>St. Mary□s Cays 1</ns0:locationByText>
            <ns0:textLat>46° 42' 58.000" N</ns0:textLat>
            <ns0:textLon>54° 13' 17.000" W</ns0:textLon>
            <ns0:referenceSystem>4622</ns0:referenceSystem>
        </ns0:locationReference>
        <ns0:interpolationType>densification</ns0:interpolationType>
        <ns0:pointType>defined</ns0:pointType>
        <ns0:geometry>
            <gml:Point srsName="urn:ogc:def:crs:OGC:1.3:CRS84">
                <gml:pos>-54.221388889 46.716111111</gml:pos>
            </gml:Point>
        </ns0:geometry>
    </ns0:BaselinePoint>
</gml:featureMember>
```

The following image shows the WMS outputs from ESRI and Geomatys endpoints overlaid on each other within an ESRI client.
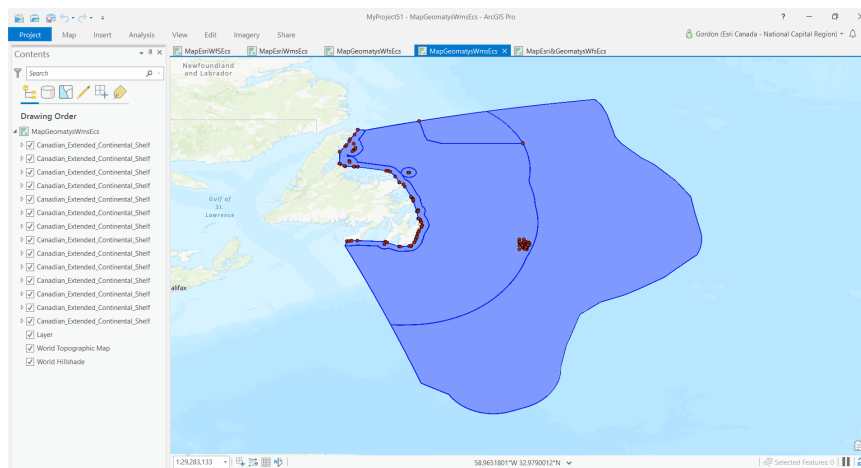
Figure 34. WMS outputs from ESRI and Geomatys endpoints [Note]

7.2. Client SDI

This section details how the various participants have implemented their Client SDIs. These are also detailed in the GitHub repository.

7.2.1. Teledyne CARIS

One of the main focus areas of the project has been the development of the S-121 GML exchange format and the associated GML/XML application schema (XSD). CARIS has been involved with proposals, testing, feedback, discussions on weekly conference calls and attendance of a productive face-to-face meeting in Ottawa.

Teledyne CARIS offers a limits and boundaries option to desktop and enterprise applications which is being extended to support S-121. Support for S-121 including the latest GML format and Feature Catalogue will be rolled out via the regular release schedules of related applications such as BDB, HPD, and EasyView.

7.2.1.1. Opening and Importing GML

CARIS has two paths for opening the S-121 GML content. The standard GML parser and the S-100 implementation. The standard GML parser is based on the GDAL library and allows the data to be opened as one layer per feature class and information type. At this point the retention and navigation of complex hierarchies and relationships is limited. The GMLAS extension is not available currently but is being investigated.
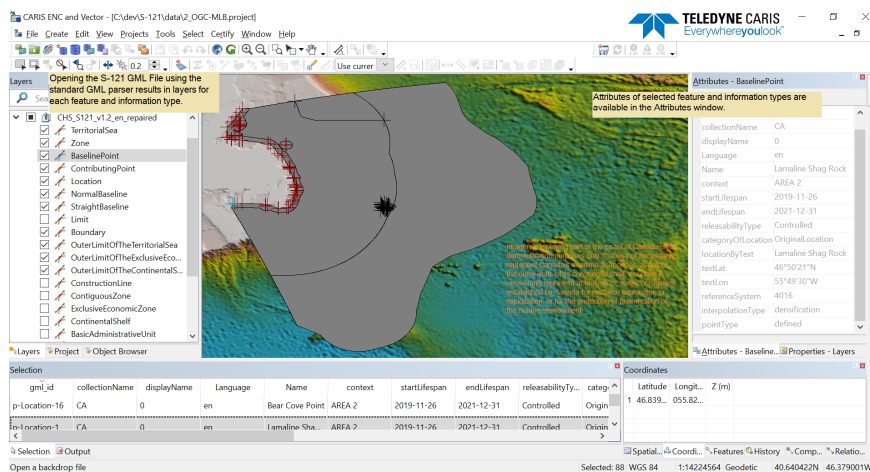
Figure 35. Teledyne-Caris GML Default Open Interface [Note]

The S-100 parser reads the S-121 data and validates the input against expected S-100 structures and fields as well as validating the content against an associated S-100 XML Feature Catalogue (FC). During the project Caris has developed an S-121 draft xml Feature Catalogue for use in testing.
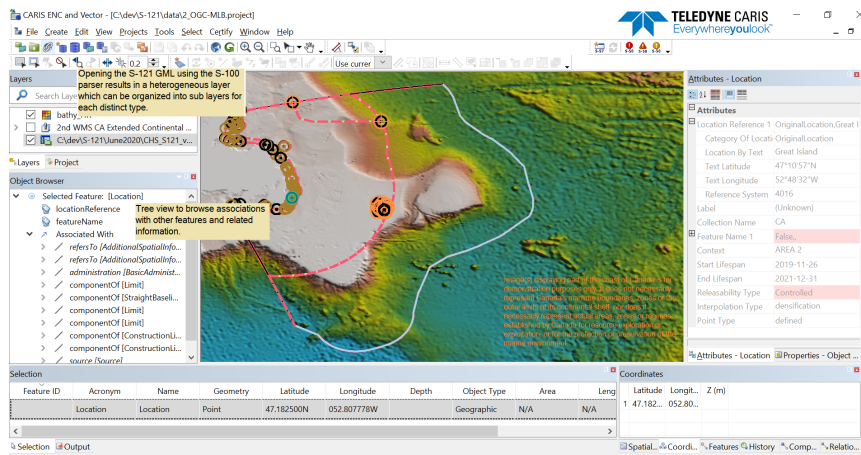
Figure 36. Teledyne-Caris Opening S-100 GML [Note]

Over the course of the project, the definition of the content of the S-121 GML format has been a moving target as this was a significant portion of the study to find ways to make content that is interoperable and to prove that with testing. The content definitions have been primarily implemented through the use of a GML/XML validation schema file (*.xsd). Since none of the other vendors were using the specified S-100 FC in their implementations, the FC was mostly ignored by the project team. Over the course of the project CARIS has been maintaining a draft S-100 XML FC mostly consistent with the gml xsd definitions.

The latest draft S-121 XML FC prepared by CARIS has been posted in the deliverables folder as well as an html version for easier browsing. (repository deliverables/d3-GML-Application-Schema/schema/current/v4.0/)

Many of the changes to the GML schema also made software changes necessary in order to implement the parsing/import and export of data in compliance with the GML schema adjustments. The substantial changes to the schema over the project amounted to significant resources and rework on the import/export code which effectively limited the amount of resources available to spend on other aspects.

7.2.1.2. Opening/Browsing WMS and WFS

Testing was carried out to confirm the ability to open and browse the available WMS and WFS services. Initially there were some issues with associated CRS, layer discovery and namespaces but these were resolved over the course of the project.

The default GDAL discovery functionality had to be extended to recognize the S-121 feature layers as the default was looking for specific substitution groups "_Feature" or "AbstractFeature" but the S-121 GML xsd was constructed using inheritance instead of substitution groups.

Images 1-5 show connection and browsing of WMS layers integrated with WFS and other data sources using the CARIS Common Application functionality.
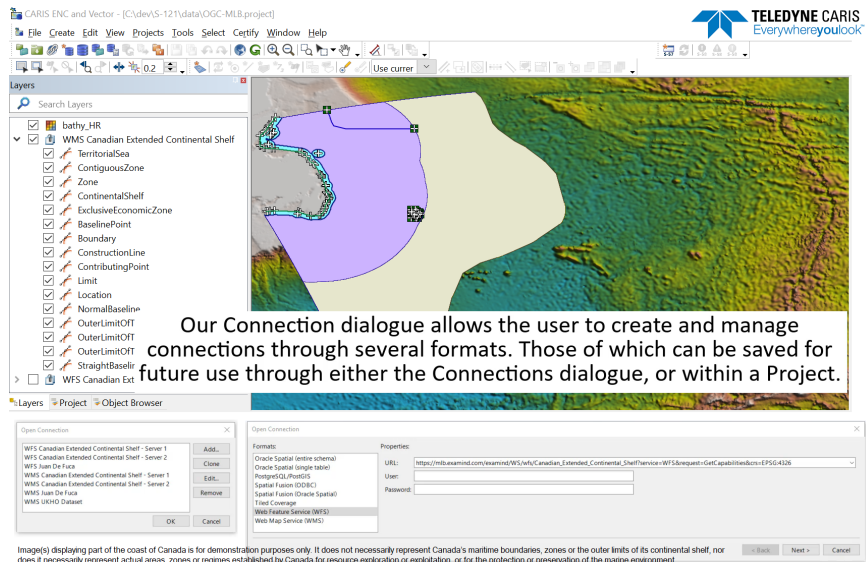
Figure 37. Teledyne-Caris Connection Tool [Note]
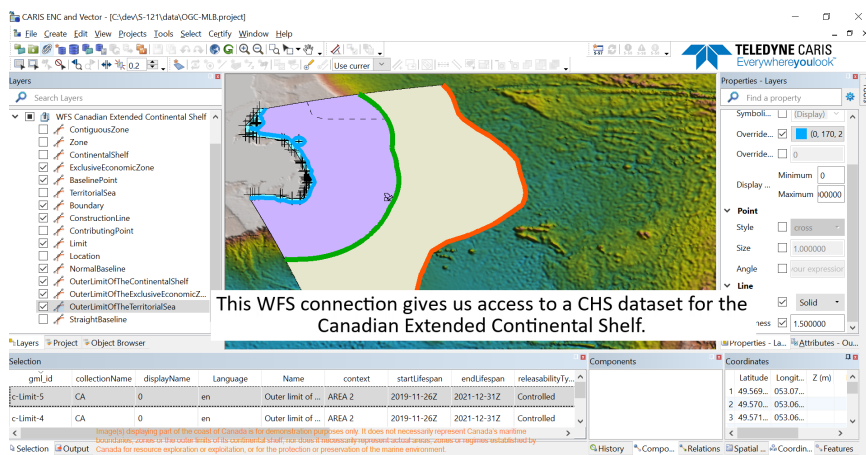
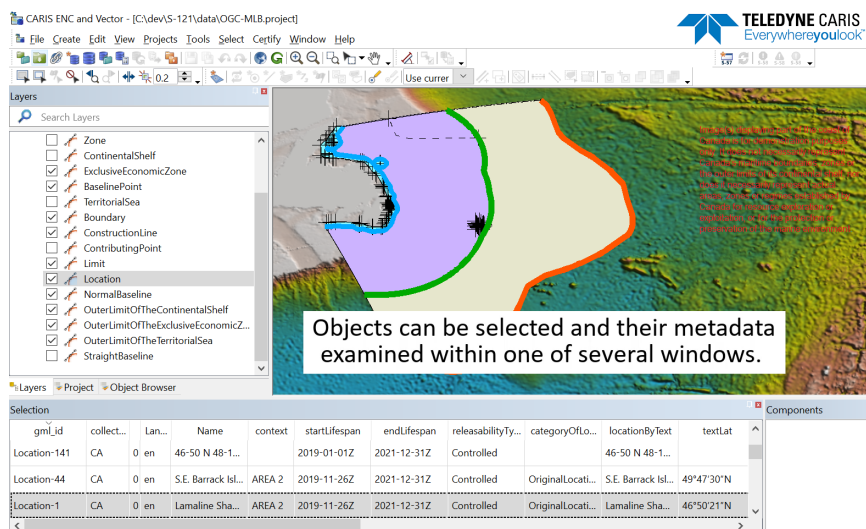*Figure 38. Teledyne-Caris WFS Connection to CHS Extended Continental Shelf* [Note]

Figure 39. Teledyne-Caris Object and Metadata Select and Query [Note]
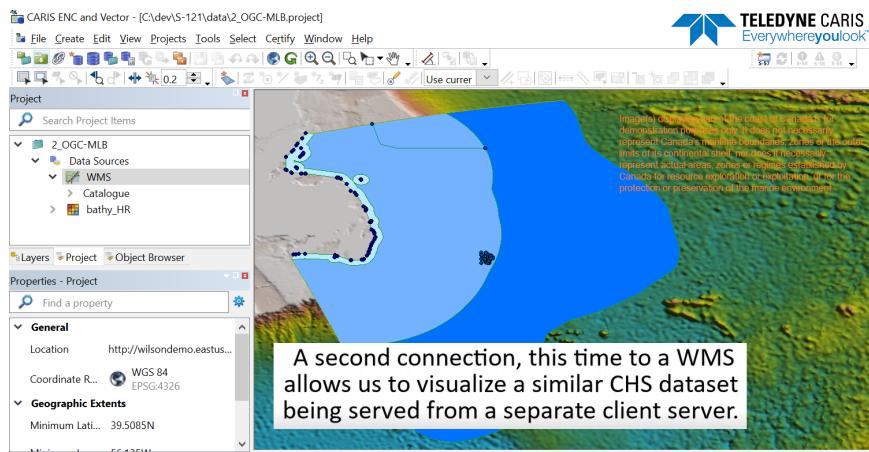
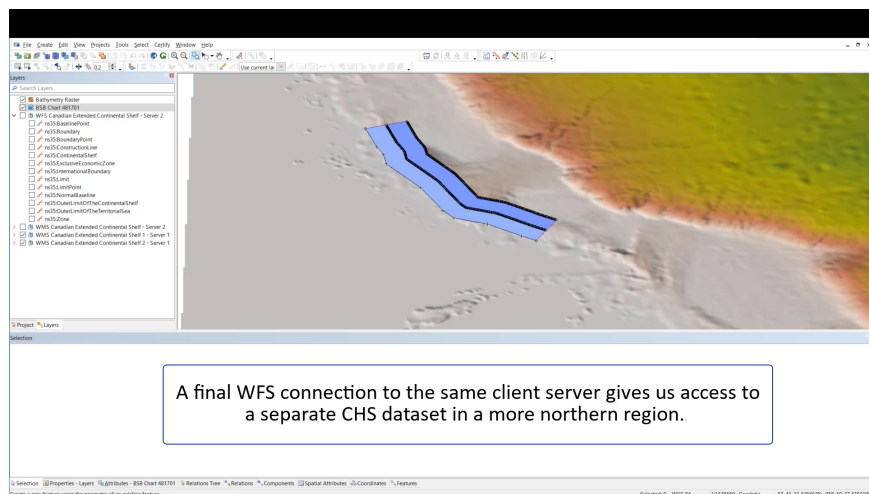*Figure 40. Teledyne-Caris WMS Visualisation* [Note]

Figure 41. Teledyne-Caris Additional WFS Connection [Note]

7.2.2. Safe Software FME

Over the course of the MLB Pilot, Safe Software was actively involved in a number of critical areas of Pilot development including the provision of FME as a client for data consumption using OGC standards and services such as GML, WMS, and WFS. Clients should be able to perform the following:

- Consume S-121 data from a WFS following the S-121 GML Application Schema;

- Consume data created by GIS Applications that follow the S-121 GML Application Schema; and

- Demonstrate integrating data from different sources in Phase 2 available from the Catalog.

The clients helped demonstrate the end-to-end scenario and toolset interoperability with Canada's SDI, including the FGP, Canadian Surveyor General, Marine Spatial Data Infrastructure based on FGP requirements (Fisheries and Oceans Canada, Canadian Hydrographic Service), and with the Arctic SDI.

For the OGC MLB pilot, FME's generic GML reader was extended to support the evolving requirements of S-121 GML. FME's GML reader has an application schema mode that can dynamically read most GML including complex GML, given a well-defined and complete application schema. FME's WFS reader invokes the GML reader in order to process the results of a GetFeature request.

7.2.2.1. Challenges

S-121 GML and the related S-100 family of schemas posed some unique challenges:

- S-100 is not prescriptive on all aspects of GML specification and therefore there is inevitably some ambiguity in the data encoding;

- An atypical GML container (S121:Dataset, rather than typical gml:FeatureCollection);

- An atypical object model (AbstractGML for information types, typical GML uses AbstractFeature with no geometry);

- Hard dependencies on S-100, GMD and other external schemas;

- Complexities of the IMO/S-100 context: Extensions are not extensions of S-121;

- Relationships important to the regulatory significance - trace back to Source;

- additionalSpatialInformation is both a feature type and a property (*fixed in later schema iterations*);

- Both geometry and non-spatial features;

- Support for both in-line and referenced geometry;

- Inclusion of GML 3.3 namespace via S100 even though not used within S121 schema (Updated local copies of GML schemas to v3.2.2); and

- Axis order / CRS.

7.2.2.2. Modifications for Pilot

FME has strong support for OGC GML and WFS, including support for complex GML formats such as EU INSPIRE, CityGML, UK OS Mastermap, and AIXM. Still, S-121 GML and the related S-100 family of schemas posed some unique challenges that required some enhancements to FME's core GML support. Initially, many of these enhancements were possible with minor modifications to the GML reader/writer configurations files. All changes listed below were eventually incorporated into FME 2019.2 and will be supported going forward for FME 2020.0 and subsequent versions.

- Addition of support for AbstractGMLType / AbstractInformationType

- Recognition of atypical dataset containers

- Recognition of S-121 name spaces

- Map s121:informationType elements into FME features

- Distinguishing between reference types and feature types

- Support for referenced geometries. Note that this was tested on limited sample datasets near the end of Phase 1 and was not used extensively afterwards in the CHS test datasets.

- Bug fix: Don't exclude gml polygon as a substitution group member of AbstractSurface on writing

- Improved support for GML 3.3 (this namespace was dropped to simplify compatibility with other clients, but FME now has no problem supporting 3.3)

- WFS Reader – make more tolerant. Improved multi-namespace handling. Make sure namespace prefixes are taken from typeNames that have QNames in content – not just namespace header

- Allow xml elements to define their own namespaces rather than depending on parent namespaces

- FME 2020 add Application Schema based XML R/W – improved support for metadata

- Allow for selection of multiple application schemas for reading, writing and validation

7.2.2.3. GML Reader

Early in the project there were some difficulties reading S121 GML with FME. The first main challenge was the atypical dataset container which FME initially did not recognize. The second challenge was the use of AbstractInformationType / AbstractGML for non-geometry feature types. FME, like most GIS applications, expects feature types to be of type AbstractFeature. Even non-geometry features are typically of type AbstractFeature albeit with null or no geometry. Given the unique nature of some of the information types, and some name collisions, filtering was added to the GML configuration files to recognize and map S-121 and S-100 namespaces so objects from those namespaces could be handled correctly by default.

Most of the above changes were made late in Phase 1 and early in Phase 2. Some of these problems came up due to the moving target of continual development occurring with the S121 and extension schemas. Also, in Phase 1 Safe worked with relatively simple, small datasets – basically handmade sample datasets and a small International Boundary Commission dataset with limited feature types and attribution. Once we started early in Phase 2 to attempt translations of the much more comprehensive CHS Extended Continental Shelf database to S121 GML a number of other challenges were discovered. Work on the georeg extension schemas caused additional issues related to multiple namespaces and WFS support.

By the mid point in Phase 2, FME no longer had any major problems reading S121 GML. One advantage in using the FME client was that it was continually tested as part of the GML generation development process. Every time a change was made to GML generation, such as adding a new feature type, immediately there were results available from validating against the schema, and the GML output was constantly being reviewed by reading it back in with the FME Data Inspector client. This meant that any problems with the GML writing or subsequent GML reading quickly became apparent. Most often all it took was validation against the application schema to determine whether the problem lay with the writer or the reader / client.

By the end of the pilot there were no known outstanding problems with FME's Data Inspector client for reading S-121 GML. Note that there are some requirements for correctly configuring the GML reader to consume S-121 GML. The following is an example of the GML reader settings used.
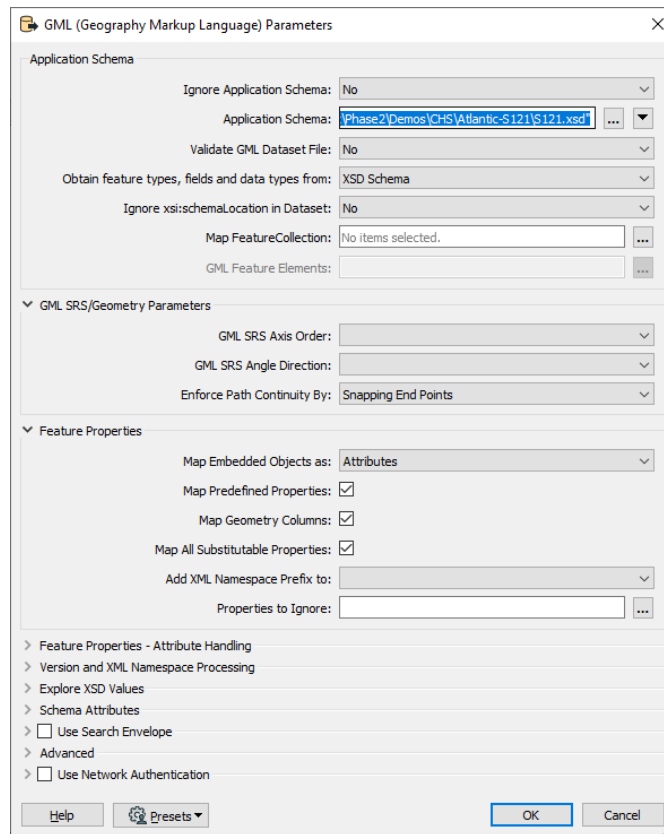
Figure 42. FME Client GML reader settings

Note that if the xsi:schemaLocation is set in the GML header as follows:

```
xsi:schemaLocation="http://www.iho.int/S-121 S121.xsd
```

Then the application schema path reader parameter does not need to be set, so long as the S121.xsd is available in the same directory as the GML dataset. The S121.xsd is now available at the public S121 website http://www.s-121.com/schemas/v4.0/S121.xsd.

Also note that if no schema is available, the client can use the 'Ignore Application Schema' set to Yes to read virtually any GML including S121 GML. However, because S121 does not use a standard featureMember container for its features, this Ignore Schema mode requires that the client specify the GML feature type elements, for example:

```
GML Feature Elements: Location Limit Zone BasicAdministrativeUnit Governance Source
Dataset Party
```

This approach allows FME to read virtually any complex GML even without an application schema, both directly from a dataset and via WFS. This can also be useful for troubleshooting WFS services since often the GetFeature response may have namespace or schema discrepancies when compared with the DescribeFeatureType response.
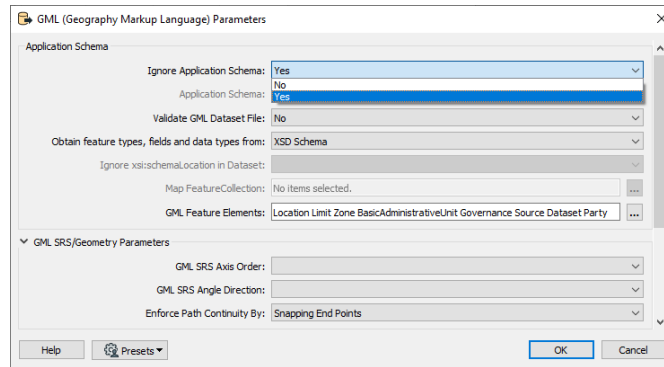
Figure 43. FME Client parameters for reading GML

7.2.2.4. S121 Extension Support and Other Issues

Reading the extension schemas was initially challenging since it introduced problems with name collision related to the dataset container. As well, given that multiple applications schemas were required, ways had to be found to scan multiple schemas to discover the required element and type definitions. One key approach was to use XSD import to use elements from an external namespace in order to keep them in different namespaces, and include to bring in foreign elements to use within the namespace of the app schema calling the include. Within FME, using the reader option 'add namespace prefix' helps keep track of elements with the same name from different namespaces.

Other issues that did cause some difficulties early on were CRS and axis order. Some clients can confuse the axis order when the CRS is defined as an EPSG. In the case of FME, the axis order can be specified – 1,2 in this case. Once the GML dataset was generated with the OGC CRS definition of "urn:ogc:def:crs:OGC:1.3:CRS84" this no longer appeared to be a problem, since this CRS explicitly defines the axis order. Also, there was the issue of GML 3.3. S121.xsd including the GML 3.3 namespace.

```
xmlns:gmlce=http://www.opengis.net/gml/3.3/ce
```

This initially caused some problems with the FME client due to element definition collisions. Updating local copies of GML schemas to v3.2.2 seemed to alleviate this problem.

7.2.2.5. Dataset Examples in FME Client
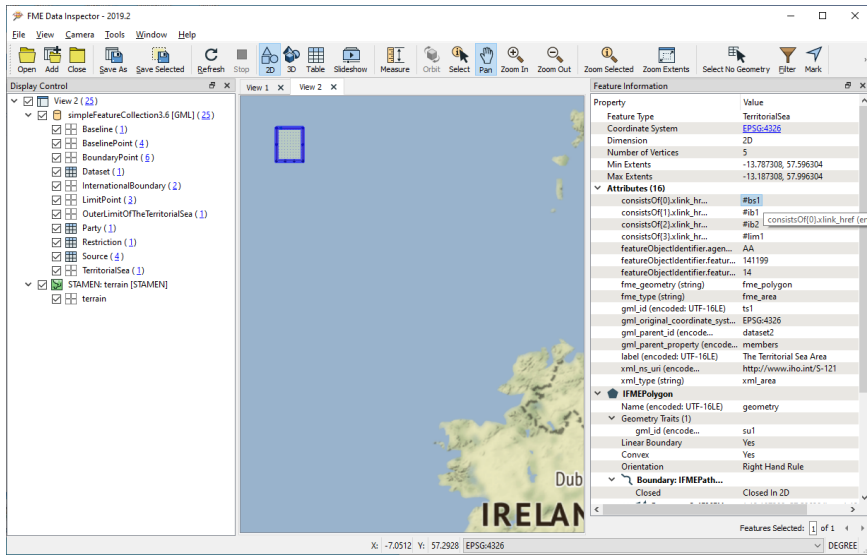
Figure 44. FME Client initial test dataset
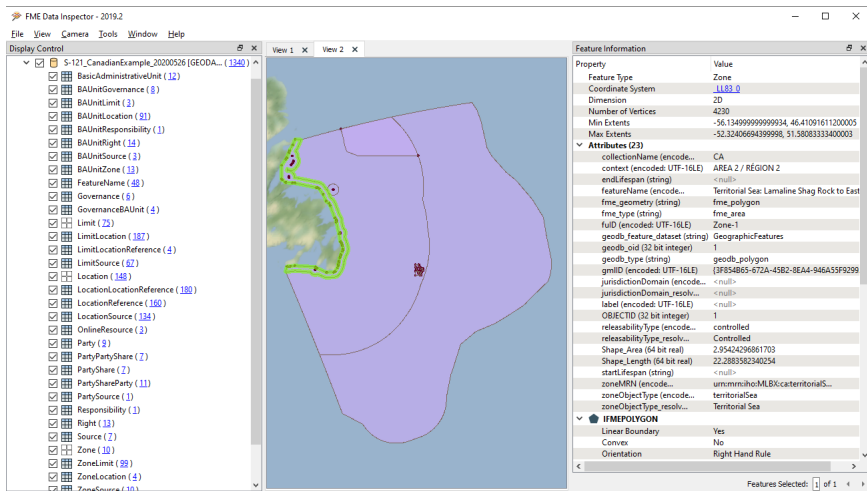
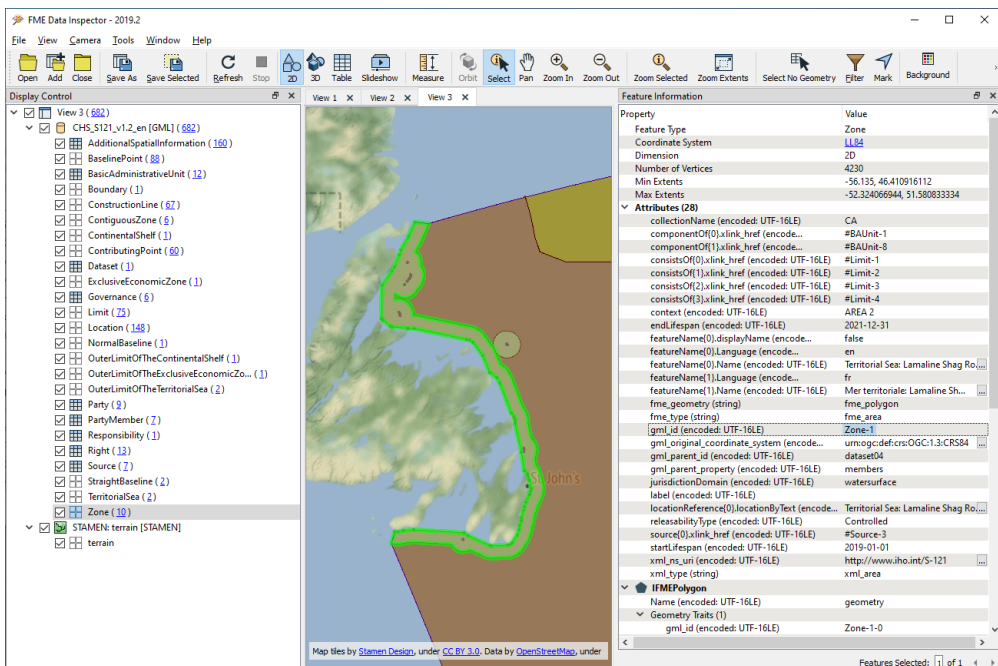*Figure 45. FME Client CHS Extended Continental Shelf source geodatabase* [Note]

Figure 46. Extended Continental Shelf S121 GML: CHS_S121_v1.1_en.gml showing Zone-1 componentOf, consistsOf, source xlink linkages and multi-lingual featureName. [Note]
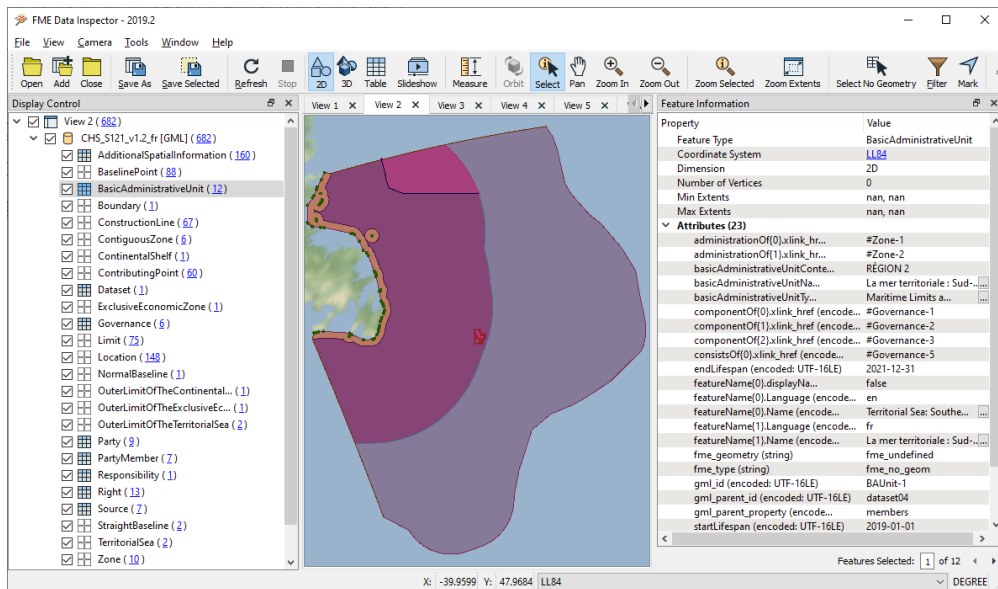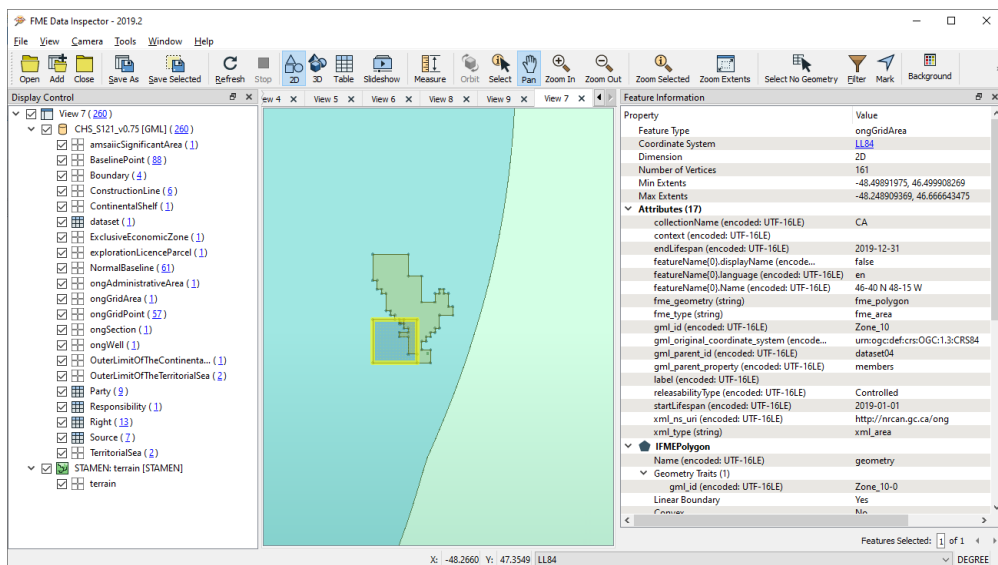
Figure 47. Extended Continental Shelf S121 GML: CHS_S121_v1.1_fr.gml showing BAUnit-1 componentOf, consistsOf, administrationOf linkages and multi-lingual featureName. [Note]

Figure 48. Combination of Georeg extension and S121 elements in CHS_S121_v0.73.xml – with ongGridArea selected [Note]

this is an example of reading the extension schema data within a COTS client, described in Extensions of the GML Application Schema.

7.2.2.6. WFS

FME's WFS reader supports OGC WFS 1.0, 1.1, and 2.0. It invokes the GML reader in order to process the results of a GetFeature request. Because of this, FME's WFS client can, in principle, handle any complex GML that FME's GML reader can read. The client also allows the inspection of complex attributes and geometry. This includes complex attributes such as featureName and LocationReference, and one to many series attributes that occur in S121 GML such as consistsOf and componentOf xlink_hrefs with multiple links to related features.

All this depends on a coherence between the namespace header within the GetFeature GML response and the DescribeFeatureType application schema response. When there are inconsistencies, the WFS client allows for some configuration options to work around them. This

includes adjusting settings such as CRS, axis order, selection of HTTP POST or GET, use of a local application schema when there are problems with DescribeFeatureType, and use of Ignore Application Schema when a valid application schema is not available. When all else fails, copying and pasting the WFS requests in a web browser often can help isolate problems.

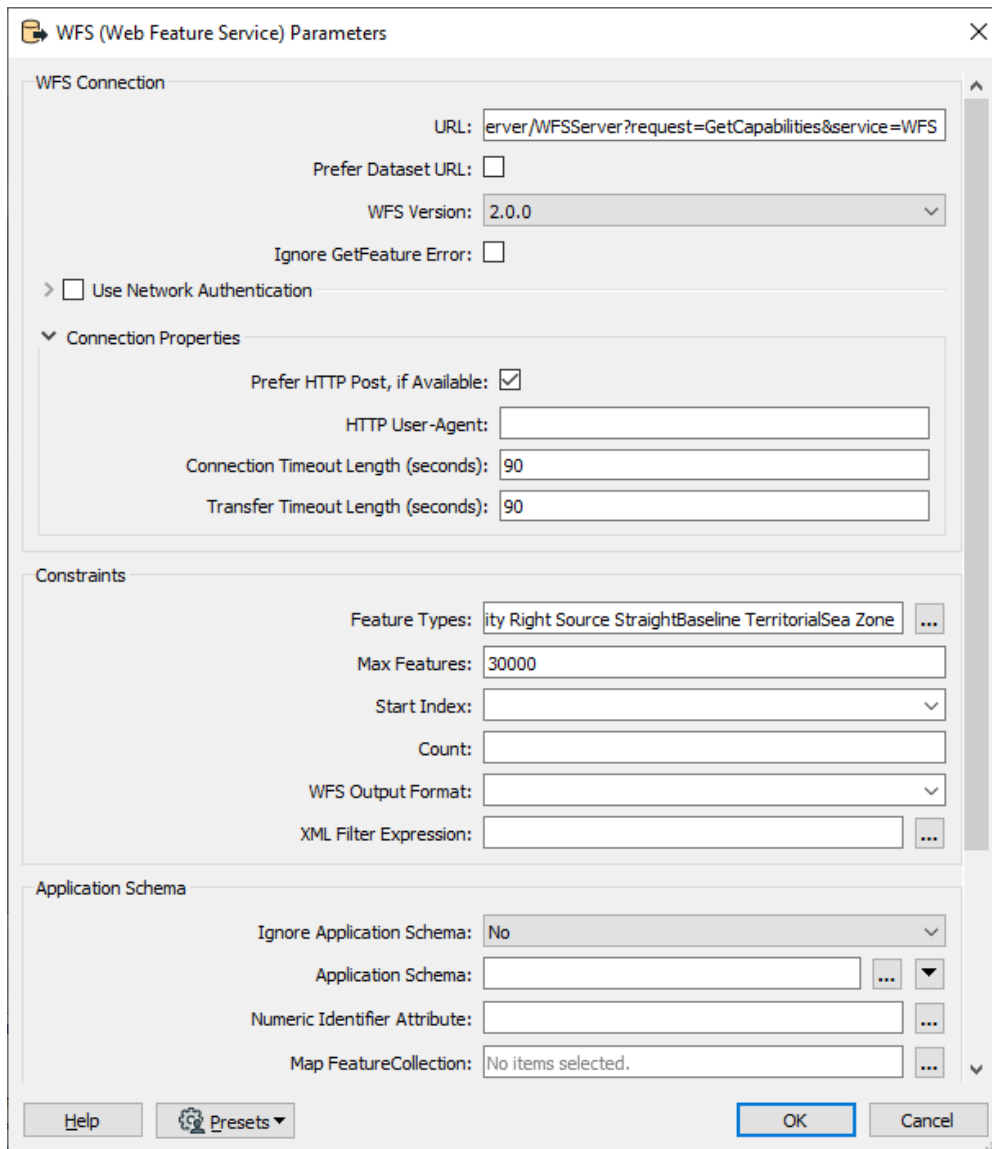For more information related to troubleshooting the use of the FME WFS client, see:

https://knowledge.safe.com/articles/19452/troubleshooting-wfs-errors.html

Figure 49. FME Client WFS Reader Settings showing Feature Type selection and Prefer HTTP POST option

The following image shows an example of the FME Client.
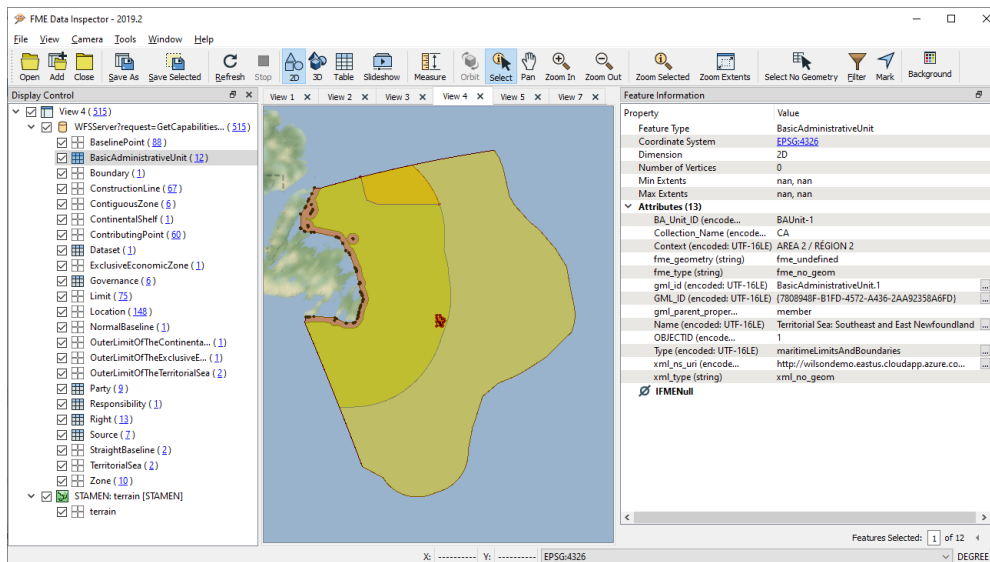
Figure 50. ESRI Extended Continental Shelf WFS Service showing BAUnit-1 non-geometry layer layer in the feature information pane. [Note]

In general, FME's WFS client was able to read from both the Geomatys and ESRI WFS servers using WFS 2.0. However, at the time of testing the Geomatys WFS server for UKHO Granville did not return any results that the client could consume. Note that the GetCapabilities request also did not return a valid XML response when invoked via a web browser. This suggests that this service may no longer be accessible at the time tested. On the other hand, some initial problems with UKHO Jersey were subsequently resolved on the WFS server side after contacting Geomatys.

In short, most WFS services associated with Phase 2 data were able to be read, including both geometries and full attribution. However, the non-geometry feature types for Geomatys CHS WFS were not accessible (Governance, Party, Source etc.).

Note that the default settings initially did not allow the FME WFS client to read the Geomatys WFS for CHS data. Changing the messaging protocol from HTTP GET to POST solved the problem. It appears that the HTTP POST protocol is more flexible for complex WFS services that require support of multiple namespaces.

7.2.2.7. WMS

FME's WMS reader supports OGC WMS 1.0, 1.1, and 1.3. The WMS client has some basic configuration options such as layer selection, split by layer, transparency, search envelope, resolution, and CRS settings. Otherwise it is a fairly simple client in that it has limited support for query, styles or symbology.
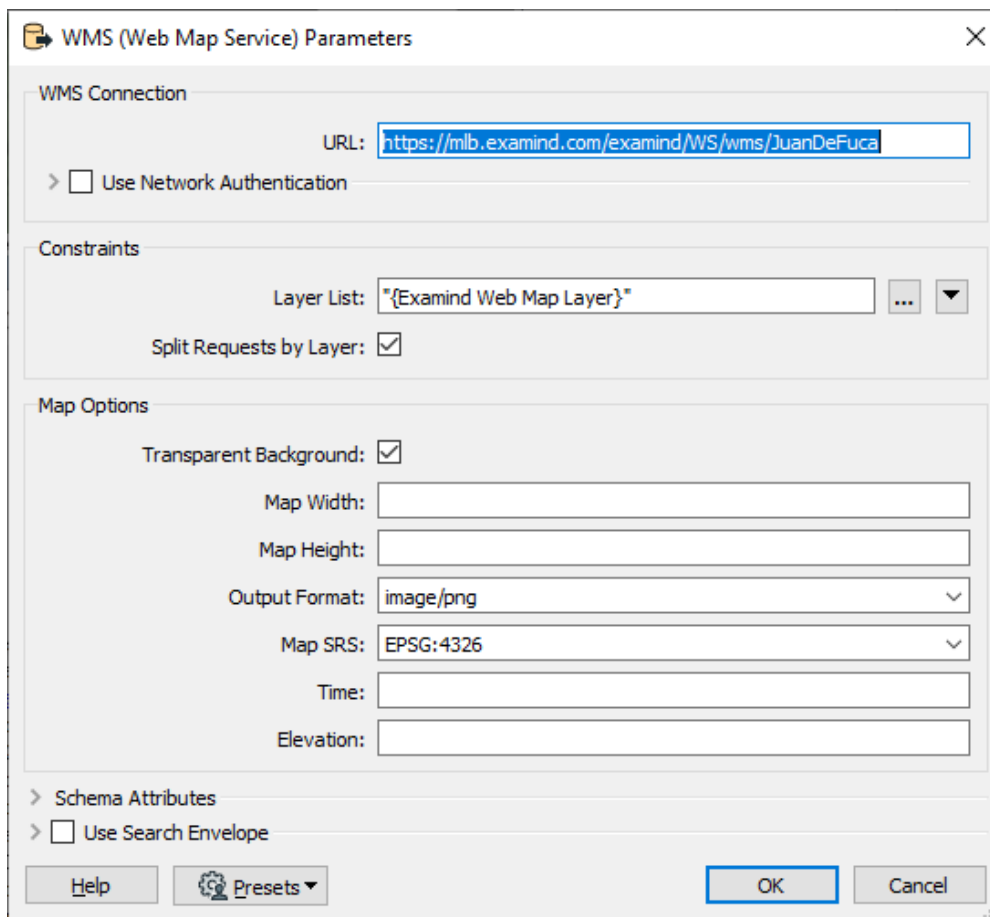
Figure 51. FME Client WFS Reader Settings showing Split Requests by Layer and Transparent Background Options
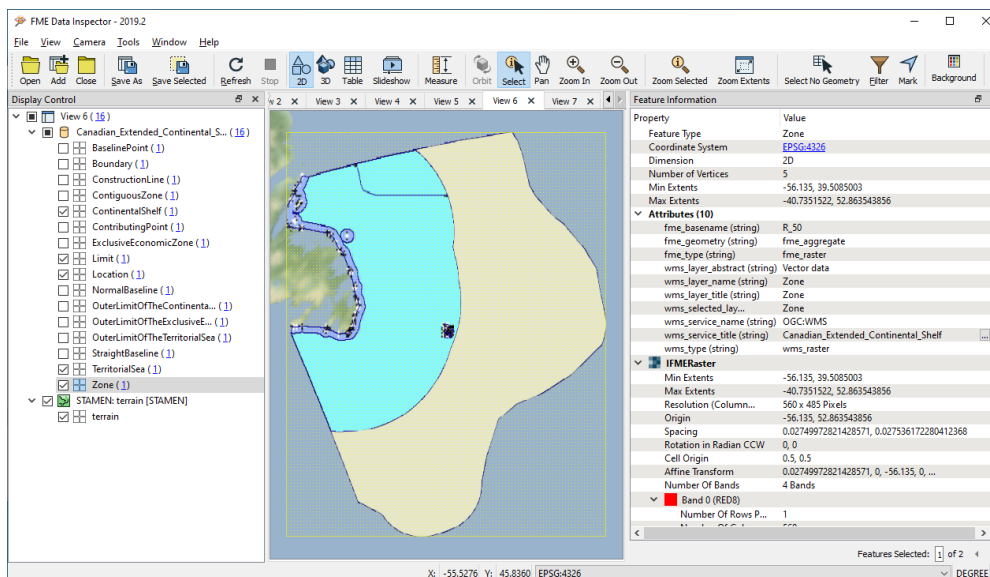

Figure 52. Geomatys Extended Continental Shelf WMS – zoom to extent with Limit, Location and selected zones enabled, displaying dataset parameters [Note]

In general, FME's WMS client was able to read from both the Geomatys and ESRI WFS servers using WMS 1.1. However, the Geomatys WMS servers for UKHO did not return any results that the client could consume. It is likely that these services may no longer be supported at the time tested, perhaps because they were only operational for Phase 1. However, all the services associated with Phase 2 were able to be read.

Given the complexity of the data being served via WMS, it is not surprising that default reader

settings which combine all data layers did not produce useful results. The best results were rendered when the 'Split Request by Layer' and 'Transparent Background' options were enabled. Overall geo-referencing appeared accurate and the labels in the CHS datasets were legible at appropriate zoom levels. This could be confirmed by enabling background basemap layers such as Stamen maps as visible in the above screen shots. However, there are likely limitations on the rendering of symbology which would need to be tested to better assess.

Also, while it is unclear whether or not feature level interrogation is supported by these WMS services, the lack of support for this from within FME's WMS client would limit the usefulness for applications such as this. In general FME's WFS client is much more flexible for working with the range of rich features typical of S121 datasets than the WMS client. Still, FME's WMS client is useful for previewing and initial inspection of available datasets.

More detail on the FME Client SDI consumption of data outputs within the project is available in a comprehensive writeup within the project repository at:

https://knowledge.safe.com/articles/112578/ogc-maritime-limits-and-boundaries-s-121-gml-pilot.html

7.2.3. Oceanwise

Oceanwise did not produce a dedicated COTS client SDI but produced a comprehensive series of tests with open source and COTS SDIs. The extensive testing of these for consumption of web services and GML are included in the WMS/WFS testing section of this ER.

7.2.4. ESRI

ESRI produced a broad range of clients for reading and writing many geospatial datasets and many of the tests that were carried out within the Pilot, as well as sponsor data creation were performed using ESRI software. In terms of the Pilot the OGC web services clients utilizing SF profile were tested extensively (included in the WMS/WFS section of the ER).

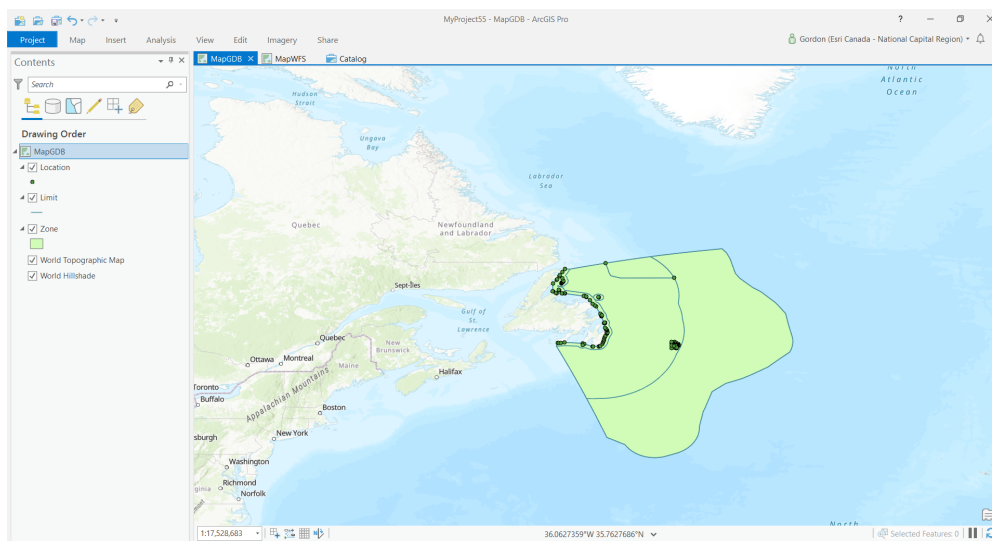A client project reading GML outputs is included in the pilot repository:

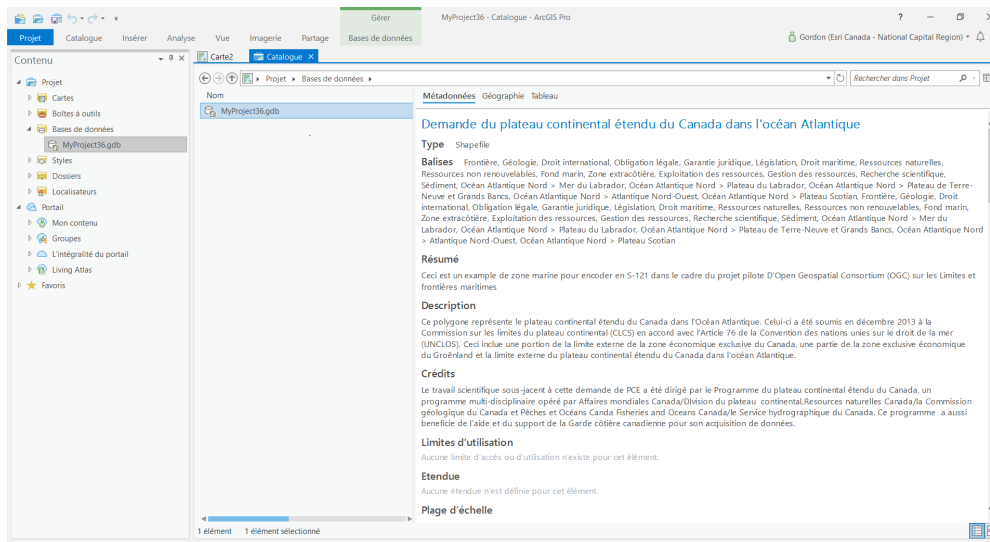*Figure 53. ESRI Client geospatial data* [Note]

Figure 54. ESRI client metadata [Note]

ESRI's implementation makes use of "flattened" schemas as described earlier in this section.

7.3. Open Source - GDAL

GDAL support for GML and complex GML is in place. As part of the project an exploration of the extent of the GDAL compatibility was explored. The key aspect of GML in respect of S-121 is its use of complex features. There are a number of complex features (indeed, complex features are an integral part of S-100 and a major innovation after IHO S-57). The metadata source information types contain an extensive set of complex features with one-to-many relationships between them. Additionally, any geographic feature (and ISO 19152 information types) can be associated with any number of ordered source information types.

In order to support this rich structure of associations and attribution within features, a comprehensive support for complex GML is required by a software system implementing S-121. Compatibility with complex GML under GDAL is provided by the GMLAS driver (http://www.gdal.org/GMLAS). This driver takes the GML data and its schema and creates a set of interlinked relational tables to resolve links and create structures representing the complex GML structure.

The only issue found with the GDAL implementation of complex GML (and its ability to parse and manipulate the data contained in a conformant GML file) is the identification of the geometric component of the data. GML does not provide a concrete method for determining which elements of a particular feature member type contains positional information (i.e., which parts of a feature can be rendered onscreen) so it is down to the implementing parser to establish which schema elements are geospatial in nature. Some later versions of the schema design were therefore not parsable by GDAL's GMLAS driver which caused ongoing issues with QGIS compatibility. The issue was traced through the GDAL source code to the namespace prefix used in the schema element representing the geometry.

This issue was raised with the GDAL moderator and an issue raised. It requires a more structured way of establishing these element types in the future. This impacts the ability of open source toolkits such as QGIS to parse and import data. Support for these tools is crucial for the project longer term in order to broaden testing and implementation by member states and other agencies.

It is hoped that the outputs such as the QGIS plugin, S121Loader will enable a greater base of support for the S-121 model.

7.4. Human Readable Text

7.4.1. Requirement Overview

Implement an XSLT that converts data following S-121 specification to a human readable format (HTML). In addition, produce output suitable for deposit to DOALOS satisfying the deposit requirement for Maritime Limits and Boundaries. This output format should encode features and attributes already defined within the S-121 feature model.

This section of the ER covers the methodology of taking data from a sponsor database and transforming it into something which can be used for both DOALOS deposit as well as the reproduction of published domestic legislation. Key issues considered are the underlying complexity of the network of relationships embedded within the GML MLB data. Although the requirements for textual rendering of S-121 data to satisfy national requirements may differ significantly between states (due to different national legislative structures) the example chosen from Canada is representative of the use case. The richness of the GML Application Schema framework and the capabilities of XSLT processing make the approach reusable in many different national contexts. This can be achieved through open source or COTS solutions, such as those used within the pilot but also require skills with XSLT technologies.

This requirement represents a core use case for the MLB project and an important "follow-on" use for S-121 data. The ability to render GML data into a definable "flat text" form underpins the driving use case for the S-121 standard itself. This poses some challenges for traditional geospatial data as representation of coordinates in textual form within legislative documents is at odds with their geospatial coordinates (expressed in decimal degrees within e.g., GML).

In response to this dichotomy, and in order to properly represent the legal source of S-121 features, the standard contains attribution specifically designed to textually hold geographic coordinates in the exact same form as in legislative documentation. These attributes, contained in a complex attribute called locationReference, have text fields which can hold any state-specific reference to a location in textual form.

The following extract from (https://laws-lois.justice.gc.ca/eng/regulations/C.R.C.,_c._1550/page-1.html) the Canadian Territorial Sea Geographic Coordinates Order shows some of the challenges, namely:

- Production of bilingual versions of the text;
- DMS representation of coordinates;
- A semi-structured document layout, with formatting arranged hierarchically;
- A mixture of free text and structured tabular coordinate lists; and
- References to charts (and, later, different reference systems).

SCHEDULE I

(ss. 2, 3 and 4)

AREA 1

Labrador

ANNEXE I

(art. 2, 3 et 4)

RÉGION 1

Labrador

	Column I	Column II	Column III	Column IV		Colonne I	Colonne II	Colonne III	Colonne IV
	Locality	Latitude	Longitude	C.H.S. Chart		Localité	Latitude	Longitude	Carte S.H.C.
1	Double Island	52°15′30″N.	55°32′58″W.	4701	1	Double Island	52°15′30″N.	55°32′58″O.	4701
2	Spear Point	52°26′37″N.	55°37′40″W.	4701	2	Spear Point	52°26′37″N.	55°37′40″O.	4701
3	Eastern Twin Island	52°40′20″N.	55°44′43″W.	4702	3	Eastern Twin Island	52°40′20″N.	55°44′43″O.	4702
4	Cooper Island	52°54′37″N.	55°47′26″W.	4702	4	Cooper Island	52°54′37″N.	55°47′26″O.	4702
5	Eddystone Island	52°58′55″N.	55°44′34″W.	4702	5	Eddystone Island	52°58′55″N.	55°44′34″O.	4702
6	S.E. Raven	53°12′23″N.	55°41′19″W.	4702	6	S.E. Raven	53°12′23″N.	55°41′19″O.	4702
7	Roundhill Island	53°25′58″N.	55°36′22″W.	4703	7	Roundhill Island	53°25′58″N.	55°36′22″O.	4703
8	North Wolf	53°43′06″N.	55°55′10″W.	4703	8	North Wolf	53°43′06″N.	55°55′10″O.	4703
9	Outer Gannet Island	54°00′00″N.	56°32′12″W.	4732	9	Outer Gannet Island	54°00′00″N.	56°32′12″O.	4732
10	Southeast Rock	54°14′58″N.	56°48′22″W.	4732	10	Southeast Rock	54°14′58″N.	56°48′22″O.	4732

Figure 55. Example extract from Canadian Terriotorial Sea Geographic Coordinates Order.

7.4.2. Methodology

It was determined two primary outputs were required.

1. **Human Readable Text (HRT), Legislative Output**, corresponding to the Legislative Documents that are produced by the Government of Canada (this could be used as an example for other countries). At a minimum, an HTML version of this text will be created using XLST.

2. **Human / Machine Readable Text, DOALOS Output**, that would be acceptable to UN DOALOS. This is a simplified version of the information contained in the **Legislative Output** and can be an exchange format using encoding instructions described in Annex B of the S-121 Product Specifications. This output must be verifiable against the original GML content.

The output for HRT is transformed from the GML produced by the data production tools.

This is reliant on:

- Schema conformant GML;

- Fully populated features including textual positions;

- Populated LADM features representing the textual preamble, headers and hierarchical structure of the text form; and

- Correctly attributed relationships between features and information types.

Proposed minimum content of DOALOS Output:

1. Metadata fields;

2. A single governance block with Title, Date and Description (set of text) attributes;

3. Basic Administrative (BA) Unit;

4. One (or more) Features (such as Territorial Sea Outer Limit);

5. One or more Spatial geometry objects (i.e. the geometry of the feature in a tabulated layout;

6. One or more Additional Spatial Attributes (i.e. Location by text);

7. One Party (e.g., Canada);

8. One Right (e.g., Sovereign Right); and

9. One or more Source objects.

It was decided to use, as an example, a dataset representing the Territorial Sea, extracted from a complete GML dataset and containing the required elements, by reference, including 'governance.' It was also determined that it would be an asset to the project that additional text output be produced in a variety of formats including, but not limited to, PDF and HTML. These formats may be created using Asciidoc as an intermediary format along side a simple report generator interface.

The workflow for the HRT is shown in the diagram below.
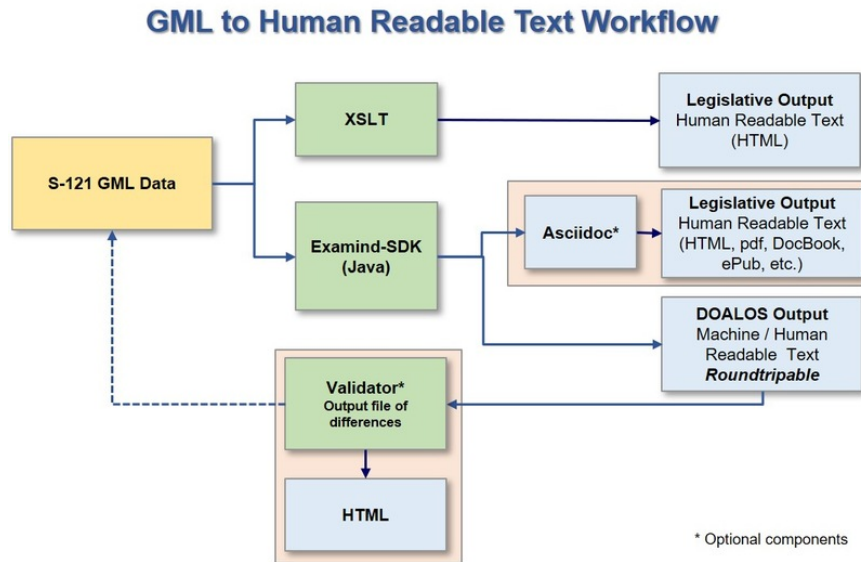
Figure 56. Human Readable Text Workflow

7.4.3. Deliverables

1. The **Legislative Output** can be generated using the XSL file (Canadian Legislative Output EN.xsl / Canadian Legislative Output FR.xsl) with the CHS S-121 data. It is possible to execute this XSL using the S-121 tools produced by Geomatys.

2. The **DOALOS Output** can be generated using the XSL file (DOALOS_EN.xsl / DOALOS_FR.xsl) with the CHS S-121 data. It is possible to execute this XSL using the S-121 tools produced by Geomatys.

An extract from the textual output is shown below.

```
Metadata:
Title: dataset04
Abstract:   Territorial Sea Geographical Coordinates Order
Topic: boundaries
Language:   en


Governance:
Reference Number:   C.R.C., c. 1550, ss. 4(1)
Title: SCHEDULE I
Description:    In respect of the portions of Area 2 for which the geographical
coordinates of points are listed in Schedule I, the baselines are straight lines
joining the points so listed.
Approval Date: 2019-01-01


Basic Administrative Unit:
basicAdministrativeUnitName:    Straight Baseline: Southeast and East Newfoundland
basicAdministrativeUnitType:    Maritime Limits and Boundaries
basicAdministrativeUnitContext: AREA 2


Limit Identifier:   Limit-1
Label:
Name:   Straight Baseline: Lamaline Shag Rock to Black Head North
Releasability   Controlled
Type Of Arc:    geodesic
The limit is defined by the following curve:    Limit-1-0
Vertex Identifier   Latitude    Longitude   Datum
Location-1  Lamaline Shag Rock  46°50⎕21⎕N   55°49⎕30⎕W  4016
Location-2  Shag Rock   46°50⎕17⎕N  55°44⎕51⎕W  4016
Location-3  Offer Island    46°51⎕21⎕N  55°37⎕25⎕W  4016
Location-4  Ferryland Head  46°52⎕16⎕N  55°23⎕04⎕W  4016
Location-5  Cape St. Marys  46°49⎕14⎕N  54°11⎕54⎕W  4016
Location-6  S.W. Bull and Cow   46°46⎕34⎕N  54°06⎕13⎕W  4016
Location-7  Shoal Point 46°36⎕50⎕N  53°35⎕12⎕W  4016
Location-8  Cape Freels 46°36⎕42⎕N  53°33⎕30⎕W  4016
Location-9  Mistaken Point  46°37⎕29⎕N  53°09⎕48⎕W  4016
```

The French version is shown below:

```
Métadonnée:
Titre:  dataset04
Description:   Décret sur les Coordonnées Géographiques pour la Mer Territoriale
Topique:   boundaries
Language:   en


Governance:
Numérode référence: C.R.C., ch. 1550, art. 4(1)
Titre:  ANNEXE I
Description:    En ce qui concerne les secteurs de la région 2 pour lesquels les
coordonnées géographiques de points sont établies dans l'annexe I, les lignes de base
sont des lignes droites reliant les points qui sont ainsi indiqués.
Date approuvée: 2019-01-01


Unité administrative:
nom:    Ligne de base droite: Sud-est et Est de Terre-Neuve
type:   Maritime Limits and Boundaries
context:    RÉGION 2


Limit Identifiant:  Limit-1
Label:
Nom:    Ligne de base droite: Lamaline Shag Rock à Black Head North
Releasability   Controlled
Type d'arc: geodesic
La limite est définie avec les courbes suivantes:   Limit-1-0
Identifiant Latitude    Longitude   Datum
Location-1  Lamaline Shag Rock  46°50'21"N  55°49'30"O  4016
Location-2  Shag Rock   46°50'17"N  55°44'51"O  4016
Location-3  Offer Island    46°51'21"N  55°37'25"O  4016
Location-4  Ferryland Head  46°52'16"N  55°23'04"O  4016
Location-5  Cape St. Marys  46°49'14"N  54°11'54"O  4016
Location-6  S.W. Bull and Cow   46°46'34"N  54°06'13"O  4016
Location-7  Shoal Point 46°36'50"N  53°35'12"O  4016
Location-8  Cape Freels 46°36'42"N  53°33'30"O  4016
Location-9  Mistaken Point  46°37'29"N  53°09'48"O  4016
Location-10 Cripple Rock Point  46°38'29"N  53°06'08"O  4016
Location-11 Big Johns Point 46°38'36"N  53°05'51"O  4016
Location-12 Unnamed Peninsula   46°38'45"N  53°05'32"O  4016
```

A version of the data rendered into HTML is shown in the following image.

Figure 57. Human Readable Text Rendered into HTML

Note the retention of place names in their default language.

7.4.4. Summary of Deliverables

The deliverables show how GML encoded S-121 data can be used to generate textual output for both the major use cases under consideration. Detailed outputs are contained in the project deliverables and, even as a proof of concept, are powerful examples of how the data can be manipulated to form the correct outputs required.

Key challenges overcome in the creation of the HRT versions of the data are as follows.

- The use of textual positions overriding those found in the geospatial position elements of the GML. This transform focused on the locationReference attributes of the geospatial features.

- A much broader requirement was identified for multi-lingual features. Centered around the LADM features (where much of the content within the textual output is defined) the requirement is to support multi-lingual translations of any element of the dataset which may be output to an end user, for example names for other features, description fields, column headings, and metadata. These are not all supported in a multi-lingual context within the current feature catalogue or GML schema and may require some further analysis to arrive at an optimal solution.

- Establishment of an appropriate set of relationships between the various geospatial features and information types is an important aspect of structuring the database forming the source for the HRT. This generated many iterations and discussions within the project team. ISO 19152 defines several use cases, but those are predominantly based on land cadastral use cases and there is, as yet, no currently defined set of norms which can be used to guide implementers. The legislative text is a prime example of where documentation in support of the standard would assist greatly implementing data producers. Ultimately a very flexible scheme based on the ability to form arbitrary hierarchies using consistsOf/composedOf relationships was implemented. This allows implementing states to define hierarchies and collections which mirror their preferred structures.

7.5. Metadata

Metadata was approached within the Pilot with the express intent of implementation, via OGC CSW a metadata service. Metadata provided by sponsoring organizations was examined and processed into such services and then tested for search capabilities by selected COTS and open source clients. Many issues with metadata are generic to OGC web services and there are, of course, state specific

implementations which were researched by participants. The results provided here focus on the metadata requirements contained within the Call For Participation.

7.5.1. Overview

The S-100 Universal Hydrographic Model requires that all compliant Product Specifications have a section on Metadata. The purpose is to allow for data discovery, enables data exchange and describes essential aspects of the data. S-121 only requires the discovery metadata.

The S-121 metadata profile complies with S-100 V4 and is similar to the equivalent section in S-101 ENC Product Specification. The original version of S-100 was based on the ISO metadata standard 19115:2003. The S-100 standard has been revised in a compatible manner to address ISO 19115-1:2014, and the XML encoding is based on ISO 19115-3:2016.

Note: ISO has since decided that XML encodings will be derived directly from the base standards and since separate standards such as 19115-3 get out of step with the base standards, but S-100 is still based on 19115-3. This means that there will be another change of encoding coming within the next two years.

S-121 makes use of only the elements of ISO 19115-1. Imagery metadata from ISO 19115-2 and quality metadata from ISO 19157 are not used. S-121 contains only Data Discovery metadata. This narrows the scope of what metadata is needed for S-121.

All other attributes are described explicitly in the standard. There is no need for Quality Metadata because official data, such as S-121 Maritime Limits and Boundaries is declarative data and as such has absolute accuracy.

The S-121 Discovery Metadata is in conformance with the latest version of S-100, which now makes use of the ISO 19115-1 standard from 2014. This is a change from the metadata that the S-121 project originally started with because the older version of S-100 used the older ISO 19115:2003 standard.

Most of the world still uses the older ISO 19115:2003 standard because it is very difficult to update metadata in legacy data; however, since S-121 is a new standard it should use the newer metadata standard. Basically, legacy data rarely gets updated. Few have the budget to fix old data. Once one has acquired the right data it can be converted to a newer format, but for discovery older data remains coded using the older version of the standards in which it was created. This means that catalogue services need to be able to process both old and new encodings to find data.

The IHO S-100 based GML schema uses the new standard so S-121 must use the new metadata to make use of the S-100 GML schema.

In the OGC MLB Pilot there was a decision to make use of the Canadian Government community profile of ISO 19115:2003 (Harmonized North American Profile - HNAP) which is based on the older ISO standard. This defines the 22 metadata fields that need to be used, but not necessarily the XML encoding.

This was initially thought to be fine for the Pilot. However:

- Since the Pilot needs to use the latest S-100 GML schema, the XML tags therefore must align with

the new ISO metadata encoding; and

- The Metadata fields that are required are defined in S-100 Part 4a and the equivalent S-121 clause 12.

The list of metadata fields was defined by the HNAP. These are the 22 fields listed below (list provided by Canadian Hydrographic Service, from DFO Metadata Training Report).

1. Title
2. Abstract
3. Reference
4. ISO Topic Category
5. Extent
6. Descriptive Keywords
7. Spatial Representation
8. Dataset Language
9. Status
10. Resource-Maintenance
11. Reference System
12. Resource Constraints
13. Contact
14. Distribution Information
15. File Identifier
16. Language
17. Character Set
18. Hierarchy Level
19. Parent Identifier
20. Metadata Standard Name
21. Metadata Date Stamp
22. Locale

Additionally, **Dataset URI** can be optionally included.

The following is a list of the 22 HNAP metadata fields coded in the modern ISO 19115-3:2016 encoding.

S-121_Dataset Discovery Metadata (as required by S-100)

Name	Path (in ISO 19115-1)	Obligation	Type	Metadata element per Pilot Project list	Note
Metadata file identifier	MD_Metadata.metadataIdentifier > MD_Identifier.code	M	CharacterString	15 File identifier	Backward Compatible – revision adds extra options
Metadata language	MD_Metadata.defaultLocale > PT_Locale.language	C (documented if not defined by the encoding process)	Code lists for : Language, and optionally CountryCode to identify dialect,	16 Language of the Metadata	Backward Compatible – uses Locale approach
Metadata character set	MD_Metadata.defaultLocale > PT_Locale.characterEncoding	C (documented if ISO 10646-1, is not used and not defined by the encoding process)	Code list from IANA (see S-100 B.3.14) Default "utf-8"	17 Character Set	Backward Compatible – uses Locale approach
Metadata file parent identifier	MD_Metadata.parentMetadata > CI_Citation.identifier	C (documented if the hierarchy of a higher level exists)	Class MD_Identifier	19 Parent Identifier	Backward Compatible – revision adds extra options
Party responsible for the metadata information	MD_Metadata.contact > CI_Responsibility.CI_Individual (table 4a-2) or MD_Metadata.contact > CI_Responsibility.CI_Organisation (table 4a-3)	M (either organization or individual must be documented)	Class CI_Individual (S-100 table 4a-2) or CI_Organisation (S-100 table 4a-3)	13 Contact for metadata	Backward Compatible – revision restructured
Date(s) associated with the metadata	MD_Metadata.dateInfo > CI_Date	M (creation date required, other dates may be provided)	Class DateTime data type and DateTypeCode	21 Metadata Date Stamp	Backward Compatible – revision restructured
Resource title	MD_Metadata.identificationInfo > MD_DataIdentification.citation >CI_Citation.title	M	CharacterString	1 Title	Backward Compatible
Resource reference date	MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.date > CI_Date.date	M	Class DateTime data type	3 Reference Date	Backward Compatible
Resource reference date type	MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.date > CI_Date.dateType > CI_DateTypeCode	M	Code list DateType	3 Reference Date	Backward Compatible
Abstract describing the resource	MD_Metadata.identificationInfo > MD_DataIdentification.abstract	M	CharacterString)	4 Abstract	Backward Compatible
Resource default language	MD_Metadata.identificationInfo > MD_DataIdentification.defaultLocale > PT_Locale.language	M	Code lists for : Language, and optionally CountryCode to identify dialect,	8 Dataset Language	Backward Compatible – uses Locale approach

Figure 58. Metadata Fields coded in ISO19115-3:2016

Name	Path (in ISO 19115-1)	Obligation	Type	Metadata element per Pilot Project list	Note
Resource default character set	MD_Metadata.identificationInfo > MD_DataIdentification.defaultLocale > PT_Local.characterEncoding	C (documented if UTF-8 is not used)	Code list from IANA (see S-100 B.3.14) Default "utf-8"	22 Locale character Set	Backward Compatible – uses Locale approach
Topic category	MD_Metadata.identificationInfo > MD_DataIdentification.topicCategory	M	Topic Category from Code List	4 ISO Topic Category	Backward Compatible
Geographic location of the resource (by description)	MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent > EX_GeographicDescription.geographicIdentifier > MD_Identifier.code	C (See notes 3 and 4)	Decimal	5 Extent	Backward Compatible
West longitude	MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent > EX_GeographicBoundingBox.westBoundLongitude	C (See notes 3 and 4)	Decimal	5 Extent	Backward Compatible
East longitude	MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent > EX_GeographicBoundingBox.eastBoundLongitude	C (See notes 3 and 4)	Decimal	5 Extent	Backward Compatible
South latitude	MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent > EX_GeographicBoundingBox.southBoundLatitude	C (See notes 3 and 4)	Decimal	5 Extent	Backward Compatible
North latitude	MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent > EX_GeographicBoundingBox.northBoundLatitude	C (See notes 3 and 4)	Decimal	5 Extent	Backward Compatible
Name of the scope/type of resource for which the metadata is provided	MD_Metadata.metadataScope > MD_MetadataScope.resourceScope > MD_ScopeCode (codelist – ISO 19115-1)	M (default = "dataset")	Code list	18 Hierarchy Level	Backward Compatible – revision restructured
Description of scope of resource for which the metadata is provided	MD_Metadata.metadataScope > MD_MetadataScope.name	O	CharacterString		Optional addition from S-100

NOTE1 ISO 10646-1 - Information technology — Universal Multiple-Octet Coded Character Set (UCS).
NOTE2 MD_ServiceIdentification may be used instead of MD_DataIdentification if hierarchyLevel = 'service'.
NOTE3 For a geographic dataset, include metadata for the geographic bounding box (West longitude, East longitude, South latitude and North latitude) or the geographic description identifier (The use of geographic bounding box is recommended - see Section 5.6.3).
NOTE4 If any one of west longitude, east longitude, south latitude or north latitude exists, then the remaining three must also be completed.

Figure 59. Metadata Fields coded in ISO19115-3:2016

Name	Path (in ISO 19115-1)	Obligation	Type	Metadata element per Pilot Project list	Note	Defaulted to
Descriptive Keyword	MD_Metadata.identificationInfo > MD_DataIdentification.descriptiveKeywords >MD_Keywords	O	CharacterString	6- Descriptive Keyword	Backward Compatible – made optional	Defaulted to: "Maritime Limits and Boundaries" Thesaurus reference: "UN Convention on the Law Of the Sea (UNCLOS)"
Spatial Representation Type	MD_Metadata.identificationInfo > MD_Identification.spatialRepresentationType > MD_SpatialRepresentationTypeCode	O	Code List Value	7- Spatial Representation Type	Backward Compatible – revision restructured	Defaulted to: "001 Vector"
Status	MD_Metadata.identificationInfo > MD_Identification.status > MD_ProgressCode	O	Code List Value	9- Status	Backward Compatible	Defaulted to "complete"
Resource Maintenance	MD_Metadata.identificationInfo > MD_Identification.resourceMaintenance > MD_MaintenanceInformation. maintenanceAndUpdateFrequency > MD_MaintenanceAndUpdateFrequencyCode	O	Code List Value	10-Resource Maintenance	Backward Compatible	Since the data represents a UNCLOS deposit the value would be 009 "data is updated as deemed necessary"
Reference System	MD_Metadata.referenceSystemInfo > MD_ReferenceSystem > MD_ReferenceSystemTypeCode	O	Code List Value	11-Reference System	Backward Compatible - Restructured	Defined within the dataset per element
Resource Constraints	MD_Metadata.identificationInfo. resourceConstraints > MD_Constraints.useLimitations	O	CharacterString	12-Resource Constraints	Backward Compatible - Restructured	To be included when the data is stored in a system such as Canada FGP. Not needed for general Marine Limits and Boundaries data
Distribution Information	MD_Metadata.identificationInfo. distributionInfo > MD_Distribution	O	class	14-Distribution Information	Backward Compatible	To be included when the data is stored in a system such as Canada FGP Resource constraints to be defined by the producing nation.
Metadata Standard Name	MD_Metadata.metadataStandard > CI_Citation.title	O	CharacterString	20-Metadata Standard Name		Defaulted to "IHO S100 Part 4"
Locale		intrinsic	Code lists for : Language, CountryCode, CharacterSet	22- Locale	Backward compatible – now an intrinsic part of 19115-1	Defaulted to code from ISO 629-2 such as "eng" for English or "fra" for French, optionally a code from ISO 3166-1 for country code to identify dialect, and code from IANA for character set defaulted to "utf-8"

Figure 60. Metadata Fields coded in ISO19115-3:2016

7.5.2. Recommendations

The recommendation is to use the 22 metadata fields including optional use of dataset URI, as identified in the Canadian Government profile of the ISO 19115 metadata, but use the modern ISO 19115-3:2016 encoding.

The experience in using the data in testing tools was that access to the metadata can be very slow. The root cause was traced to the HNAP registry which does individual searches through a large file for every codelist lookup. The production server also appears to be identifying metadata code list values by indirect reference through a registration number instead of via "reference by value" (which means the actual values are used and do not need to be looked up every time).

- S-121 should use the identified list of 22 HNAP metadata fields, but should use the ISO 19115-3:2016 encoding in alignment with S-100 V4.

- NRCAN should be approached with a view to improving performance of the HNAP codelist

server.

- Not all clients suffered a consistent slowdown which is perhaps due to caching of results. Clients should be encouraged to cache results where possible.

Chapter 8. Data Interoperability Testing

In this section of the ER, we look at the results of the interoperability testing carried out by project participants. We aim to show how the evolving GML application schema was used to encode sponsor data and how compatible that data is with different software configurations and packages, as well as to what extent the data produced exercised the individual sections of the S-121 model.

As of the time of writing the following datasets were available to the project participants.

- Canada - Territorial Sea Portion

- Canada - Arctic Phase 2

- Canada - Baseline

- Canada - Extended Continental Shelf

- Canada - International Boundary Commission

- Australia - Continental Shelf and Treaty with New Zealand

- Australia - Master Version B with Examples

- Australia - Theoretical Examples

- UK - Jersey and Granville

- UK - Guernsey

- UK - Isle of Man

- UK - Territorial Sea Limits

Additionally, during development of the main GML Application Schema, a test dataset representing a minimum set of representative S-121 features was compiled and used to test initial compatibility with participant software and conformance with both GML and the S-100 GML profile.

8.1. Construction of Test Datasets

During Phase 1 of the project many of the activities centered on examples which are representative of broad categories of S-121 data. It is important to realize that sponsor data was submitted to the project in bespoke formats - due to the nature of S-121 no comprehensive encoding for data exchange existed and therefore individual transformations into the S-121 GML Application Schema had to be constructed for each dataset and then metadata attributed individually. Indeed, many of the activities in Phase 2 of the project revolved around such transformations.

Due to the nature of S-121 and its associations with information types often source, governance and LADM structures were inserted manually as well or needed transformation from existing bespoke database structures.

For these reasons there is a sliding scale of "compatibility" with the GML Application Schema for IHO S-121 which should be appreciated. Although a dataset can be fully conformant with the schema, each of the following individual packages making up the product specification are optional:

1. Representation of Geographic, geospatial features present in Sponsor data, e.g., Territorial Sea Areas, Extended Continental Shelf, Exclusive Economic Zone;

2. Attribution of Source information types to each geospatial feature; and

3. Inclusion of ISO 19152 LADM associated features to datasets.

Therefore, a dataset can be conformant and yet have no ISO 19152 elements within it. Initial testing focused on data made with bespoke methods. Later on the pilot, Phase 2 focused on construction of data as described in Data Creation - Conversion of Sponsor Data to GML.

In Phase 1 of the Pilot the following datasets were created.

- Canada Extended Continental Shelf was transformed by multiple participants into conforming datasets under versions 3.6, 3.7, and 3.8 of the GML Schema and also to demonstrate backward compatibility with v.3.2.1 of GML. This version was also used for testing metadata attribution and RRR structures with sponsors giving valuable assistance.

- UKHO datasets. UK Territorial Sea Limits attributed with metadata supplied and guided by UKHO staff.

- Canada (IBC) Juan De Fuca dataset, converted via FME.

- simpleFeatureCollection - although a simple dataset this was the first to be created with multiple versions of the Application Schema and represents one of the simplest examples possible. This basic example was also used to test geometry by reference and compare that with inline geometry encoding.

Each dataset used different facets of the GML Application schema, namely:

- Geospatial features (All datasets should contain these);

- Dataset metadata (All datasets should have a minimum set of these but not necessarily all of them);

- Geometry - either fined purely by reference or inline geometry;

- Source information - Source information fully attributed;

- Additional Spatial Information - later renamed to "locationReference:" this is where geospatial locations are described textually alongside their GML location encoding;

- Governance and BAU relationships fully linked; and

- LADM rights, restrictions and responsibilities - both simple and complex examples.

8.2. Encoding of LADM RRR Data

A substantial part of the S-121 standard is concerned with the implementation of ISO 19152 structures, representing the Land Administration Domain Model (LADM). The ISO 19152 standard is a conceptual model which represents the domain of land administration, specifically rights, responsibilities and restrictions (RRR) affecting land and water. It has two overall goals:

- An extensible basis for the development of land (and, by implication, water) administration systems; and

- The enablement of communication between interested parties of those structures contained within the model, based on a shared vocabulary.

Many of the Information Types defined in S-121 are concerned with the implementation of ISO 19152 and the attribution which reflects the details of the ISO 19152 model. The primary reference dataset used within the project is the Canadian continental shelf data. For the project, the team worked extensively with the sponsors to refine the structure, fit it to the S-121 model and build the relevant structures which reflect the RRR associated with the data. These are implemented in a relational model using table joins in a GIS and are translated, via the FME workspace, to named relationships as defined in the IHO S-100 framework. An updated, draft IHO feature catalogue, received from Teledyne Caris at the conclusion of the pilot, has refined the relationships from an S-100 standpoint, and will form part of an update to the S-100 feature catalogue for S-121 in the forthcoming point release of the standard.

The RRR structures in the S-121 standard are embodied in the following information types:

- Basic Administrative Units;
- Governance;
- Rights, Responsibilities, and Restrictions;
- The Geospatial features which comprise the individual UNCLOS-representative entities; and
- Party and Party Group entities.

An example structure, constructed to illustrate these relationships, is shown in the following diagram.
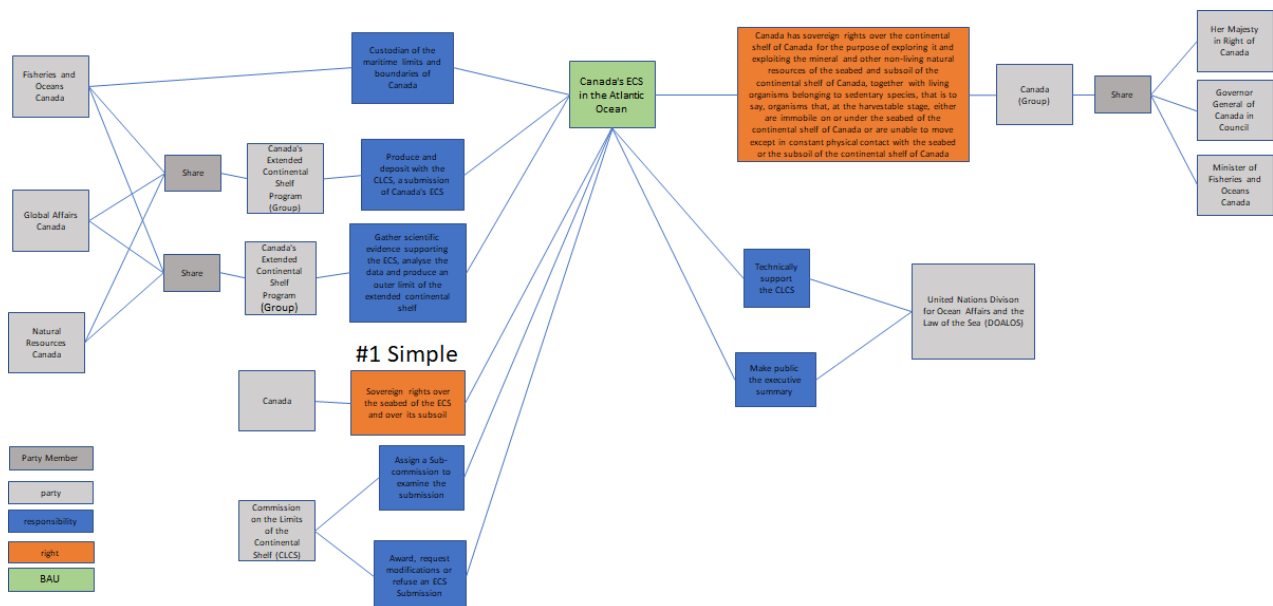
Figure 61. Relationship Structure

The RRR structure implements ISO 19152. In this structure the Basic Administrative Unit collects a number of geospatial features together and expresses the associated rights, Responsibilities and Restrictions associated with it along with an association to a number of individual (and group) parties. The diagram shows the RRR structure within the CHS dataset. In it a number of relationships are defined which conform to ISO 19152. Additionally, the aggregation of features allows, for example, Limits to be associated with their individual component parts.

8.3. Testing Overview

The rest of this section of the ER presents tangible results of the interoperability testing done within the project. This tests the compatibility of created data with more generic OGC services and tools. Much of the exhaustive testing presented here was performed by Oceanwise and comprehensively recorded visually. The goal of this section is to show the possibilities of how S-121 data can be opened up via web services once it is transformed into open standards.

The GML Schema and encoding, together with the transformation of sponsor data presents an enormous potential, particularly once extended components are added into the data content. The aim is to show how complex datasets representing not just geospatial features but their source and LADM RRR relationships can be distributed and graphically examined. S-121's use of complex GML and its support within various clients presents some challenges in terms of configuration and optimum use. S-121 is a test of many client's capabilities because of its complex, legislative nature and the results in this section show a broad variety of capabilities and functionality.

The OGC roadmap for WFS and WMS services is in a state of development and the move to OGC Open API based services could be a topic for future developments of these services, early consideration for this is given in section *9.4* of this ER which evaluates alternative encodings such as GeoJSON which would be useful for supporting OpenAPI. The ability to distribute data containing complex relationships (not just for feature naming and textual positional representation but relationships relating to ownership, responsibilities, and international transboundary cooperation and treaties) presents challenges which stretch the capabilities of spatial data and its producers.

This section presents a consolidated view of how data is able to be imported through web services from OGC web services using S-121 transformed from sponsor data sources in accordance with the GML Application Schema.

8.4. Web Services Testing

Test clients: QGIS is a Free and Open Source Geographic Information System. It can open most GIS format files and is generally fairly flexible so ideally suited to testing various sources of data. Testing was undertaken in QGIS Madeira 3.4.10 and 3.4.14.

https://qgis.org/en/site/forusers/download.html

8.4.1. Reading GML

QGIS can read GML both directly and through an Application Schema Plugin (GMLAS). The GMLAS plugin allows for the complex (i.e., nested) feature and information types defined in the Schema to be viewed, the simple GML reader does not.
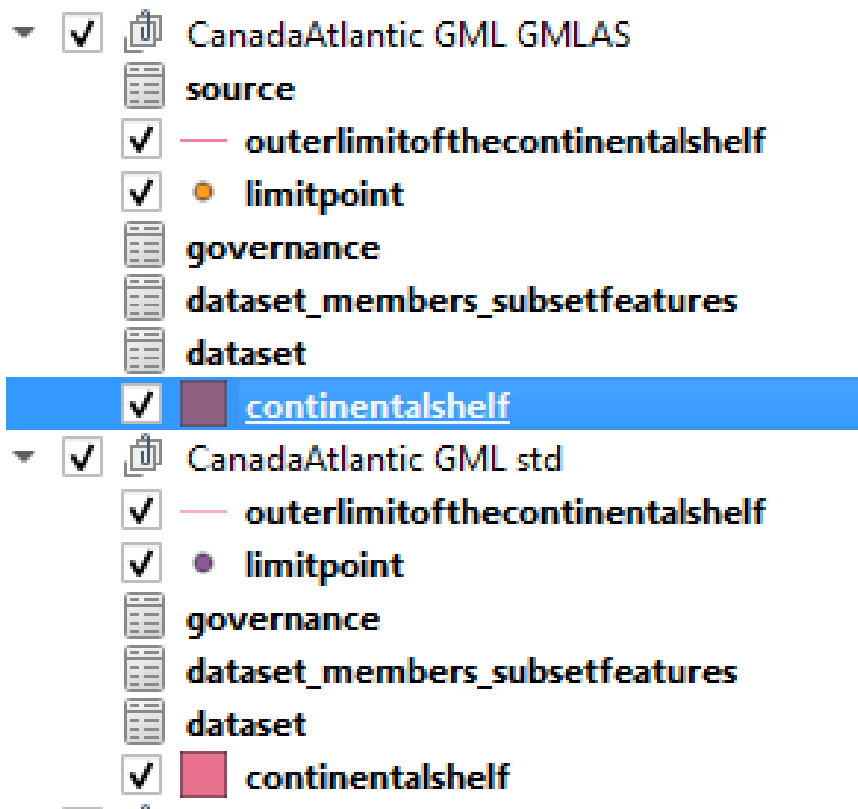
Figure 62. QGIS Layers

The effect this could have is that users who are unaware of the GMLAS plugin might load the GML and not be able to see some of the information about parties, responsibilities, source, etc. which are only available in the information types. The impact could be that the data is used incorrectly and decisions based on such data.
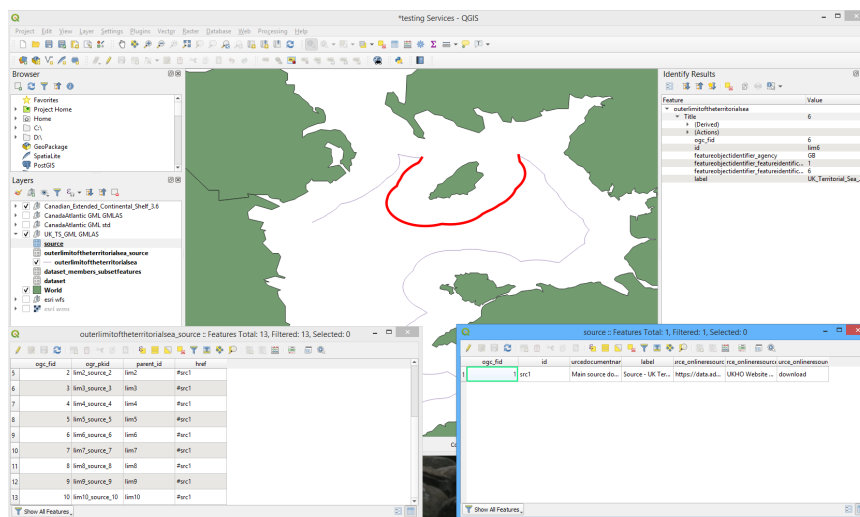

Figure 63. QGIS with GMLAS plugin data import

Using the CHS v1.0 of the GML, which has all the relationships encoded within it, and the GMLAS plugin, the data available is fairly comprehensive. The relationship tables allow for the relationships to be rebuilt and tables joined to the geometry. This allows the data to be analysed, queried and labelled based on the data from the information types. It is generally fairly obvious how to do this rebuilding based on the names of the tables, however it is beneficial to understand the S-121 structure.
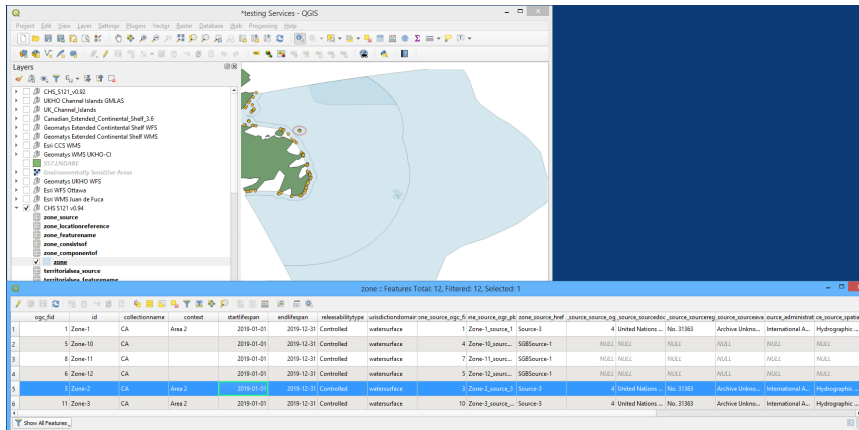
Figure 64. QGIS feature interrogation [Note]

In addition, QGIS has the concept of Relations within a Project. The GMLAS plugin QGIS can establish the relations between the geometry table and the complex attributes at the feature level i.e., it understands the relationship between the limit geometry and the limit_source table so these are available in the attribute view when interrogated (requires choosing table view to make it easier to see).
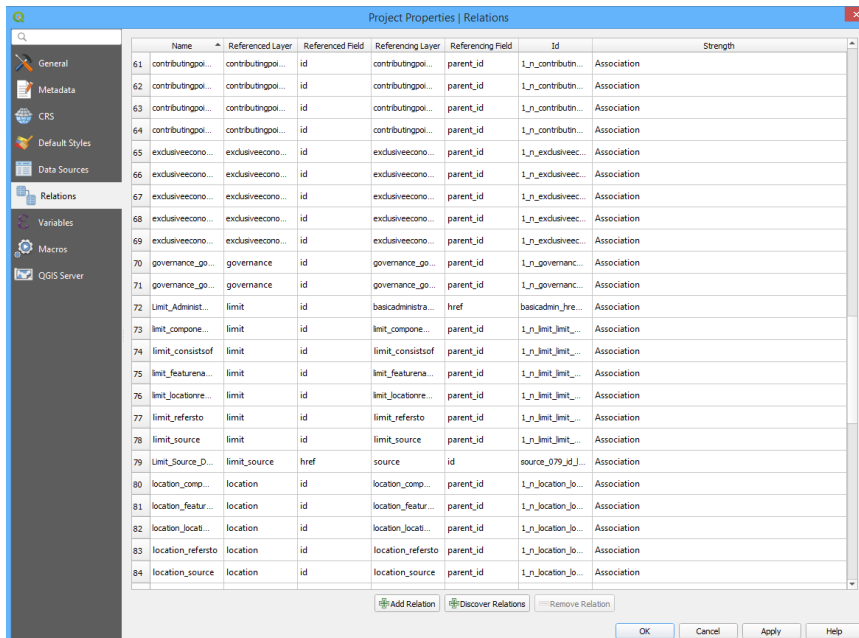
Figure 65. QGIS Relations Configuration [Note]

What doesn't appear by default seems to be the relationship between each feature and the InformationType it relates to. For example, the limit_source table and the Source do not have a relationship defined so the user has to manually do a join before the full details of each source are available on the Limit feature. This could be addressed by the custom plugin making the relations as the user imports the data.
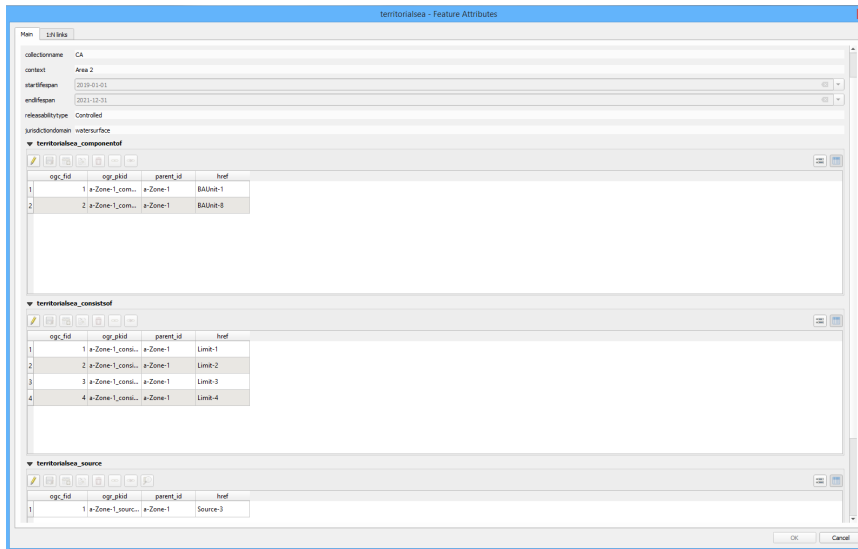
Figure 66. QGIS Relation editing

8.4.2. WMS

OGC WMS services can be read into QGIS. As these are received as images there is little further customization available. Each layer within the WMS can be loaded separately.
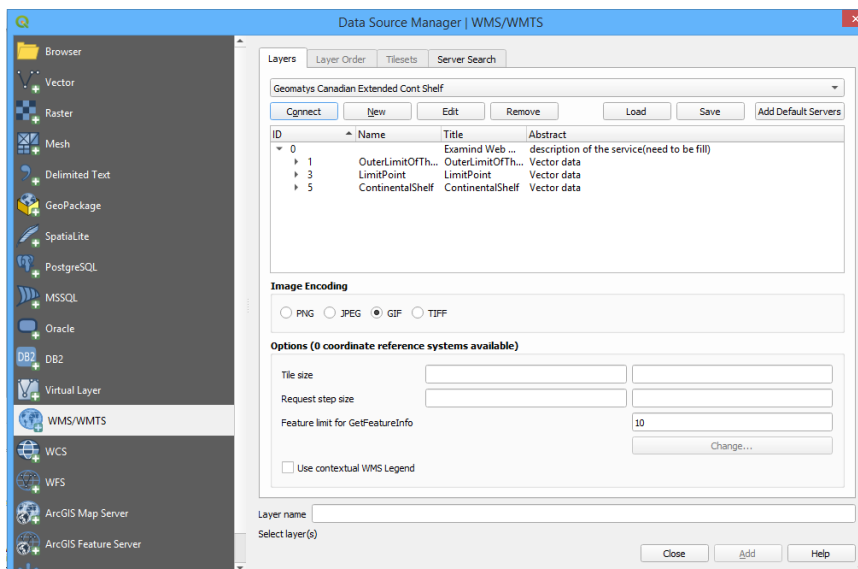
Figure 67. QGIS WMS Loading

QGIS does not have the option to force the layer to be transparent where there is no data. This then affects some services depending on image Format selected. This can be resolved by setting the Format to PNG or GIF but isn't necessarily the default setting for the service.
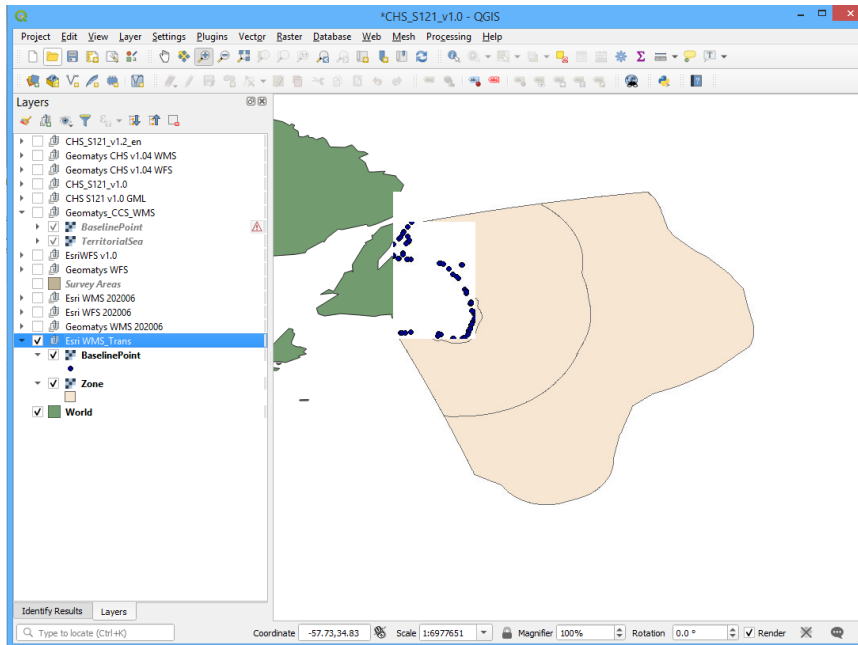
Figure 68. QGIS WMS Transparency Settings [Note]

Symbology for WMS is fixed by the serving organization, the end client cannot control this so standard symbology should be created for S-121 to make WMS from any organization interoperable. A discussion over symbology within the project assessed various candidate symbols for the feature scope of IHO S-121. Although the symbology of such layers is not within the scope of the MLB project it was recognized that in order to establish web services, such as WMS, there may be a requirement for some neutral symbols to be defined to assist implementation.

WMS has the additional benefit of being 'clickable' so an end user can click on a feature and find its attribute values. These are not queryable as part of analysis tools but at least provide access to some information on the features. For the S121 feature types this will mean that the associated Information Types may not be available to the end user as they will only see the attributes on the feature, not the additional tables e.g., Parties, Governance, Source. The only solution to this is for the serving organization to flatten the relationships in the source dataset before serving.
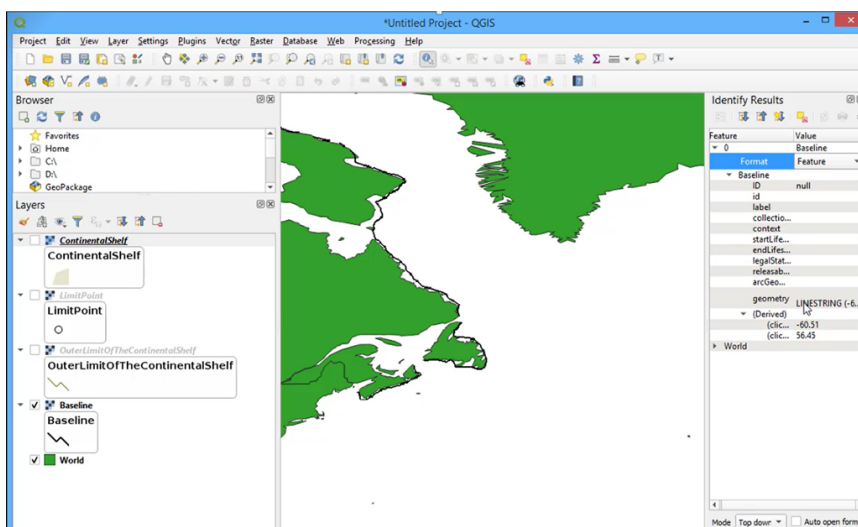
Figure 69. QGIS WMS Feature interrogation [Note]

8.4.3. WMS Example Services

8.4.3.1. Geomatys WMS

SOURCE: CHS_S121_v1.0.gml Geomatys have set up multiple WMS for various source datasets including Canadian Extended Continental Shelf and UKHO/Channel Islands Fisheries limits.
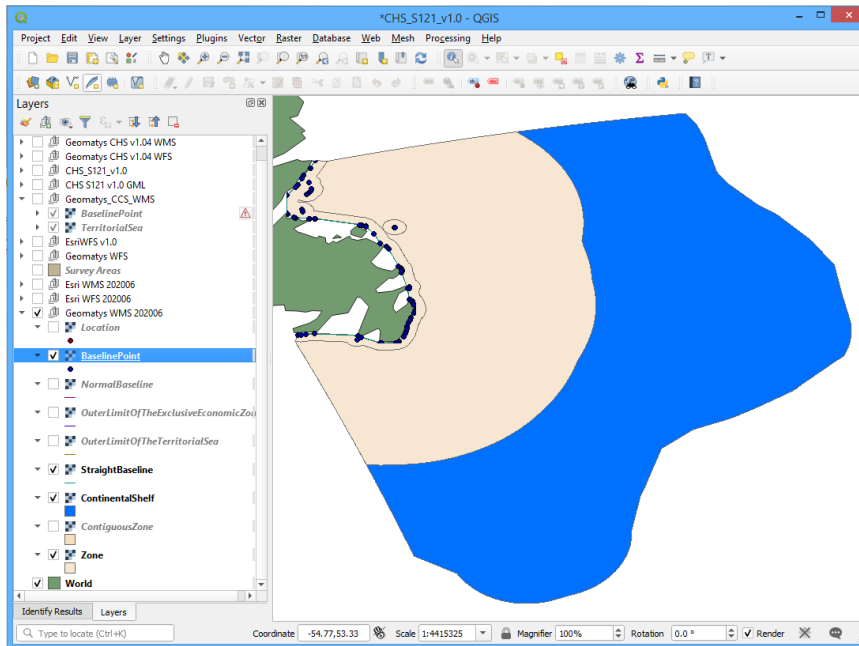
Figure 70. Geomatys WMS Extended Continental Shelf data. [Note]
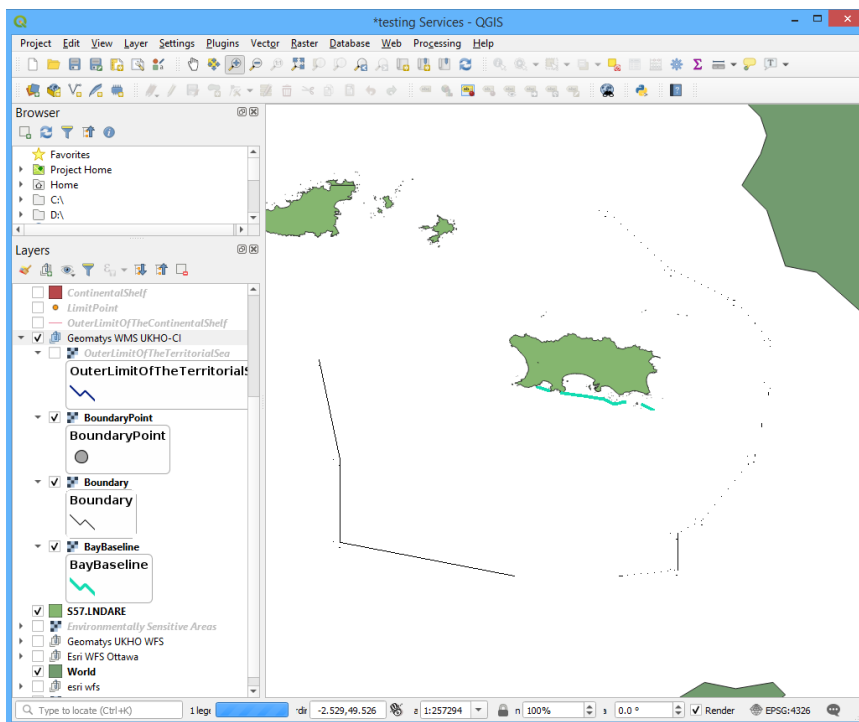
Figure 71. Geomatys WMS UKHO Channel Islands Sample data.

These load easily into QGIS and display the data in the correct location. The default settings for the layers are GIF so are transparent when loaded and can be combined with other datasets.

With version 1.0 the comprehensive attributes are available within the CHS WMS. They are slightly complicated to access and only available through the Identify tool. When the feature is selected the

Format needs to be set to Text and the full attributes can be viewed. As is the case with WMS the information cannot be easily queried so it makes it more difficult to use for any kind of analysis but at least the information is accessible to the end user.
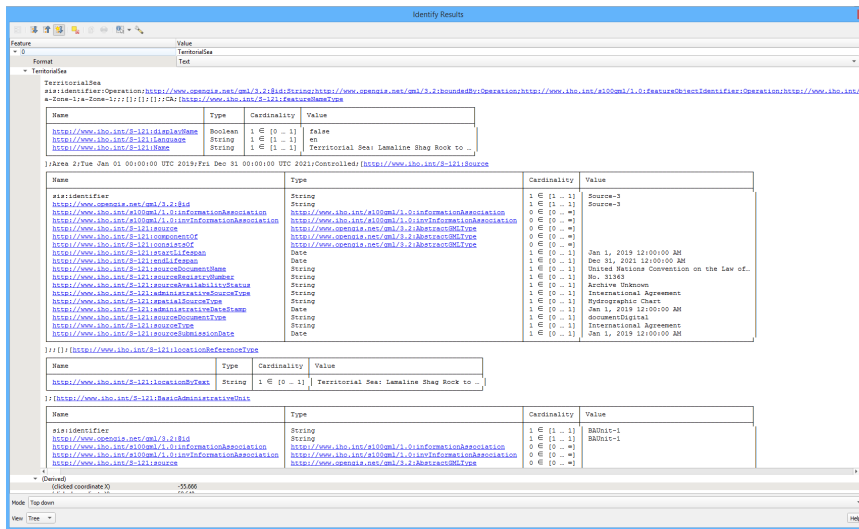
Figure 72. QGIS WMS Attribution access

8.4.3.2. Esri WMS

SOURCE: CHS_S121_v1.0.gml converted to Geodatabase.

Results: ESRI have setup the WMS for the pilot example datasets. Both load successfully into QGIS, however using the default settings (JPEG format) each layer was not transparent so hid other data layers. To fix this PNG or GIF format needed to be selected.

GetFeatureInfo requests to the WMS returned attribution. However not all S-121 populated in the dataset. This is because the source does not have those attributes populated and relationships built in.
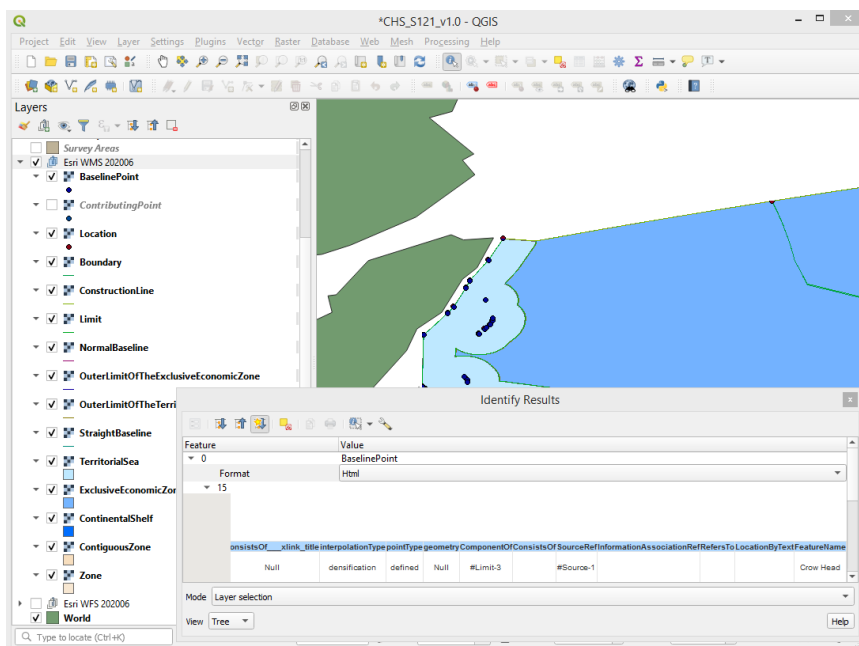
Figure 73. QGIS, ESRI EMS Attribution

As can be seen, the Continental Shelf Service displays correctly for all geometry types.
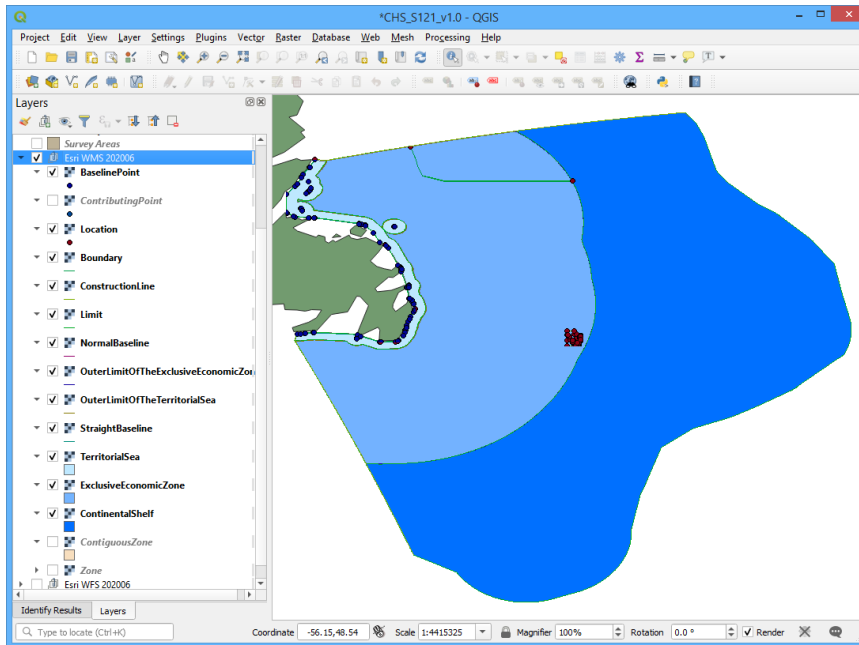
Figure 74. QGIS ESRI WMS CCS dataset display

This dataset is sourced from a flattened version of the S121 schema so displays the related attributes as concatenated attributes e.g., the consistsOf, componentOf attributes shown.
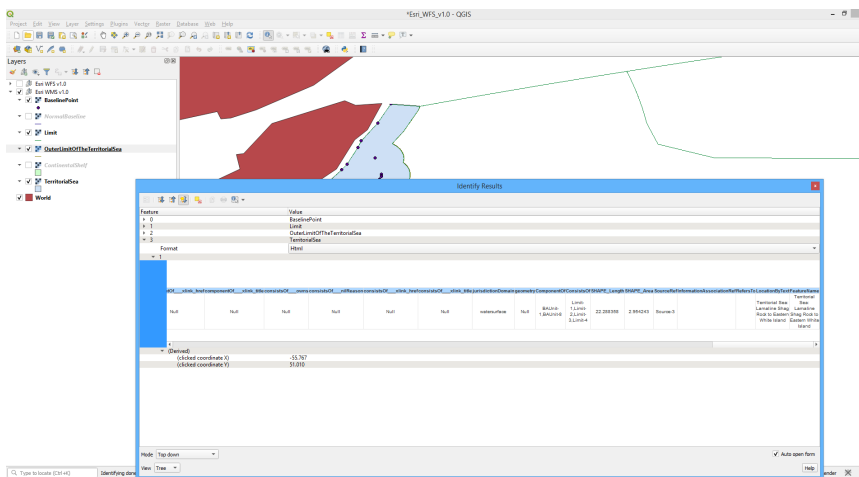

Figure 75. QGIS ESRI Flattened attribution.

It is not possible in QGIS to join non-spatial data to a WMS layer so although the end user can, for example, see the Source of Limit-3 is Source-1 they cannot link the actual Source data through to the layer to see the full details of Source-1.

A WMS does not allow non-spatial data to be viewed so the Information Types (Governance, Party, Source, etc.) are not available to the end user. This is a limitation of the WMS standard, rather than a limitation of QGIS.

A previous version of the ESRI WMS included a legend to identify the different feature types which makes it much easier for an end user to understand. This would be a useful addition to the v1.0 service.

8.4.4. WFS

Datasets streamed via WFS can be read into the software, with some customization available (e.g.,

Axes inversion), if required. Each layer within the WFS can be loaded individually. The WFS has the benefit that the Information types in the GML would be handled by the system serving out the WFS, therefore the Client does not need to handle it.

WFS can serve both spatial and non-spatial data. QGIS is able to handle the non-spatial data and make it available in tables. One issue with this is the actual relationships between the spatial and non-spatial (e.g., Party, Responsibility, AdditionalSpatialInformation) are not easily visible to the end user. This can mean it is difficult to interrogate and understand the intended usage of the data. This is common to WFS clients.

WFS become vector features within the client system so can be interrogated and used for analysis. This also means the end user can apply their own symbology to the data.

8.4.4.1. ESRI WFS

SOURCE: CHS_S121_v1.0gml converted to FGDB
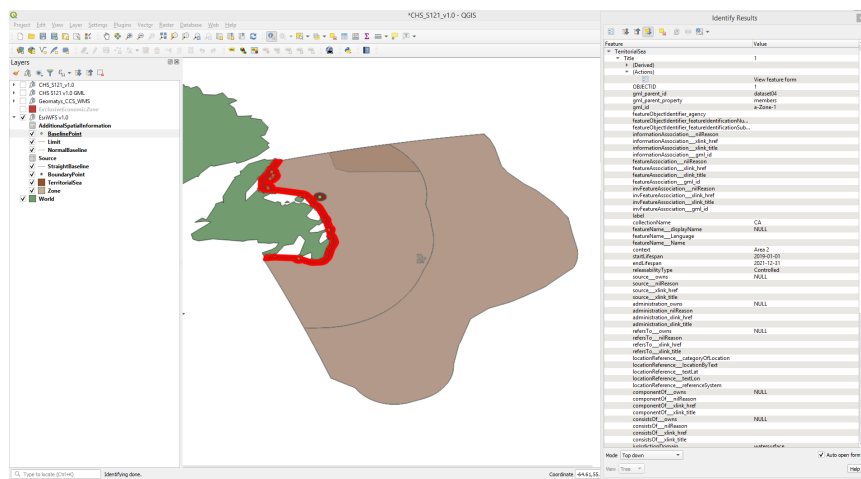
Figure 76. QGIS ESRI WFS converted to File GDB. [Note]

Esri have set up WFS for the Canadian Continental Shelf data. In earlier versions of the WFS using default QGIS settings the axis order served by the Esri system was read incorrectly, so it appears in the Southern Ocean. This is a known issue when other Clients read in Esri Services. This can be rectified by selecting Invert Axis Orientation in the Connection settings for the WFS. The Service will then load into the correct location.

For S-121 dataset version 1.0 the CRS definition was changed from EPSG to an OGC URI which explicitly denotes axis order and allows the data to be displayed automatically in the correct location.
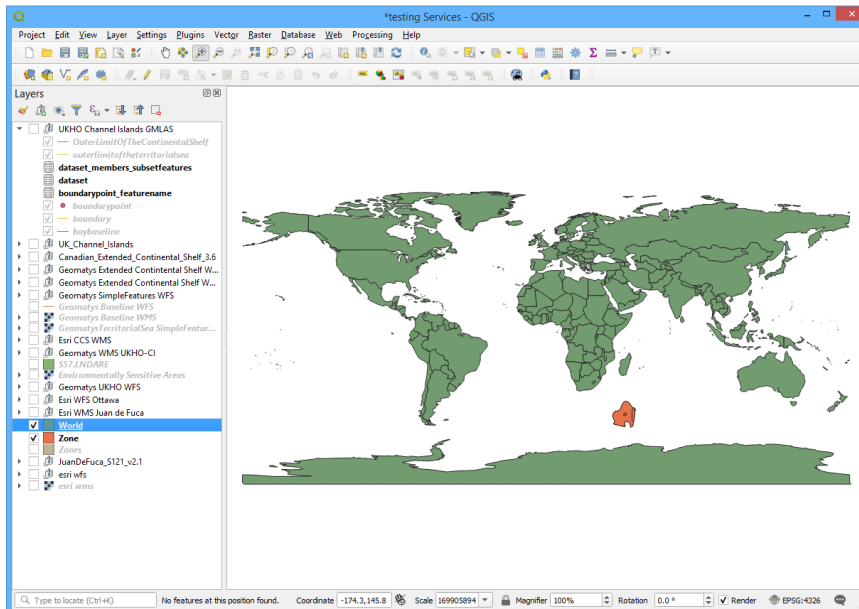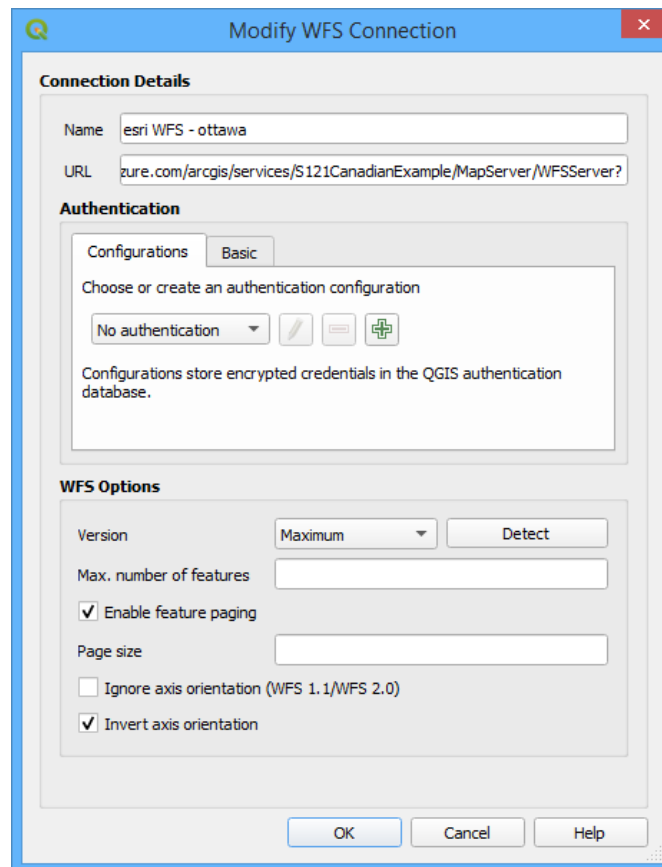
Figure 77. ESRI WFS Axis Issues
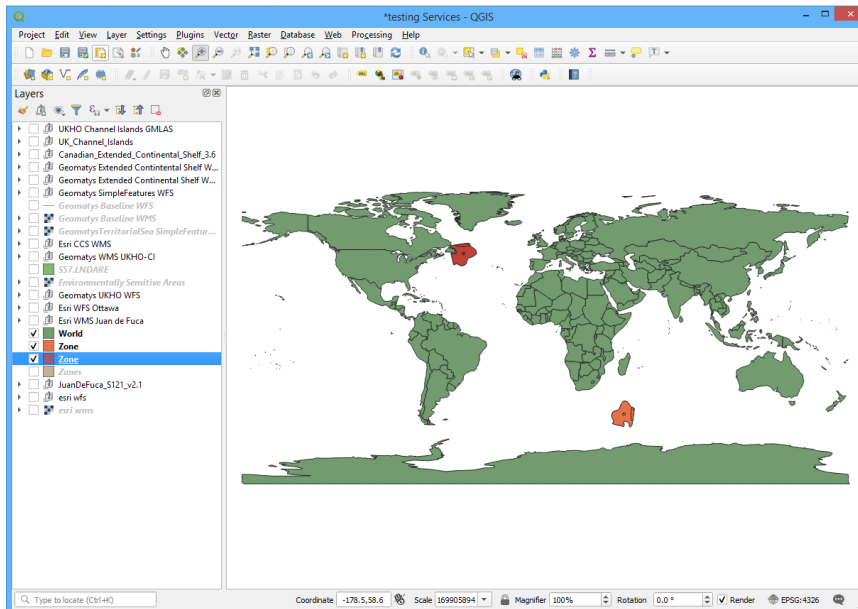

Figure 78. ESRI WFS Inverted Axes

Figure 79. ESRI WFS Corrected Axes [Note]

ESRI WFS also includes non-spatial data which appear as tables in the Layer list. These can be interrogated and analyzed, but do not obviously link to the spatial features as the relationships are not made available through the Service. With an understanding of the S121 structures the joins can be made to connect up the data in the same way they can for the GML source.
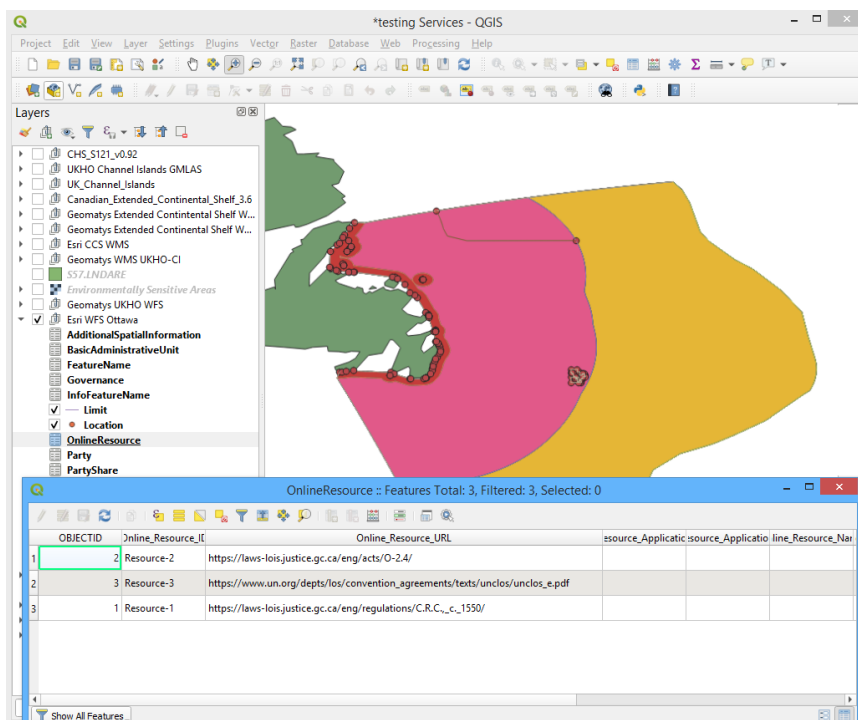

Figure 80. QGIS ESRI WFS joins with information types [Note]

ESRI have flattened the relationships and concatenated the reference information from the GML before serving so the basic attributes can be seen when a geometry feature is selected. This is a good way of combating the issues around the GML data being loaded using the standard loader and information being missing. As the WFS has to be Simple Features this method does mean that joining the Information Types is not possible as the Simple Features cannot handle the one-to-many relationships.

Figure 81. QGIS ESRI WFS concatenated attributes
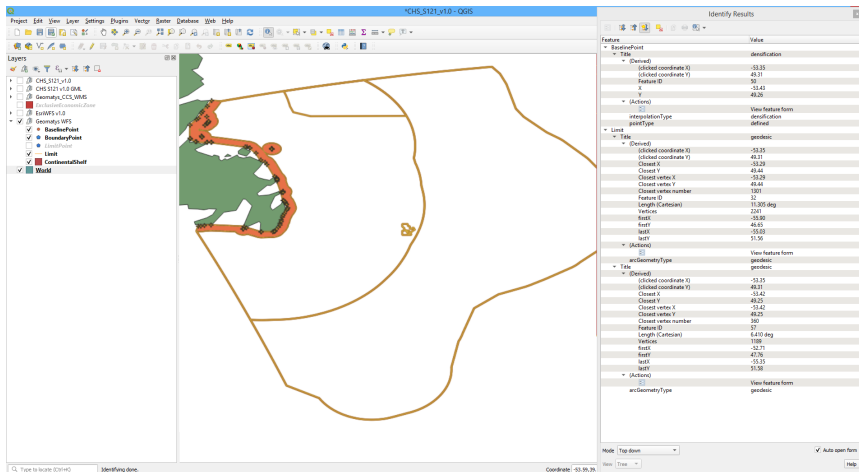
8.4.4.2. Geomatys WFS

SOURCE: CHS_S121_v1.0.gml

Figure 82. QGIS Geomatys WFS service [Note]

Geomatys set up WFS for CHS and the UKHO Channel Islands datasets. The WFS loaded in the correct place without Axis Inversion and the default format is GIF and therefore loaded with transparency enabled. Each layer within the WFS can be loaded separately (points, lines and polygons) and was easy to display all the information a user may require. The WFS features loaded as Vector items so can be used for analysis e.g., "Is the location I propose working in within the Exclusive Economic Zone of Canada?" It can also be symbolized by the user.
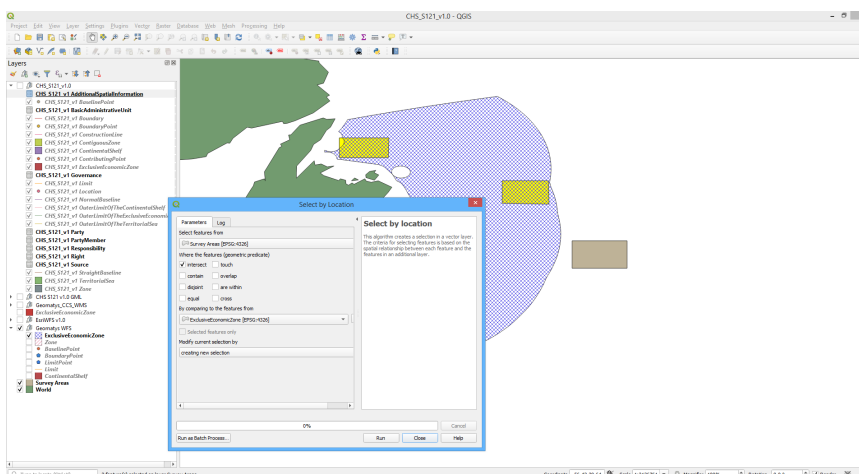

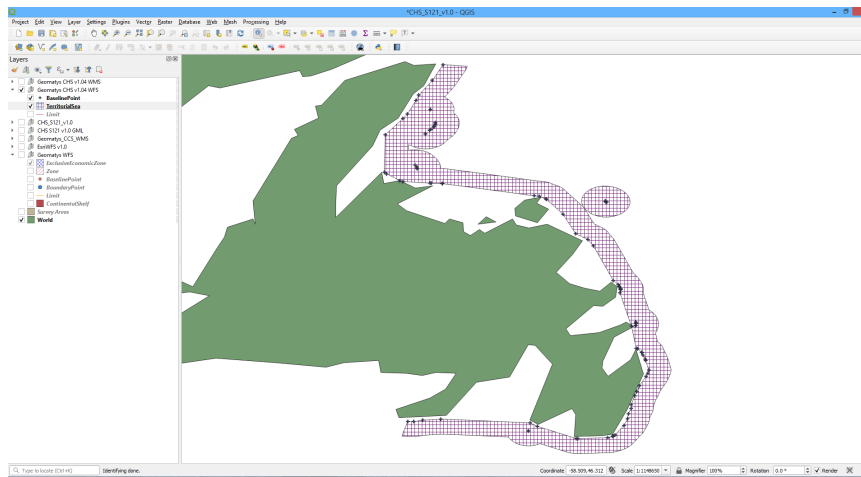*Figure 83. QGIS Geomatys WFS Vector Analysis* [Note]

Figure 84. QGIS Geomatys WFS Layers [Note]

Attribution in most of the source datasets used is limited, so doesn't display all S-121 attributes and InformationTypes. The CHS_S121_v1.04.gml sourced dataset would have all the S121 attributes and InformationTypes populated, however the ESRI implementation of the WFS standard only allows Simple Features so the complex relationships cannot be displayed. Non-spatial data is not available from the Geomatys WFS.
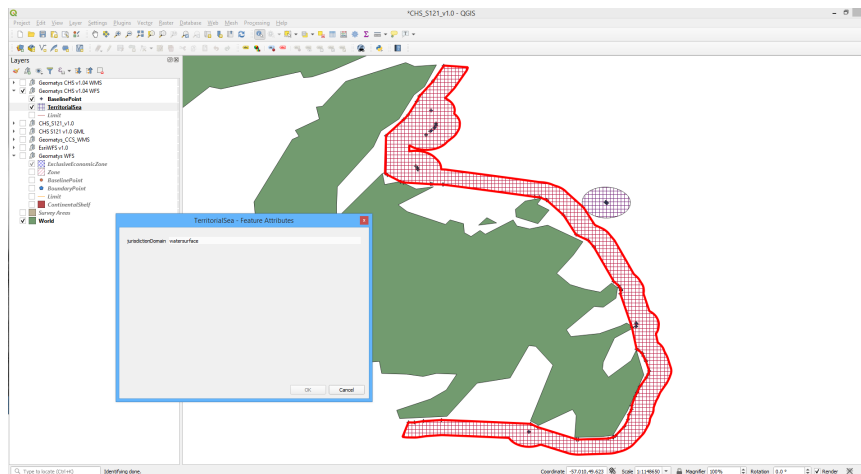

Figure 85. QGIS Geomatys WFS Attribution [Note]

8.5. Summary

Serving a WFS is the most useful way for a Hydrographic Office (HO) to manage and share their MLB data as the HO will be in complete control of the data and attributes made available. There would be a case for creating a flattened S-121 schema so the full information and relationships can be made available through simple features based WFS such as ESRI's implementation (web services such as WFS have the additional advantage that the HO will be able to ensure it is always kept up-to-date with any changes as they will be in control of it.)

8.6. CSW - Catalogue Service for the Web Testing.

CSW is a Catalogue Service for the Web. This allows a user to search online for Services and Datasets that organizations have published within a particular domain. It is a metadata search facility which predominantly uses structured, predefined metadata standards and relies on datasets being attributed correctly with structured metadata conforming to those standards.

In QGIS CSW can be accessed using the MetaSearch Plugin (this needs to be activated and then available under 'Web' menu item).

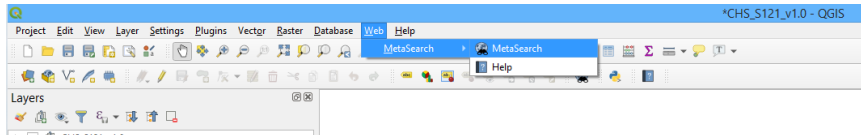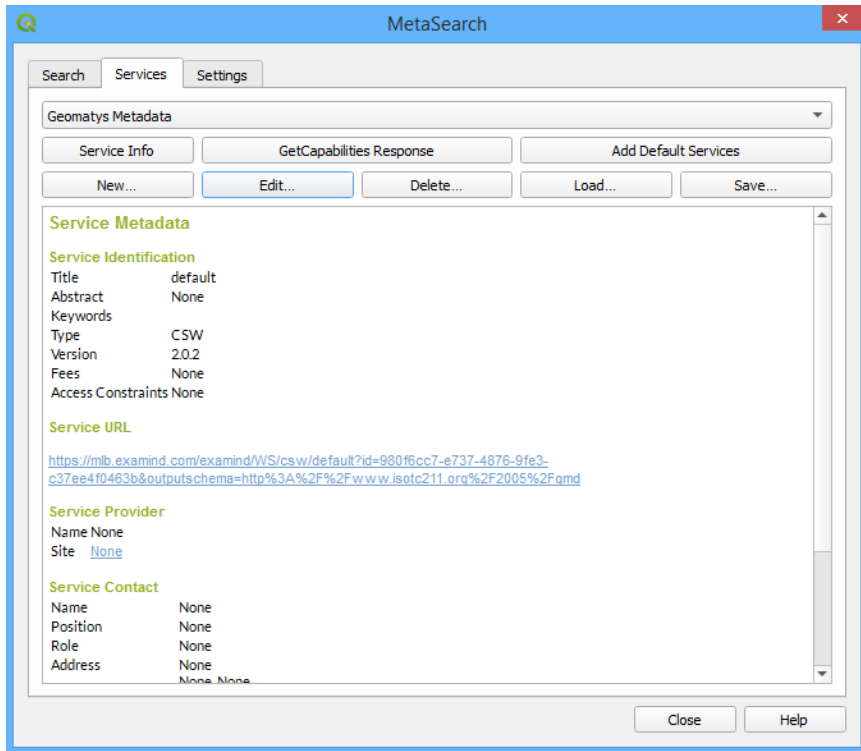*Figure 86. QGIS Access to metasearch plugin*

Figure 87. QGIS Metasearch configuration

This allows searching by Keyword and Bounding Box. However, it only returns Dublin Core metadata records, which then technically do not comply with the requirements of Phase 2 of the MLB project which requires HNAP or ISO. This is a limitation of the QGIS tool, however, not the data.
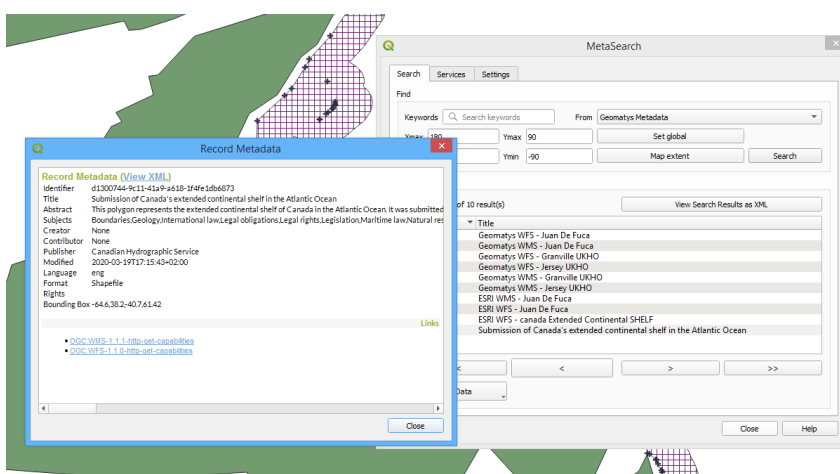
Figure 88. QGIS CSW Record Metadata
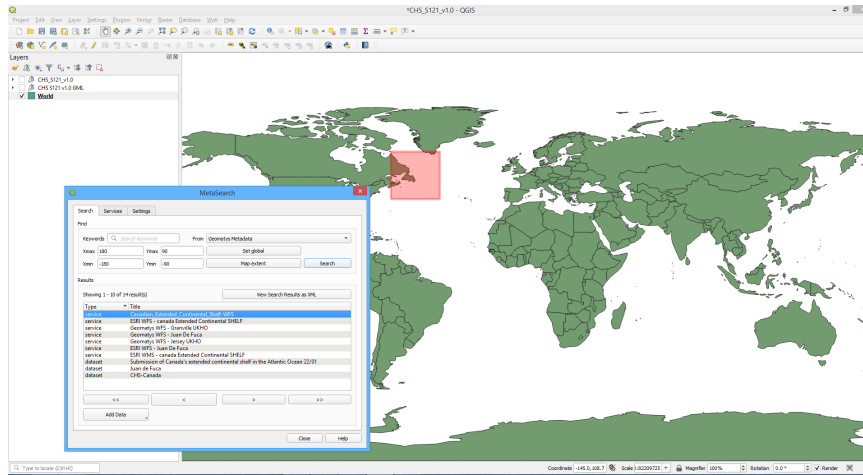
Figure 89. QGIS CSW Bounding Boxes

8.6.1. Geomatys CSW

During the pilot, Geomatys also published a CSW service containing records for all web services available through Phase 2 of the project. This includes both Geomatys and Esri services. The services can be filtered using keywords for a specific feature type, area or type of service. Wildcards can be used in the filtering so searching **Continent** will return records containing "Continental."

Figure 90. QGIS Metasearch CSW Options

Figure 91. QGIS Metasearch Search Example
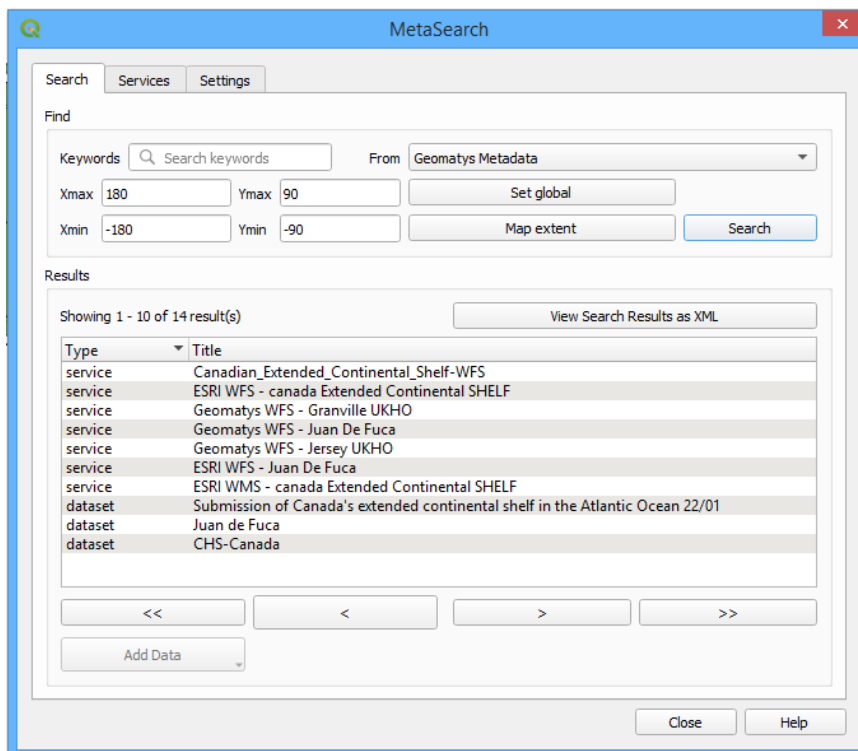
Figure 92. QGIS Metasearch CSW Results
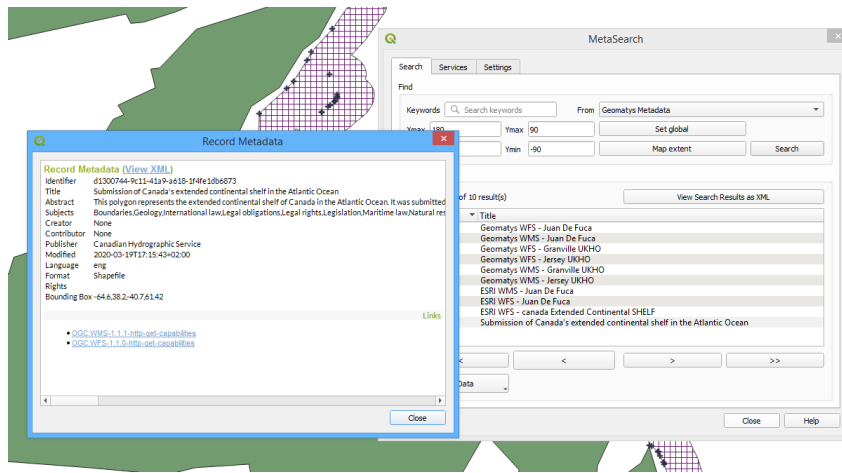
Figure 93. QGIS CSW Record Metadata
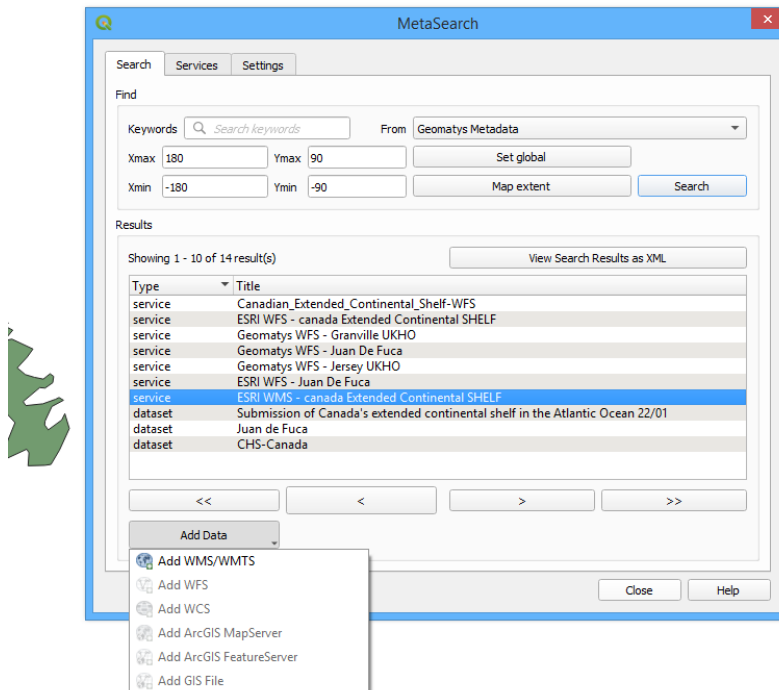
Figure 94. QGIS CSW Add WMS data from search to project
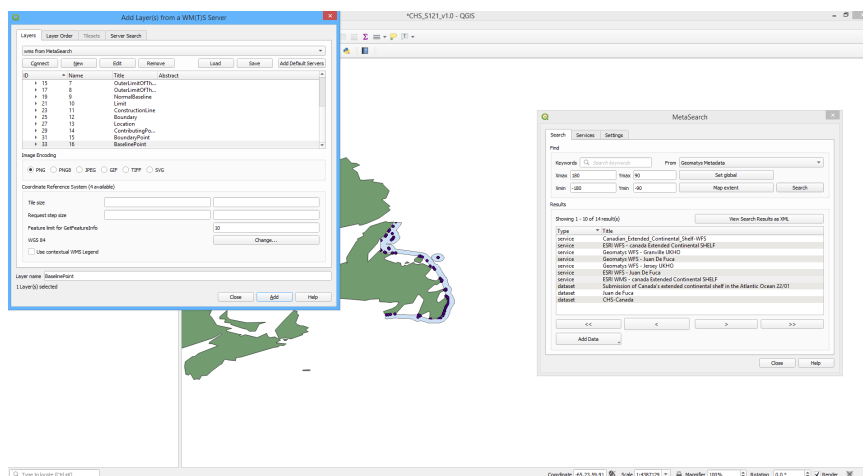
Figure 95. QGIS Add ESRI WMS from CSW query
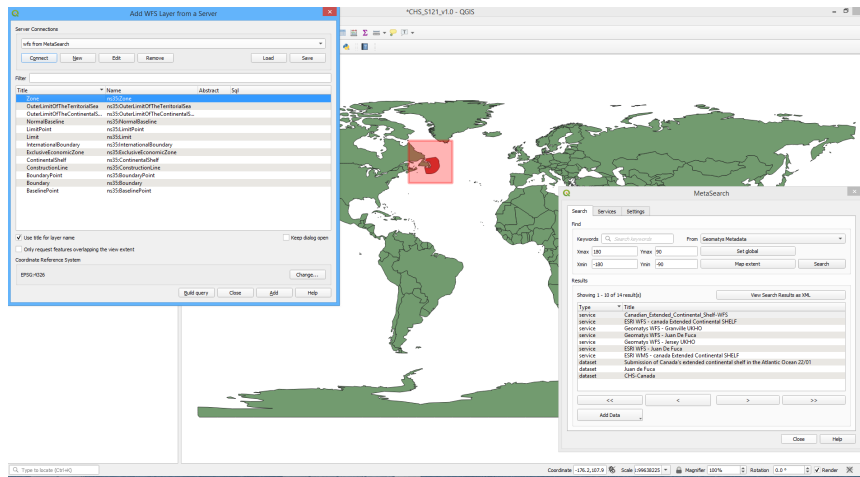
Figure 96. QGIS Add WFS from CSW

8.6.2. FME Support for CSW Processing

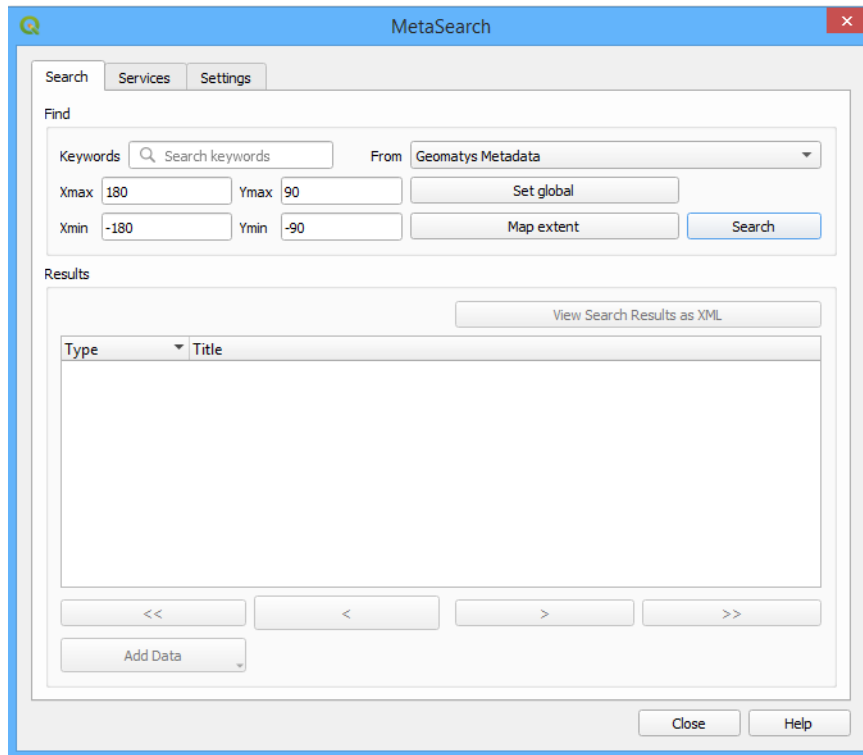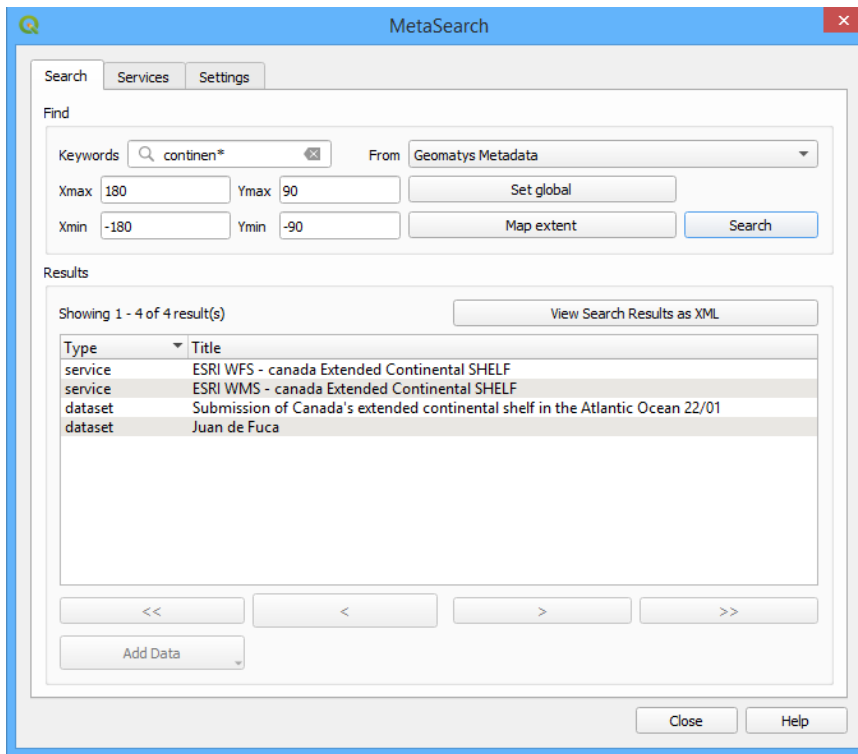FME does not have a specific CSW reader per se. Nevertheless, for well documented web services such as any OGC service, it is relatively easy to create a client using an FME Workspace that invokes service calls based on templates associated with the service (CSW in this case) and passes these to a series of HTTPCallers. The results of this can in turn can be used to drive down stream processing such as automated data retrieval.

The FME client workspace CSWclient_MLBpilot_v1.0.fmw was created to demonstrate the functionality of CSW processing.

The basic process steps for CSWclient_MLBpilot_v1.0.fmw are as follows:

1. Prompt for query term or feature ID;

2. Compose CSW Query;

3. Query CSW for all available records via HTTPCaller;

4. Parse CSW response XML into CSW records;

5. Apply query filter to CSW records;

6. Retrieve CSW details for selected records;

7. Filter CSW details by format;

8. Feed CSW detail records to FeatureReader for selected format;

9. Retrieve dataset using webservice information from CSW.URI.
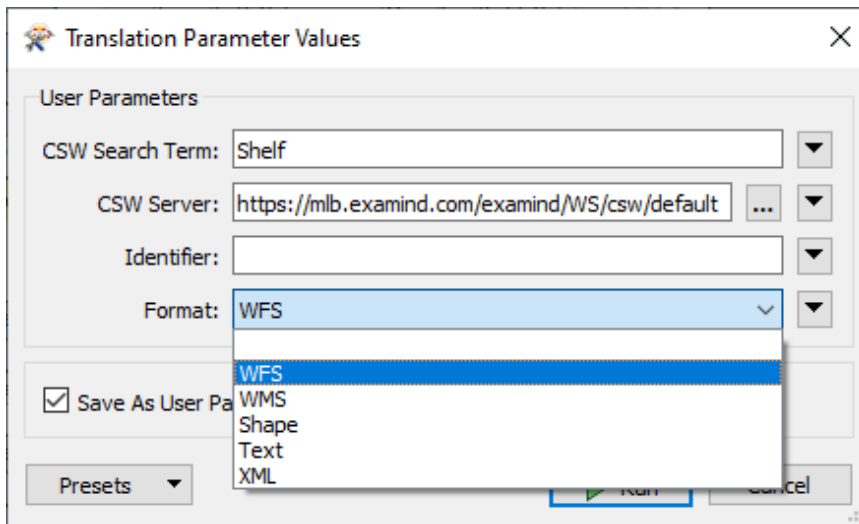
Figure 97. FME CSW Client workspace runtime parameters showing query and format selection
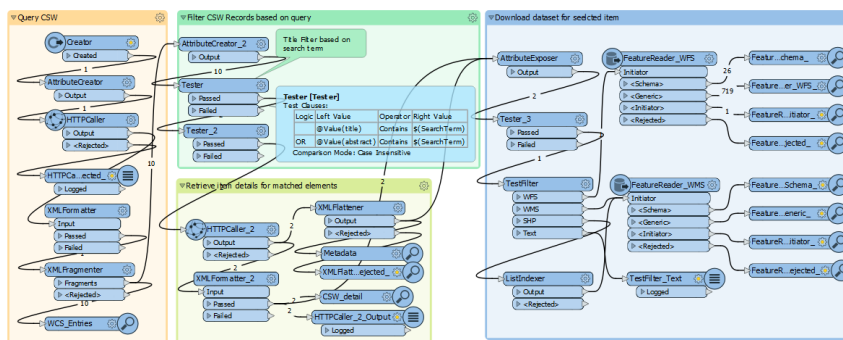
Figure 98. FME CSW Client workspace: CSWclient_MLBpilot_v1.0.fmw
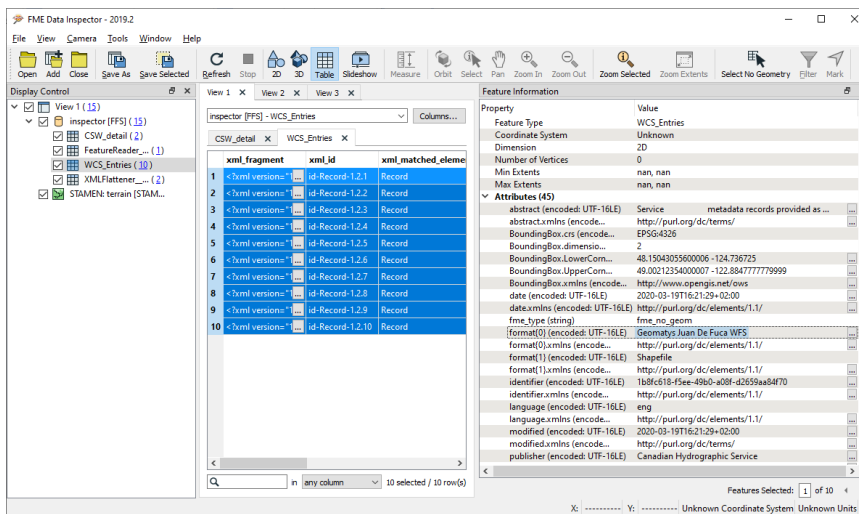
Figure 99. Available CSW entries
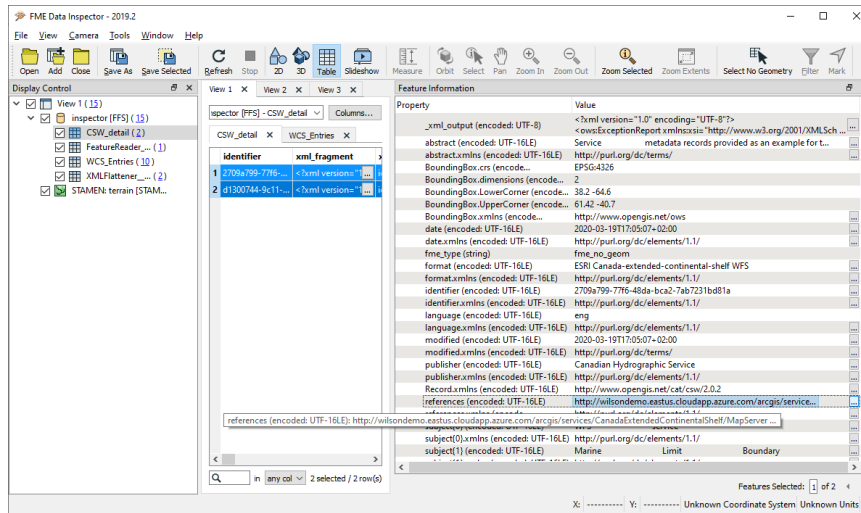
Figure 100. CSW details for record query of interest: 'Shelf'
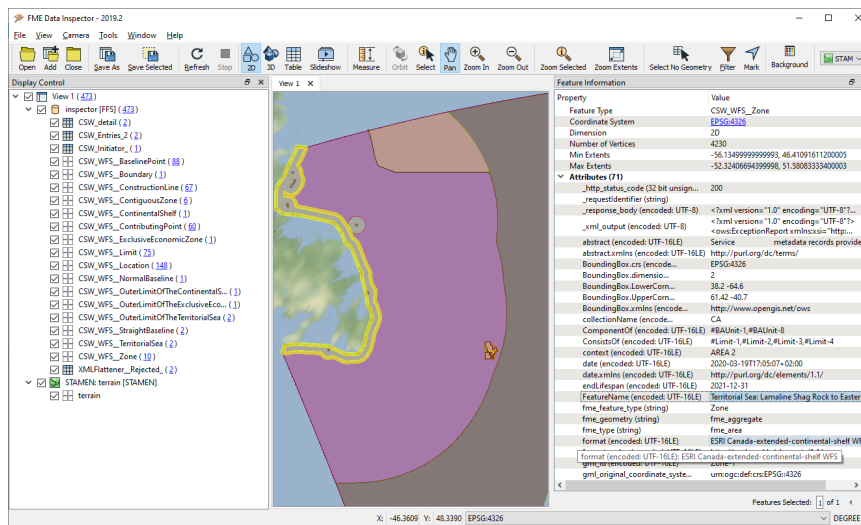
Figure 101. Automated WFS retrieval for CSW query result for query='Shelf' and format = WFS [Note]
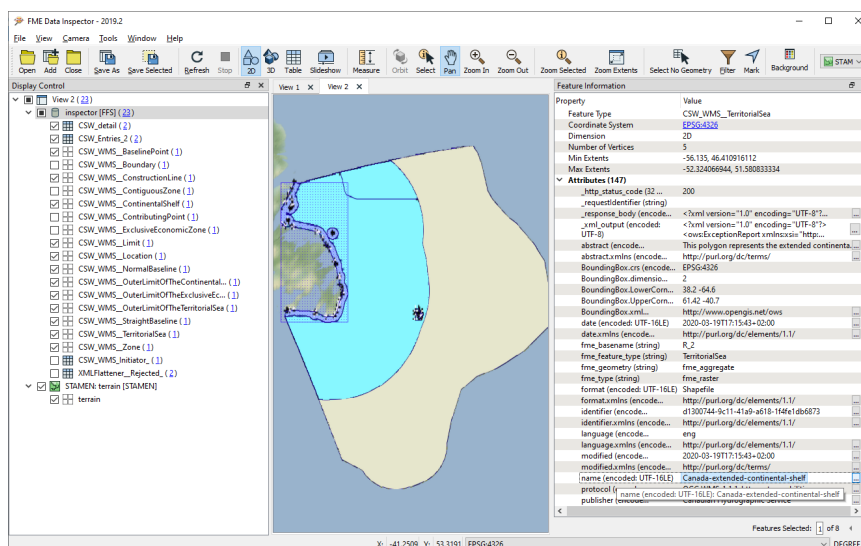
Figure 102. Automated WMS retrieval for CSW query result for query='Shelf' and format = WMS [Note]

The FME CSW client using the workspace CSWclient_MLBpilot_v1.0.fmw was able to interact with and retrieve CSW records and associated datasets. Further, it was relatively easy to extend the client and add additional search terms and service or format types for retrieval.

Given the basic nature of the user interactions with the service, this type of CSW client would be

more conducive to CSW integrations of a more automated nature. For example, a scheduled FME Server process could query a CSW daily and only download datasets that meet certain criteria, such as recent updates, or the presence of certain metadata keywords. It would also be worth exploring the options for formalizing FME's support for OGC CSW via FMEHub (https://hub.safe.com/) and eventually as a new FME reader format.

Chapter 9. Project Outputs

9.1. Input to IHO S-100 Working Group

The outputs recommended for the S-100 Working Group are listed below: these have been identified within the project activities as being of interest to the S-100 Working Group and represent areas where the participants felt there may be room for improvement in the IHO architectural and standards baseline. These are over and above any outstanding "issues" remaining in the schema which are detailed in Outstanding Issues and Ongoing Work Items with the GML Schema.

- An Improvement of guidance on mapping feature catalogue elements to GML Application Schema elements. This could be more explicit than in the current guidance given in the standard even to the point of mapping specific S-100 types to their GML equivalents. The aim should be to make the production of a GML Schema as automated as possible leaving little open to interpretation. If there is a need for more sophistication, then this could be done by adopting a finer grained approach to the levels of "conformance" with the profile.

- More concrete guidance on how geometry should be approached for modeling within GML Application Schemas, in particular, simple geometry types such as point, curve and surface. There would seem to be a case for adding some kind of GML Topology structures into the S-100 GML Profile. Although vendor support for these is likely to be patchy, it would at least allow a thorough investigation to ascertain how suitable they are for modeling of S-121 data without possible ambiguity. The issue of topology is a sensitive one as it is a sophisticated notion. Referenced geometry enables some of the benefits of topology such as common points for e.g., Baseline Points and Boundary points but elements like oriented curves defined by reference to points are still difficult to implement and only a fuller topology model would satisfy the requirement entirely.

- More specific guidance on the creation of top level dataset "Collection" elements, their types, and overall structure should be considered and made explicit. The current standard is not normative in this respect and it should be considered whether a generic S-100 "Dataset" element should be introduced eliminating the ambiguity. This also has benefits in the context of the "extension" Schema described in this document, giving it a location and normative definition. The project also settled on a dataset structure which is unordered and uses inheritance rather than subgroups to specify members. This, although unconventional, streamlines the GML Structure. S-100 should adopt a structure which allows for maximum interoperability with existing clients, while retaining the richness of schemas necessary to model specific domains, and also enables extensions such as those defined in this document.

- Dataset ordering should be defined as part of the structure even if external validation is required to enforce it. The observation of the project is that relationships between features can take place between information types and therefore any ordering of feature groups would need some external validation to ensure the "define before use" rule is kept to. The other option is to abandon such constraints abandon such constraints. It is worth noting that by convention, most XML and GML schemas do not implement define before use, and doing so can lead to problems with implementation. This debate needs to be had within the relevant communities and noted in the revised Part10b.

- Enumeration values or indexes should be optional but with a dataset metadata attribute in place to show which are being included. This is alluded to in the standard but not realized.

- Multilingual facilities in most S-100 product specifications are limited to feature naming. It may be desirable to include other multi-lingual attributes such as descriptions and the S-100 method should be reviewed to establish if it is fit for purpose in its current state. The alternative is ISO style locales or some other (tbd) mechanism. It should be recognized that S-100 is likely to have a broad reach and the requirement for language specific implementations will be a common one - but that it will extend to other categories of attributes as well as "names" (e.g., descriptions, categories etc.). This is also relevant in the context of the enumeration values noted in the previous point.

- The IHO DSID and DSSI GML Profile elements do not currently derive from GML types (and probably should) and their content should probably be brought up to date with the latest edition's approach to metadata. The current elements are possibly obsolete in some ways. They should reflect the metadata methodology in the rest of the standard and, likely, implement some form of extensibility for future proofing them. The DSSI feature, point and edge totals are probably not of any use to implementing systems as they are non-standard in the context of GML.

- GML 3.3 is included in the default GML Profile headers. This most likely isn't required and affects interoperability with some clients. Removing GML 3.3 from the namespace header did not result in any observable loss of functionality and did allow more GML clients to consume the data.

- The generic extension mechanism, although simple by nature, potentially provides a way of IHO S-100 GML being augmented with GML from other sources, separately modeled and distributed as an integrated whole. S-100 currently allows the integration of multiple IHO product specifications as a combined exchange set but not the integration of multiple features from different domains within a single dataset. The extensions suggest the breakdown of an S-100 GML "schema" into component parts and separating them from the aggregation/metadata components. Establishing this in S-100 would be a big step forward. Tighter GML encoding might be required by ECDIS implementation perhaps (to reduce effort for S-10X implementation on the client systems), so there are other benefits as well. A more modular approach would also allow for sharing of elements such as types in extended schemas. The simple examples in the project have many benefits in the MSDI context and these mechanisms should be promoted in that community to identify potential other uses.

- Feature Object Identifiers (FOID) are still included in the standard GML features. These would seem to be obsolete. GML includes unique identifiers for all features so the FOID does not need to be a part of the generic feature. If MRN is included, making the identifier an attribute of the feature then the inclusion of FOID as an embedded identifier seems to make little sense. It potentially gives a feature three separate identifiers, MRNs, FOIDs, and GML:ID.

- An ability to override and add to generic features within the profile would make sense so that it is possible for a product specification to enhance metadata content as necessary.

- Little guidance exists for those data producers who wish to adapt an existing geospatial relational database to a GML Application Schema. The transformation process prototyped in the Pilot could be abstracted to form such guidance which would be useful to data producers wishing to follow the processes outlined in this ER.

9.2. Input to IHO S-121 Project Team

The project has enjoyed a substantial cross-representation from the OGC S-121 Project team. The outputs most relevant to that team are listed in this section.

The main deliverable for the IHO product specification is the GML Schema: the existence of a GML Schema conforming to S-100 Part10b and suitable for incorporation in the product specification is now a reality. The project team has expressed a willingness to embed the GML Schema within the standard as an 'annex' - conforming to the S-100 product specification format. The GML Schema would require reformatting and some elements to be assessed prior to its incorporation but it certainly forms a best-view of how S-121's conceptual model can be solidified within a concrete GML Schema for use in exchange between implementers.

Elements requiring consideration and review prior to their incorporation in S-121 would be as follows.

- MRNs. Much discussion is ongoing within the IHO S-121 project team surrounding MRNs. Within the Pilot their existence was acknowledged and added to the schema. However, it is acknowledged that there are broader issues with MRNs than those dealt with during the project lifespan. However, the inclusion of MRNs is an easy addition in the context of GML Schema (although the question of how to implement via a validated GML type is still open).

- Metadata - considerable discussion around the HNAP metadata profile generated a valuable side by side comparison with the ISO profile and an analysis of the S-100 provisions under Part 4a. This should be taken further and used to refine the metadata provision in the product specification based on the experiences of the participants. The metadata section of this ER contains more specific recommendations. S-100's metadata is about to go through a more general review and it would be advisable to keep abreast of that. Ultimately a form of pluggable metadata may be required in order to ensure the individual needs of S-121 stakeholders are preserved. This should then be reflected in the GML profile of S-100 as appropriate.

- Encoding guidance for attribution within the product specification is currently minimal. This makes representative datasets difficult to define as feature and metadata fields (particularly some of those relating to source and LADM information types) are unclear as to their meaning in context of this domain. This is ongoing - the correct location for such content in S-100 product specifications is within a Data Classification and Encoding Guide (DCEG). The more comprehensive and informative such a guide is the more interoperable and consistent data will be between different data producers.

- Changes to model - locationReference has been improved, and some changes to relationship names have been made. There is also a draft UML version of the FC/Schema which summarizes the implementation of the core model in these two encoding schemas (Feature catalogue and GML Schema).

- A thorough alignment of the feature catalogue against the model and schema should be carried out as the schema is baselined to ensure no inconsistencies are present. Some changes to feature catalogue have been drafted and a revised feature catalogue created by Teledyne Caris which will form a valuable first draft of such an update. Ideally the inclusion of the GML Schema should take place alongside the feature catalogue refresh to ensure they are consistent with each other so should the GML Schema output from the Pilot change, the feature catalogue should also change.

- More attention is required to determine if comprehensive validation checks required as part of the geometry validation because of the lack of support for topology (and hence Level3a geometry within the GML profile of S-100).

- Validation tests should be defined and developed in S-121 to ensure they are stated at a level where GML Specific versions can be defined. This would enable production of better quality S-121 data.

- Much discussion was held around the benefits of "simplified" profiles of the S-121 standard, flattened versions of schemas, and how to simplify without obscuring the underlying meaning of the features themselves. This should be investigated further. Certainly the experience within the project is that simplified versions of the schema would broaden the support for both COTS and Open Source implementations. The addition of Joins to, e.g., QGIS output, and visualization is significant in that it could make an open source visualization of the S-121 data much more comprehensible than the default rendered by the GMLAS implementation showing features in customized form. Coupled with editing capabilities this would significantly increase usability of the system in a completely open source fashion.

- The issue of data integrity has not been dealt with by the S-121 product specification. S-100 includes Part 15, which is a general approach to data authenticity which has a strong link with provenance and authorship of data, a major part of the S-121 picture. It is possible to implement S-100 Part 15 for S-121 but it requires explicit statements in S-121 to that effect. Current developments within the S-100WG around data integrity at a feature level rather than a dataset level may also have some added benefits to S-121 and should be tracked.

- It should be noted that the GML Profile under S-100 does not support full topology and therefore, although geometry by reference is included in the GML Application Schema, elements like named start/end points on oriented curves cannot be expressed using the geometry elements defined under the profile (COTS and open source support for these elements are implemented to varying degrees as well). The compromise, therefore, made by the S-121 GML Application Scheme is to allow shared geometry via reference but at a cost of requiring validation tests to ensure the topological consistency of GML datasets. This should also be taken up with the S-100WG.

9.3. Other

Outputs to GDAL development team in respect of identification of geometry fields in complex GML and resolution of included/imported schema XSDs (used for extension schemas) should be progressed through open source channels.

9.4. Alternatives to GML

An important (albeit hypothetical question) is whether alternative encodings could be used to achieve the results of the OGC MLB project. The use of GML has been ubiquitous through the project and some of the work carried out has relied on GML-specific constructs, such as schemas, identifiers, geometry, formatting, etc. Today, many geospatial professionals use encodings other than GML to represent data in web services, most notably GeoJSON. S-100 itself has another defined encoding, ISO 8211, a binary format used for compact, self-describing encapsulation of vector data for electronic charts. The use of ISO 8211 was noted in discussions between participants at the end

of the project as a possibility for an alternative encoding which is unambiguous, compact, well-tested, and is the only one to fully support the Level 3a geometry required by the S-121 product specification.

Additionally, the OGC development trajectory is very much towards the establishment of Restful OGC APIs, which are completely content neutral. These next-generation web services, spurred on by the exponential rise in the popularity of web mapping services, will therefore use any of a number of actual encodings, the most popular being GeoJSON. So, it is worth considering whether the project activities could be replicated using other encodings, particularly JSON and GeoJSON.

In terms of supporting alternative encodings the following are considerations and essential steps to consider.

- Schema design - a "schema" provides a normative syntax for the XML describing the features, their attributes, and the relationships between them. An alternative encoding, such as GeoJSON would need to support all these entities as well but the real power of using XML/GML is the rich schema description language which allows the tight specification of the syntax of the implementing grammar (XML). GeoJSON's schema evolution is not as rich currently but is advancing quickly.

- Validation - again, XML/GML has a powerfully defined concept of validation which would need to be tightly defined for other encodings.

- Support from the overarching framework. S-100 contains a GML profile which restricts the individual subsets of the GML language used by its product specifications so a simpler vocabulary is used when defining S-121 GML data. No such S-100 framework for other encodings other than ISO 8211 currently exist.

In theory a GeoJSON encoding would be possible to support. The practicalities are where much of the detail is though, the shortcomings of the schema/validation, and lack of extensive tool support would be a challenge for an implementer but, arguably, native GeoJSON could be implemented in much the same way as the MLB project has implemented GML and follow similar evolutionary paths. Much of the dialogue and evolution of the GML Schema is not primarily because of its nature as a GML Schema, but revolves around aspects of the underlying standard itself.

The core element to consider is the S-100 feature catalogue, the XML file which describes the product specification's feature/attributes and relationships. This is a central artefact and, in theory, should define the greater part of the GML Application Schema (the MLB Schema was designed using the feature catalogue as a template). Taking this process to its logical conclusion it is an interesting question to ask if the GML Application Schema could be auto-generated from the feature catalogue without any loss of expressivity (this is one of the outputs being proposed to the S-100WG). If so, then a similar argument applies to other encodings like GeoJSON, defining a template for that encoding and validation to be used as well. The marine GeoPackage initiative is trying to define similar templates within the GeoPackage encoding.

9.4.1. GeoPackage

The other major encoding worthy of mention is GeoPackage. GeoPackage provides a single integrated container for data in a relational form with multiple tables, geometry properties and supports many desirable features such as embedded portrayal, vector tiling and geospatial

indexing.

The potential of GeoPackage is enormous in terms of file-based embedding of data and OGC has an existing initiative within the marine domain, Marine GeoPackage. An example of QGIS import of a GeoPackage enabled "Maritime Jurisdiction" layer is shown in the following image.
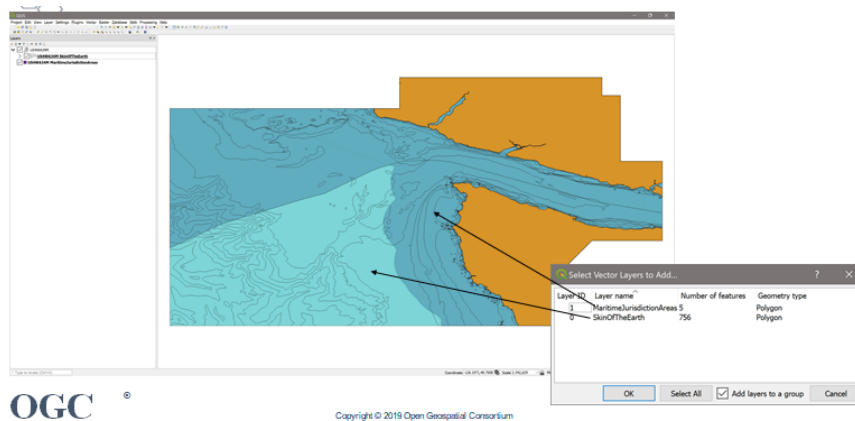

Figure 103. Marine GeoPackage

9.4.2. Creation of GeoJSON/GeoPackage with FME

The most comprehensive translation conducted in Phase 2 was from the CHS extended continental shelf geodatabase to S-121 GML using the baselined S-121 GML application schema. This GML output dataset was then used to feed a range of OGC web services. However, given the complexity of WFS and GML, there are potentially many users interested in encodings with easier access such as GeoJSON and GeoPackage. These could also be transmitted via REST services and used by a wider range of clients, web applications mobile devices.

Currently within the OGC community, there is considerable interest in the development of the OpenAPI set of web services, which includes definitions for REST feature services such as OGC API - Features. Two of the encodings often used for these services are GeoJSON and GeoPackage. Therefore, as an optional additional exercise, FME was used to generate S-121 in alternate encodings: GeoJSON and GeoPackage.

Because FME is a centralized semantic translation engine, changing from one encoding to another requires minimal reconfiguration. The main changes are typically only what is required to address destination format structural and syntax differences. For example, because JSON and XML are structurally similar – both are object oriented, nested structures – the main changes relate to the syntax differences for how XML and JSON encode complex elements like arrays and other nested attributes.

9.4.2.1. GeoJSON

Objectives

1. Generate S-121 GeoJSON that captures all the relevant content available from the sponsor database for the selected test area – in this case Area 2 off the east coast of Newfoundland and Labrador.

2. As far as possible, the S121 GeoJSON output should reflect a similar structure to the equivalent S-121 GML, since both are based on the same S-121 UML design and Feature Catalog. Note that

this data structure should be most closely reflected in the feature type and attribute schema, but not necessarily in the geometry schema since by its nature GML has complex constructs that are not required by the alternate encodings tested.

3. The S-121 GeoPackage tested was more closely structured after the source CHS geodatabase and not S-121 UML or GML since as a database it is relational and not object oriented in nature.

Inputs

- CHS geodatabase: Canadian Extended Continental Shelf
- S-121_CanadianExample_GB_NF_20200319.gdb

The input was the same as that used by CHS_Geodb_to_S121gml_v1.1.fmw described above.

Process Steps

FME uses a model oriented spatial ETL transformation process to convert data from one format and structure to another. The source database and destination GeoJSON data models are schema driven so the work is largely oriented around transforming and mapping the data structures from the source to match the schema and specification requirements of the destination.

The process steps for Geodatabase to GeoJSON is very similar to the Geodatabase to GML workflow described above. The modifications required were minimal since all that was needed was the addition of a GeoJSON writer and some minor modifications to the output schema configuration. The FME GeoJSON writer handles lists and complex attributes somewhat differently than the GML writer. The following summarizes the changes made to CHS_Geodb_to_S121gml_v1.1.fmw to produce the GeoJSON encoding workspace CHS_Geodb_to_S121geojson_v1.1.fmw.

Changes made to reconfigure CHS_Geodb_to_S121gml_v1.1.fmw for S-121 GeoJSON generation:

1. Added GeoJSON writer for JSON encoding test;
2. Added JSON object fields to handle complex attributes such as featureName, LocationReference, componentOf, consistsOf;
3. Created S121JSONConverter custom transformer to build JSON objects using JSON templates; and
4. Added logic to test list concatenation as an option for list encoding in JSON for some one to many relations such as source.xlink_href and relatesTo.xml_href.
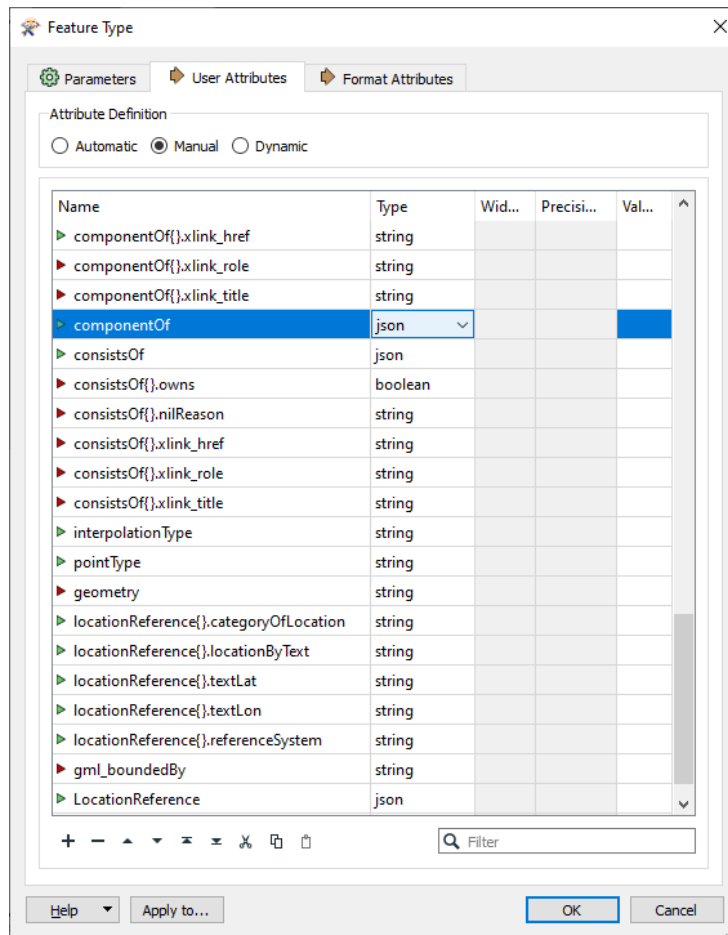
Figure 104. GeoJSON destination feature type modification:

This image shows the addition of json datatype parent fields to support json object writing for complex attributes such as componentOf, consistsOf, featureName, and LocationReference.
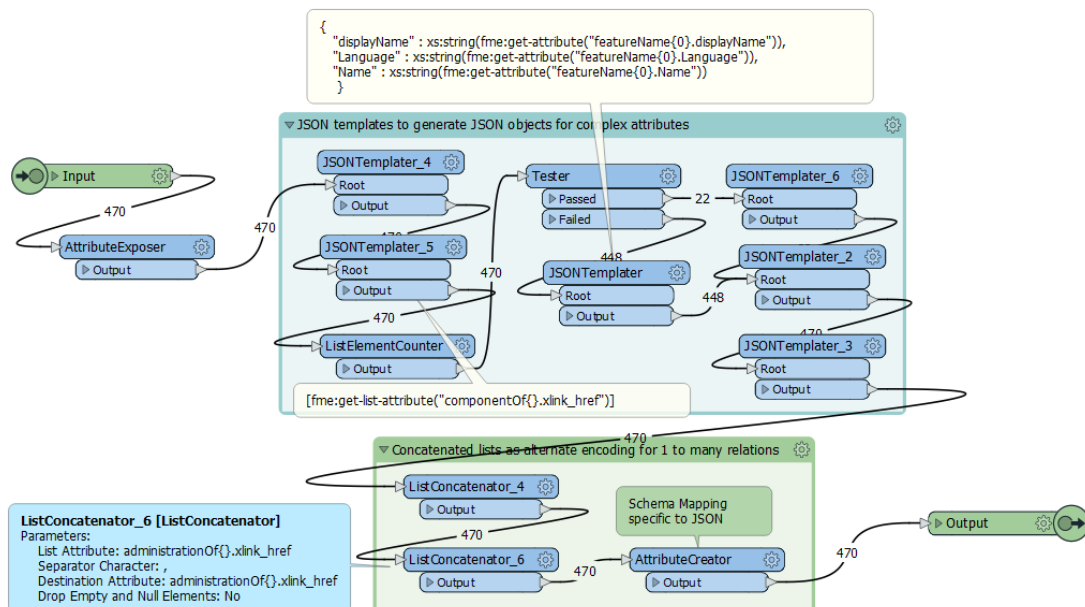
Figure 105. S121JSONConverter custom transformer definition to support conversion from geodatabase to S121 GeoJSON in FME workspace CHS_Geodb_to_S121geojson_v1.1.fmw. Note JSONTemplater transformers and the JSON templates to construct complex attribute objects for featureName, LocationReference etc.

Once the GeoJSON writer was added and the feature types reassigned to it, the basic approach for adjusting to GeoJSON encoding involved modifying the destination feature types as shown in figure

1. There is an important difference between FME's GML writer and the GeoJSON writer. The GML Writer can support the complete representation of complex attributes in the destination feature type schema. The GeoJSON Reader can represent complex schema using parent.child and list notation, but at the moment this cannot be used by the Writer to generate complex attributes.

Fortunately, there is a workaround that involves adding a parent level node attribute with datatype = 'json' that represents the top level of the complex attribute. For example, while FME can read featureName{0}.Name, and featureName{0}.Language, in order to write this we need to create a GeoJSON object called featureName that contains an array of Name and Language. LocationReference, componentOf, consistsOf, and featureName are all examples of the new top level parent attributes that were created to support complex attributes and arrays.

Once the schema was modified as above, the next step was to create the JSON objects to populate the complex attributes. This was accomplished in the new custom transformer: S121JSONconverter. For example, the following JSON template was used to generate the LocationReference object:

```
{
"categoryOfLocation" : xs:string(fme:get-attribute("locationReference{0}.categoryOfLocation")),
"locationByText" : xs:string(fme:get-attribute("locationReference{0}.locationByText")),
"textLat" : xs:string(fme:get-attribute("locationReference{0}.textLat")),
"textLon" : xs:string(fme:get-attribute("locationReference{0}.textLon")),
"referenceSystem" : xs:string(fme:get-attribute("locationReference{0}.referenceSystem"))
}
```

The fme:get-attribute() functions merge the feature properties into the JSON template.

Outputs

- GML English - CHS S-121 Canadian Sample GML - CHS_S121_v1.2_en.gml, and GML French - CHS_S121_v1.2_fr.gml.

- GeoPackage results CHS S-121 Canadian Sample GeoPackage.

- GeoJSON and FME workspaces can be accessed here: https://knowledge.safe.com/articles/112578/ogc-maritime-limits-and-boundaries-s-121-gml-pilot.html.

- Sample geojson outputs posted are: CHS_S121geojson_v1.12_fr_demo.json and CHS_S121geojson_v1.1_Zone.json.

- The script used to perform the translation is called an FME workspace.
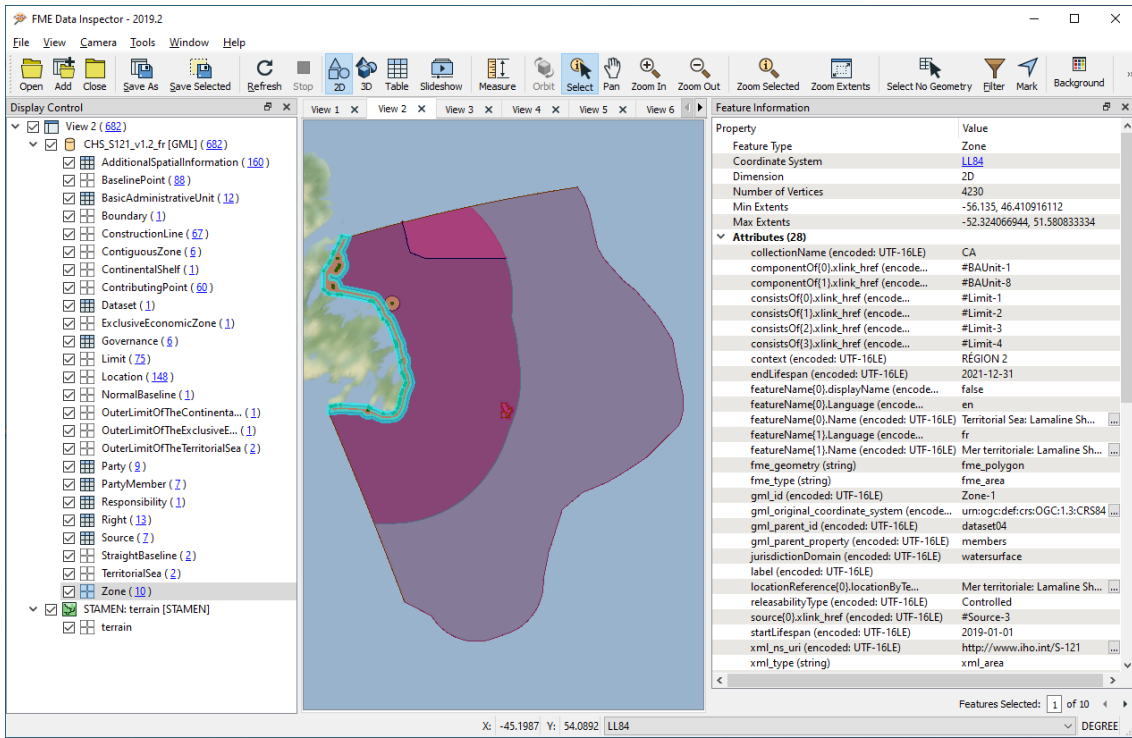
Figure 106. CHS S121.gml: Zone 1 Territorial Sea for comparison [Note]

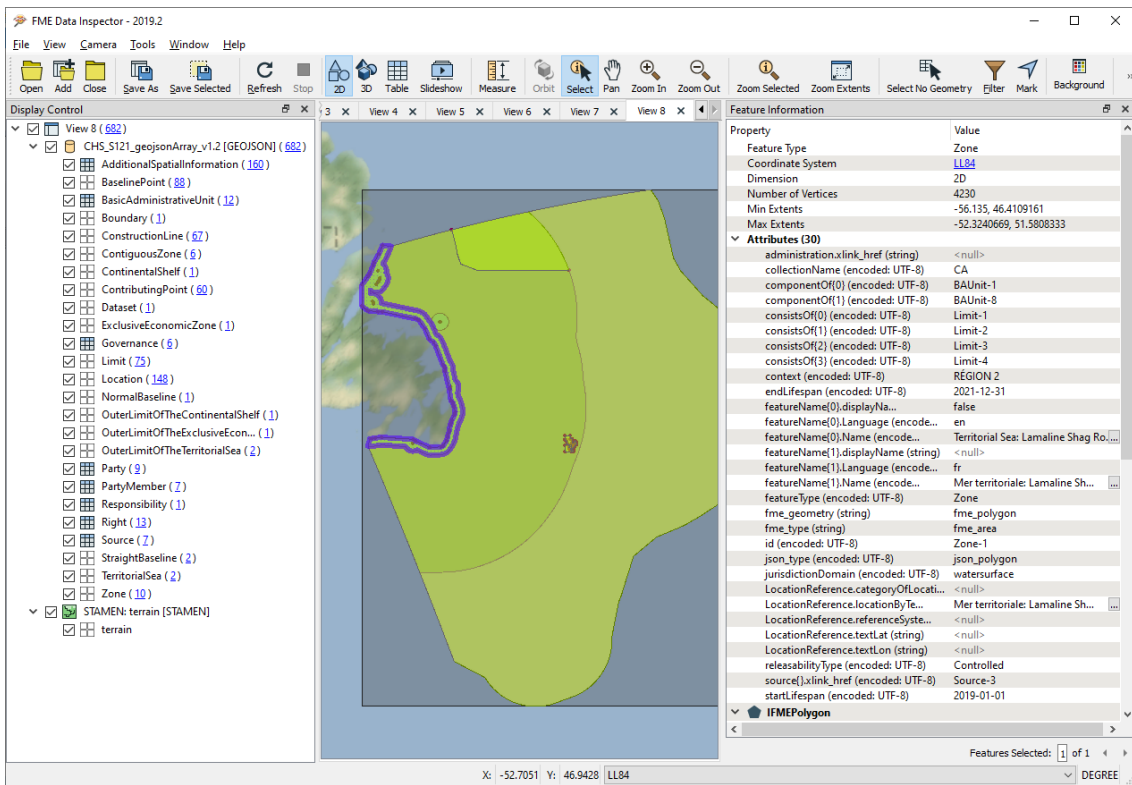*Figure 107. CHS S121.GeoJSON: Zone 1 Territorial Sea* [Note]
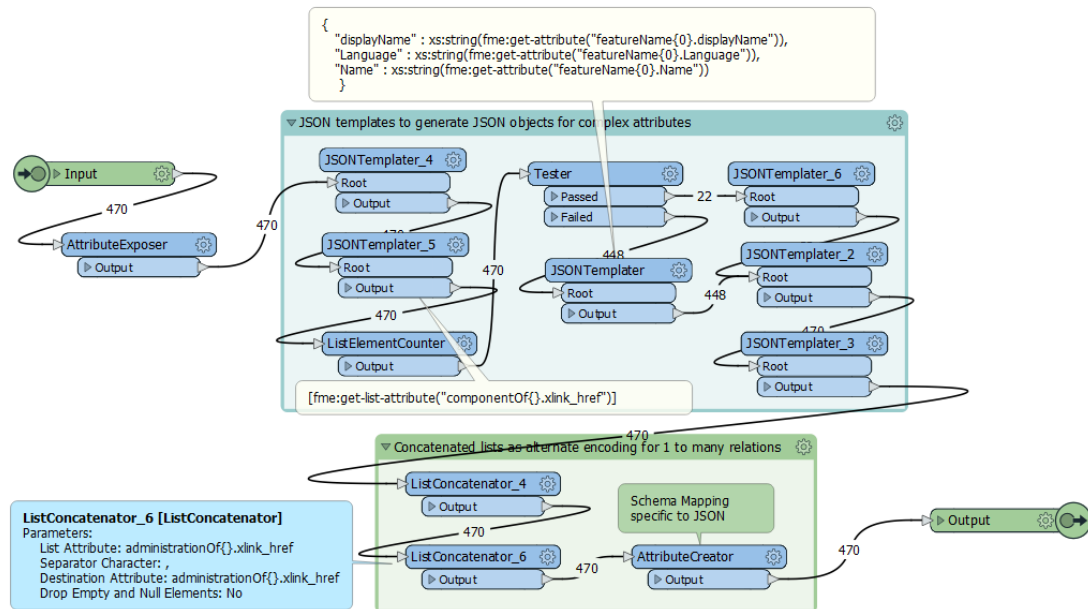
Figure 108. GeoJSON S-121 Zone feature type displayed by default renderer on MLB GitHub
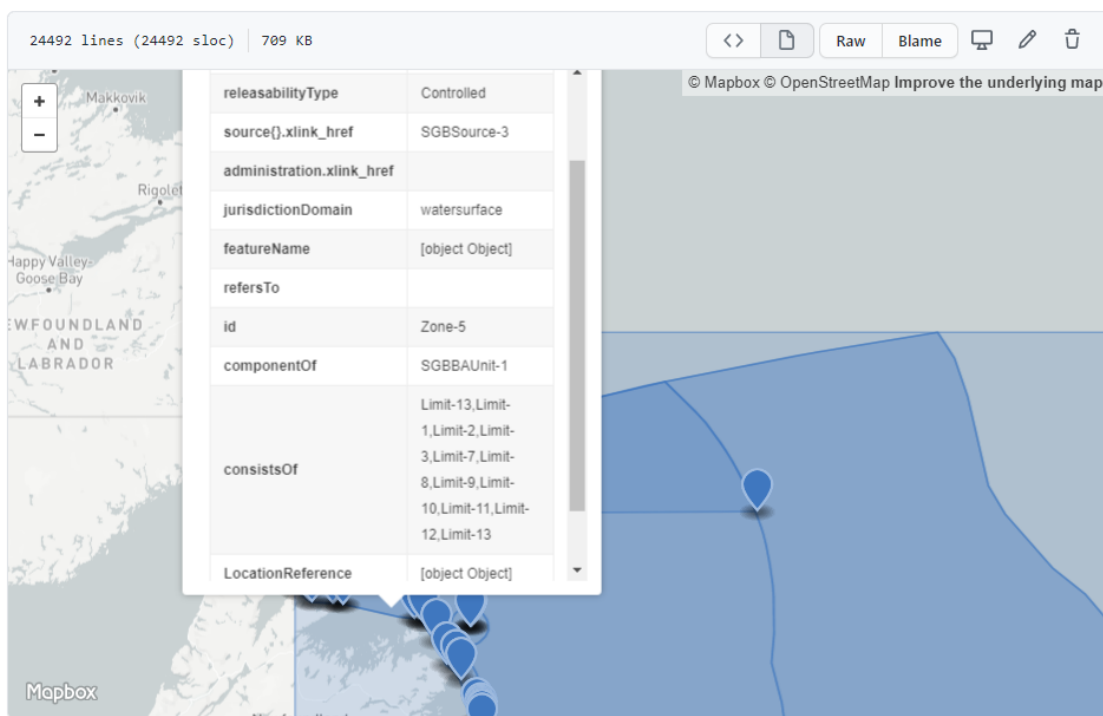
Figure 109. GeoJSON S-121 Zone feature type displayed with attribute schema on the OGC MLB GitHub. Note that relationship arrays are rendered, but not complex attributes like LocationReference or featureName. [Note]

Example GeoJSON Encoding of Location-1

```
{"type" : "FeatureCollection",
 "name" : "Location",
 "features" : [
 {   "type" : "Feature",
 "geometry" : {
 "type" : "Point",
 "coordinates" : [ -55.8239444, 46.8391389 ]},
 "properties" : {
 "id" : "Location-1",
 "collectionName" : "CA",
 "featureName" : {
 "displayName" : "false",
 "Language" : "en",
 "Name" : "Lamaline Shag Rock"},
 "context" : "RÉGION 2",
 "startLifespan" : "2019-11-26",
 "endLifespan" : "2021-12-31",
 "releasabilityType" : "Controlled",
 "source{}.xlink_href" : "Source-1",
 "administration.xlink_href" : "#BAUnit-5",
 "refersTo" : [ "Reference-1", "Reference-72" ],
 "componentOf" : [ "Limit-1" ],
 "consistsOf" : [],
 "interpolationType" : "densification",
 "pointType" : "defined",
 "LocationReference" : {
 "categoryOfLocation" : "OriginalLocation",
 "locationByText" : "Lamaline Shag Rock",
 "textLat" : "46°50□21□N",
 "textLon" : "55°49□30□O",
 "referenceSystem" : "4016"
 }
 }
 }
```

The example shows an example encoding of the Location: id = "Location-1." Note Note the 2 options for encoding one to many relationships: the simple string with concatenated list used for "source{}.xlink_href" : "Source-1," and the JSON objects used for complex attributes "featureName," "LocationReference," and relation arrays used in "relatesTo" and "componentOf."

9.4.2.2. GeoPackage

Conversion from CHS geodatabase to GeoPackage was accomplished using the FME transform workspace: CHS-S121-0205-geodatabase_file2geopackage.fmw. This is essentially a default translation from ESRI Geodatabase to OGC GeoPackage. Note that FME supports 2 different geodatabase readers: ESRI Geodatabase and the open source File Geodb Open API. The licensed ESRI Geodatabase reader was required because this is the only way to read the relationship classes required to read the S121 relationships.

Other than setting the appropriate reader / writer parameters, there were no intermediate process steps required. The destination schema was directly derived from the source CHS geodatabase. The practical reason this transformation was performed was to allow pilot participants to read an equivalent S-121 database who did not have access to ArcGIS licenses. The same FME transform which was used to convert from ESRI geodatabase to S-121 GML could be used to do the same conversion using this GeoPackage database as the source simply by adding the GeoPackage as the source dataset and connecting its feature types. This was tested during the closing stages of the pilot and represents a way of creating S-121 GML using only open standards. The added benefit of this, therefore, was the testing of the use of GeoPackage as an alternate database encoding.
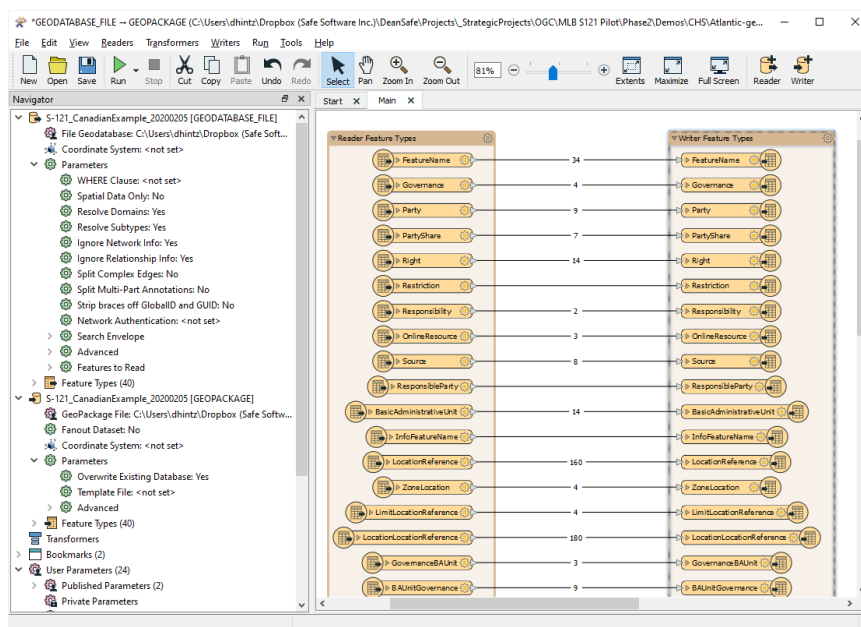
Figure 110. Conversion from CHS Continental Shelf Geodatabase to GeoPackage.
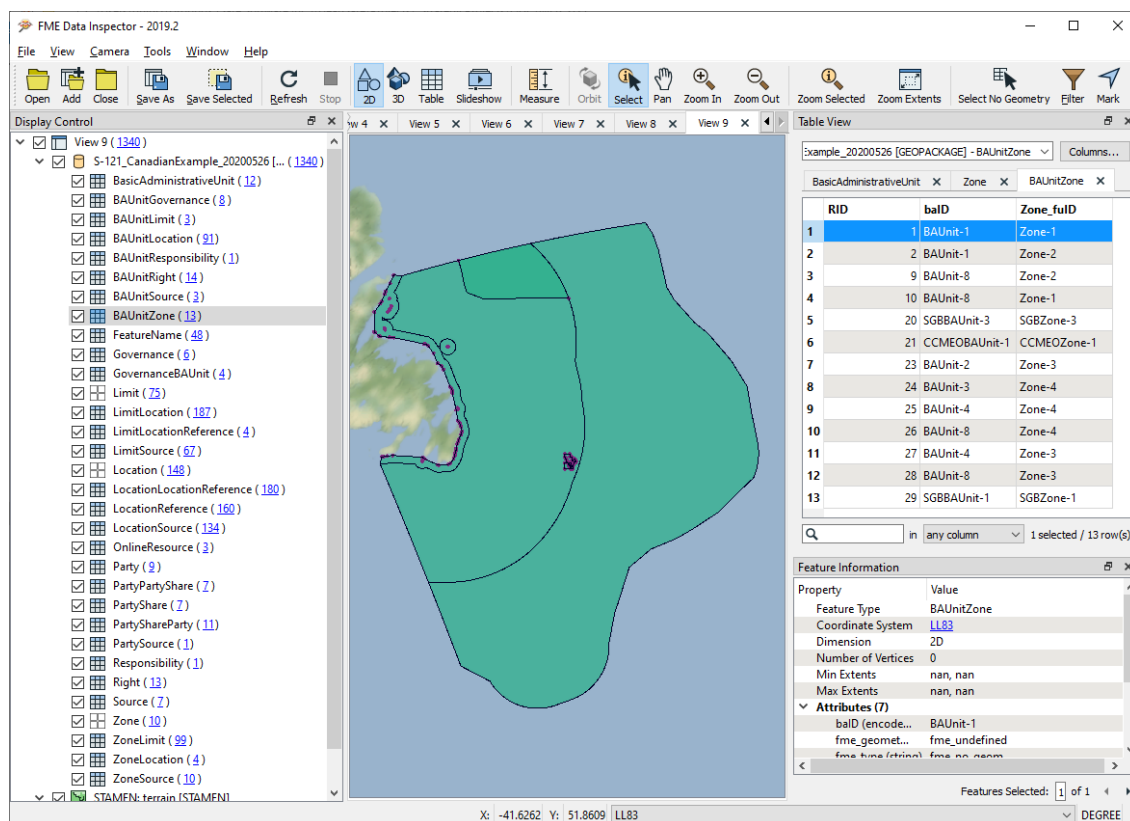
Figure 111. CHS Continental Shelf GeoPackage showing Zone/BAUnit relationship table [Note]

Challenges

There were several difficulties related to transforming data from a relational database to S-121 GeoJSON.

One was the FeatureCollection structure. One feature type is relatively straightforward to structure in GeoJSON. However, different tools approach handling of multiple feature types within one dataset in different ways. In effect, FeatureCollection in GeoJSON is more closely associated with GML feature type. Therefore multiple feature types are difficult to model within one dataset in a way that can be easily accessed by a wide range of tools. FME implements multiple feature types by building an array of FeatureCollections. However, many other tools have trouble consuming this. The workaround used here was to fanout based on feature type name and generate a separate GeoJSON file per feature type. These were uploaded to the OGC MLB GitHub which was able to render them as individual map layers – see earlier figures.

In the future, the FME workflow could be modified to handle multiple feature types within a single feature collection with a feature_type property to distinguish the feature types. This would allow a single OGC API - Features request for multiple feature types to be encoded in either GML or GeoJSON and still return a single response. If the GeoJSON encoding only allows one feature type per dataset, an equivalent response would require a response package such as a zip file containing multiple GeoJSON files, one per feature type, and this would need to be decoded before it could be displayed.

Tools such as GitHub are able to handle the display of simple and array attributes, but seem to have trouble rendering complex nested attributes such as LocationReference and FeatureNames. Note the value of [object Object] for each of these properties – see figure 6. Naturally the default rendering on GitHub is fairly basic. Custom, or more sophisticated clients should be able to parse the contents of the complex properties. Further testing with a range of GeoJSON and web clients such as Leaflet and OpenLayers would be useful, as would streaming to OpenAPI WFS3 feature service clients.
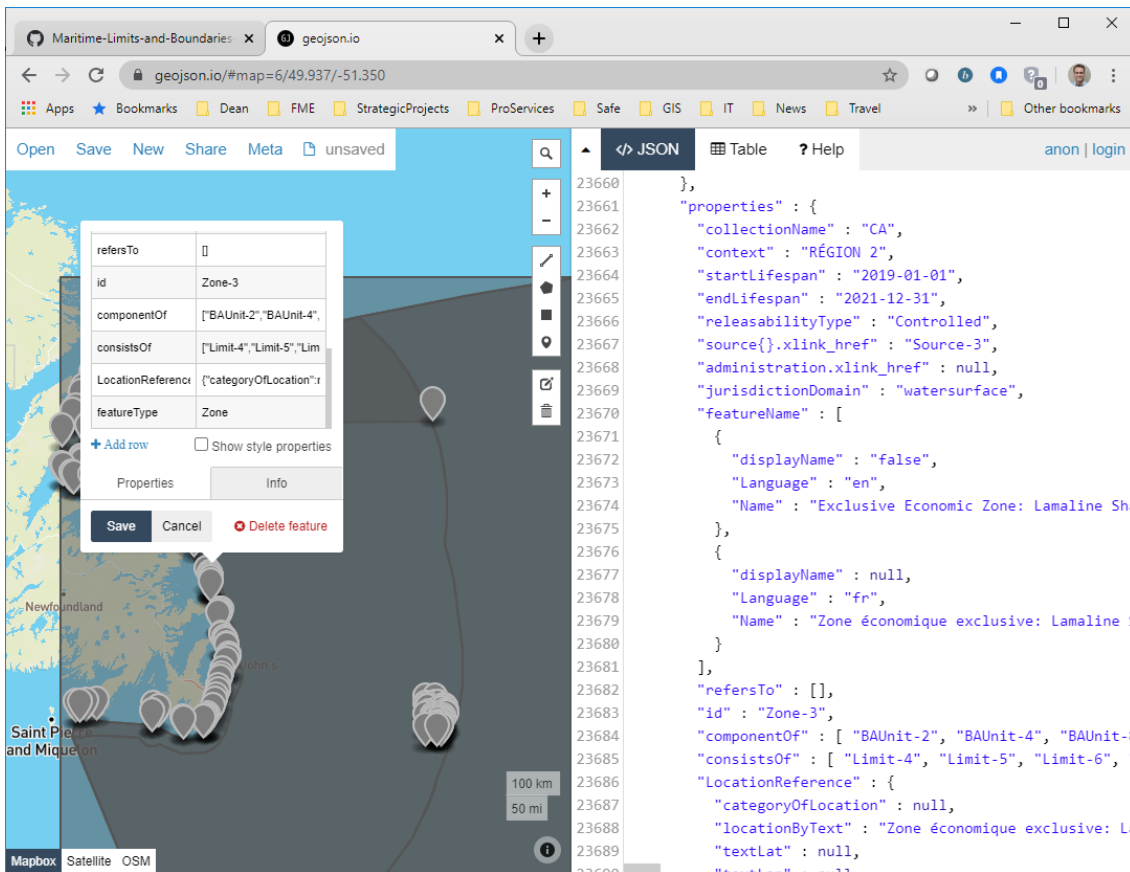
Figure 112. GeoJSON S-121 Zone feature type displayed with attribute schema using Leaflet on geojson.io. Note that relationship arrays and complex properties such as LocationReference and featureName are all read.

The image above shows a demo implementation of Leaflet with MapBox called geojson.io. Simply pasting the contents of CHS_S121geojson_v1.1_Zone.json into the data window allows this web page to render the CHS MLB S121 Zone feature type data. Leaflet appears to correctly render the geometry, simple and complex properties, including the relationship arrays such as consistsOf as well as complex properties such as LocationReference. For more on geoson.io see: https://github.com/mapbox/geojson.io/blob/gh-pages/README.md

Another challenge was the generation of relationship list structures and complex properties. This was handled by a single S121JSONConverter transformer that was inserted right before each destination feature type. Using the same custom transformer for all feature types meant that there are some empty or null value properties that show up for most feature types that do not appear when reading the equivalent GML. However, no known features or values were lost in the conversion.

Given the simple nature of the transform and similarities in relational schema, no significant challenges arose during conversion of geodatabase data to GeoPackage.

9.4.3. Conclusions

Ultimately the FME GIS Application was able to generate S121 GeoJSON that was closely aligned with the S-121 GML output generated from the same dataset. However, as there is no widely accepted way for validating GeoJSON schema, there is no obvious simple way to verify that the output generated is in fact S-121 compliant, other than by manual inspection. A business rule based approach could be used to verify valid schema, the presence of required attributes, valid field

values, code lists, and geometry types etc, though this would likely take considerable effort to implement. Other approaches using JSON schema would also be worth exploring - see: https://json-schema.org/.

It should be noted that all the geometries used to generate GML were also easily mapped to GeoJSON geometries. Default georeferencing also produced appropriately geopositioned results since both the input and output are in the same datum and projection: LL-WGS-84. Note that in this case no reprojection was required. However, GeoJSON version 3 is limited to LL-WGS84, so there would be problems with encoding data from other coordinate systems.

Simple attributes were readily generated, but some challenges were encountered with generating complex attributes and arrays. These were overcome once the differences between GML and GeoJSON were taken into account. In general, the best approach in FME GeoJSON writing for complex attributes was to use a combination of top level attributes with json datatype populated by JSON objects generated with JSON templates.

The S121 GeoJSON schema was directly derived from the S-121 GML schema. However, the list array structures and parent.child complex attribute structures are not directly supported by the GeoJSON writer. To mitigate this, parent elements were defined with data type=json to accept json objects for each complex attribute.

Validation remains a challenge. Any future testing would need to carefully address this point. One approach would be to use S121.GML as the main transport / exchange format since it is much easier to verify validity. S121.GeoJSON may be better suited as a display / view only rendering of the S121 data for web and mobile clients, but not as an auditable transport / exchange format.

GeoPackage was employed without any serious difficulties. GeoPackage should be explored further as a possible alternative encoding, given that it is a well defined and accepted open standard, and there are that there are a number of tools that can read it, it supports the encoding of an entire database in one file, it supports spatial and attribute indexes, and performs well for large datasets. One challenge which was not explored in this preliminary test was complex attributes. While it was relatively easy to perform a simple replication conversion from geodatabase to GeoPackage, we did not attempt to model complex attributes such as one to many relations or nested featureName attributes within the GeoPackage. Given that GeoPackage is a relational database, it could prove difficult to model a nested structure similar to what S121 encodings in GML or GeoJSON support. It may be worth exploring possible relational encoding guidelines for S121 data.

In general, while these results are very preliminary and only reflect testing done very late in the pilot, alternate encodings hold considerable promise in regards to making S121 and S100 extension data more readily accessible by a wide range of clients and use cases. Considering that nation states which may choose to use S-121 have considerable diversity in terms of IT resources and levels of sophistication, it is compelling to provide alternate encodings to give developers a wider range of options for building their implementations and delivering data to as wide an audience as possible. Also, some use cases such as data exchange may be more suited to S121 GML or GeoPackage, whereas use cases such as data streaming to web or mobile clients may be better supported by S121 GeoJSON. It is strongly recommended that any future work conducted in relation to maritime SDI in general, or S-121 in particular consider utilizing alternative encodings such as GeoJSON, GeoPackage, and OpenAPI services such as OGC API - Features.

Chapter 10. References

The following normative documents are referenced in this document.

NOTE	Only normative standards and conventions are referenced here from IHO, OGC and the United Nations.

- United Nations Convention on the Law of the Sea, DOALOS, 1982

- IHO S-100 Edition 4.0.0 2018, International Hydrographic Organization

- Geographic Markup Language (GML), OGC ISO 19136

- OGC Call For Participation CFP

- ISO 19152

- ISO 19115

- IHO S-57

Chapter 11. Terms and Definitions

For the purposes of this report, the definitions specified in Clause 4 of the OWS Common Implementation Standard OGC 06-121r9 shall apply. Given the domain of this report, and based on experience in the development of the S-121 standard, definitions contained in normative documents are not reproduced unless necessary, particularly in the context of the UNCLOS convention.

Therefore, terms like "Territorial Sea Area" and "Baseline" are not defined in this document but should be referenced from the normative references in the appropriate standards document and bibliography.

11.1. Abbreviated Terms

The following clause gives a list of the abbreviated terms and the symbols necessary for understanding this document.

- API - Application Programming Interface
- CHS - Canadian Hydrographic Service
- CMDS- Common Maritime Data Structure
- COTS - Commercial off-the-shelf
- CRS - Common Reporting Standard
- CSW - Catalog Service for the Web
- DCEG - Data Classification and Encoding Guide
- DFO - Fisheries and Oceans Canada
- DOALOS - Division for Ocean Affairs and the Law of the Sea
- DSSI - Dataset Structure Information
- ENC - Electronic Navigational Chart
- FOID - Feature Object Identifier
- GDB - ESRI Geodatabase
- GIS - Geographic Information System
- GML - Geographic Markup Language
- GMLAS - Geography Markup Language driven by Application Schemas
- HRT - Human Readable Text
- HTML - Hypertext Markup Language
- IHO - International Hydrographic Organization
- ISO - International Organisation for Standards
- JPEG - Joint Photographic Experts Group
- JSON - JavaScript Object Notation

- KML - Keyhole Markup Language
- MapML - Map Markup Language
- MDWG - Marine Domain Working Group
- MLB - Maritime Limits and Boundaries
- MOU - Memorandum of Understanding
- MSDI - Marine Spatial Data Infrastructure
- MSDIWG - Marine Spatial Data Infrastructure Working Group
- OGC - Open Geospatial Consortium
- OWS - OGC Web Services
- PDF - Portable Document Format
- PNG - Portable Network Graphics
- SOLAS - Safety of Life at Sea
- UKHO - United Kingdom Hydrographic Office
- UML - Unified Modeling Language
- UNCLOS - United Nations Convention on the Law of the Sea
- URL - Uniform Resource Locator
- WFS - Web Feature Service
- WMS - Web Map Service
- WMTS - Web Map Tile Service
- WPS - Web Processing Service
- XML - Extensible Markup Language
- XSD - XML Schema Document

Appendix A: Human Readable Text

Some extracts from the HR text deliverables are presented in this annex. The production of these required a full end to end pipeline within the project of:

1. Data transformation from sponsor's original form to GML

2. Validation of GML against Schema

3. Transformation using XSLT to text form in two languages

4. Validation of text against GML data source

Territorial Sea Geographical Coordinates Order

C.R.C., c. 1550

Date modified :
2019-01-01

SCHEDULE I

In respect of the portions of Area 2 for which the geographical coordinates of points are listed in Schedule I, the baselines are straight lines joining the points so listed.

AREA 2

Straight Baseline: Southeast and East Newfoundland

Original Position

	Locality	Latitude	Longitude	C.H.S. Chart
Limit-1				
Location-1	Lamaline Shag Rock	46°50′21″N	55°49′30″W	4016
Location-2	Shag Rock	46°50′17″N	55°44′51″W	4016
Location-3	Offer Island	46°51′21″N	55°37′25″W	4016
Location-4	Farmland Head	46°52′16″N	55°33′04″W	4016

Figure 113. English Version

Décret sur les Coordonnées Géographiques pour la Mer Territoriale

C.R.C., ch. 1550

Date de dernière modification :
2019-01-01

ANNEXE I

En ce qui concerne les secteurs de la région 2 pour lesquels les coordonnées géographiques de points sont établies dans l'annexe I, les lignes de base sont des lignes droites reliant les points qui sont ainsi indiqués.

RÉGION 2

Ligne de base droite: Sud-est et Est de Terre-Neuve

Position d'origine

	Localité	Latitude	Longitude	Carte S.H.C.
Limit-1				
Location-1	Lamaline Shag Rock	46°50′21″N	55°49′30″O	4016
Location-2	Shag Rock	46°50′17″N	55°44′51″O	4016
Location-3	Offer Island	46°51′21″N	55°37′25″O	4016

Figure 114. French Version

These were produced using an XSLT transform from GML encoded using the schema created within the project. An extract of the GML is shown below:

```
        <s121:BasicAdministrativeUnit gml:id="BAUnit-2">
            <s121:componentOf xlink:href="#Governance-1"/>
            <s121:componentOf xlink:href="#Governance-2"/>
            <s121:componentOf xlink:href="#Governance-3"/>
            <s121:startLifespan>2019-01-01</s121:startLifespan>
            <s121:endLifespan>2021-12-31</s121:endLifespan>
            <s121:featureName>
                <s121:displayName>false</s121:displayName>
                <s121:Language>en</s121:Language>
                <s121:Name>Exclusive Economic Zone: Southeast and East Newfoundland</s121:Name>
            </s121:featureName>
            <s121:basicAdministrativeUnitName>Exclusive Economic Zone: Southeast and East
Newfoundland</s121:basicAdministrativeUnitName>
            <s121:basicAdministrativeUnitType>Maritime Limits and Boundaries</s121:basicAdministrativeUnitType>
            <s121:basicAdministrativeUnitContext>AREA 2</s121:basicAdministrativeUnitContext>
            <s121:administrationOf xlink:href="#Zone-3"/>
        </s121:BasicAdministrativeUnit>
        <s121:Governance gml:id="Governance-1">
            <s121:startLifespan>2019-01-01</s121:startLifespan>
            <s121:endLifespan>2021-12-31</s121:endLifespan>
            <s121:label/>
            <s121:referenceNumber>C.R.C., c. 1550, ss. 4(1) / C.R.C., ch. 1550, art. 4(1)</s121:referenceNumber>
            <s121:featureName>
                <s121:displayName>false</s121:displayName>
                <s121:language>en</s121:language>
                <s121:Name>Territorial Sea Geographical Coordinates Order, SCHEDULE I</s121:Name>
            </s121:featureName>
            <s121:governanceTitle>SCHEDULE I</s121:governanceTitle>
            <s121:governanceDescription>In respect of the portions of Area 2 for which the geographical coordinates of
points are listed in Schedule I
, the baselines are straight lines joining the points so listed.</s121:governanceDescription>
            <s121:releasabilityType>Controlled</s121:releasabilityType>
            <s121:dateApproved>2019-01-01</s121:dateApproved>
            <s121:dateConsidered>2019-01-01</s121:dateConsidered>
            <s121:dateIntroduced>2019-01-01</s121:dateIntroduced>
            <s121:governanceOf xlink:href="#BAUnit-1"/>
            <s121:governanceOf xlink:href="#BAUnit-2"/>
            <s121:governanceFor xlink:href="#BAUnit-5"/>
        </s121:Governance>
```

```
XSLT Extract:

                         <!-- Each BAU in this governance -->
                         <xsl:for-each select="s121:governanceFor">
                             <xsl:sort select="@gml:id"/>

                             <xsl:variable name="_bauId" select="substring(@xlink:href,2)"/>
                             <xsl:for-each select="$root/s121:members/s121:BasicAdministrativeUnit[@gml:id = $_bauId]">

                                 <h3><xsl:call-template name="printSeparatedLocale"><xsl:with-param name="text"
select="s121:basicAdministrativeUnitContext"/></xsl
:call-template></h3>
                                 <h4><xsl:call-template name="printSeparatedLocale"><xsl:with-param name="text"
select="s121:basicAdministrativeUnitName"/></xsl:ca
ll-template></h4>

                                 <br/>
                                 <em><xsl:value-of select="document('')/*/t:dico/e[@k='originalposition']"/></em><br/>
                                 <br/>

                                 <table>
                                     <col width="120" />
                                     <col width="300" />
                                     <col width="160" />
                                     <col width="160" />
                                     <col width="120" />
                                     <thead>
                                         <tr>
                                             <th id="t1207h4" headers="t1207h0"></th>
                                             <th id="t1207h5" headers="t1207h1"><xsl:value-of
select="document('')/*/t:dico/e[@k='locality']"/></th>
                                             <th id="t1207h6" headers="t1207h2"><xsl:value-of
select="document('')/*/t:dico/e[@k='latitude']"/></th>
                                             <th id="t1207h7" headers="t1207h3"><xsl:value-of
select="document('')/*/t:dico/e[@k='longitude']"/></th>
                                             <th id="t1207h8" headers="t1207h4"><xsl:value-of
select="document('')/*/t:dico/e[@k='chart']"/></th>
                                         </tr>
                                     </thead>
                                     <tbody>
                                         <xsl:for-each select="s121:administrationOf">
                                             <xsl:sort select="@xlink:href"/>
                                             <xsl:call-template name="featureAsRow">
                                                 <xsl:with-param name="gmlId"><xsl:value-of
select="substring(@xlink:href,2)"/></xsl:with-param>
                                                 <xsl:with-param name="categoryOfInformation">OriginalLocation</xsl:with-
param>
                                                 <xsl:with-param name="schedule"><xsl:value-of
select="$_ogvTitle"/></xsl:with-param>
                                             </xsl:call-template>
                                         </xsl:for-each>
                                     </tbody>
```

Appendix B: Revision History

Table 1. Revision History

Date	Editor	Release	Primary clauses modified	Descriptions
September 23rd, 2019	J. Pritchard	0.8.9	all	Initial version, end Phase 1
March 31st 2020	J. Pritchard	0.9.5	all	Initial Draft final, end Phase 2
April 22nd 2020	J. Pritchard	0.8.9	all	Reviewed Draft, Pre OGC TC Version