Paper for Consideration by S-100WG5

Realtime data and S-100

Submitted By:	IIC Technologies
Executive Summary:	Support for real-time data within S-100 requires some enhancements to S-100 and its constituent parts. This paper explores the nature and size of the task required and proposes some potential solutions for harmonising the most appropriate elements of S-100 within a web services framework for real-time data support..
Related Documents:	IHO S-100, IHO S-57, OGC SWE, OGC API (Core)
Related Projects:	OGC Marine Domain Working Group.

Introduction / Background

The IHO S-100 Universal Hydrographic Data Model provides the framework for the development of various digital data products intended to represent maritime world. It serves as a foundation which can used by various national agencies and user communities to capture, produce and exchange information in a well-structured, interoperable and reliable way. Until now, S-100 initiatives have not focused on developing support for delivering real time information from different broadcast points to an end system. S-100 includes a section 14 – Online Data Exchange, which serves as a conceptual basis for streaming of S-100 datasets and introduces a data model to describe protocol-based stacks and data services but without many of the essential details for implementors.

There are currently some S-100 based product specifications under development whose data model describes entities which can be captured and distributed most usefully in real-time, and the aspiration is to provide a useable framework which enables such facilities under the S-100 umbrella. For example, data encoded in IHO S-104 could be broadcast in to show water level at a specific location and its changes for a specific area. With use cases for real-time data exchange emerging it is desirable to add support for real-time data feeds at the S-100 level to establish a common framework. This paper makes a number of observations on the problem prior to the definition of new content within S-100 specifically aimed at support for such data types.

Analysis/Discussion

Generally S-100 focuses on providing a way of describing data products which are relatively static in nature with product update intervals usually measured in weeks or months rather than minutes or seconds. Data products are generally “compiled” through a variety of means from data originating in a number of sources and sensors. S-100 provides a way, through its product specifications, of mapping real-world phenomena to a set of instances (and interrelationships) of conceptual entities defined in the IHO Geospatial registry. These “compiled” instances are packaged into self-contained data exchange sets and distributed, usually, using traditional file-based data exchange mechanisms.

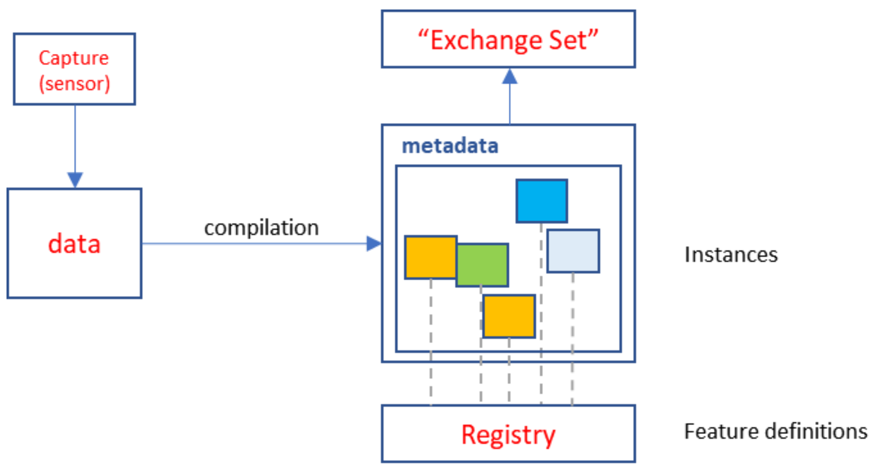

Figure 1: Exchange set compilation and packaging from raw data

In this model packaged data in the form of exchange sets are generally "issued" – assigning a timestamp or other meaningful temporal mark within the metadata and locating the entire exchange set to some point (or period) in time. The package of data content can then be transferred by some means to a recipient system – indeed the model presupposes the existence of an "end user system" compatible with the data form and within which some form of user interaction is intended.

This model reflects S-100's origins in chart compilation and embodies its applicability for establishing reliable exchange of data between originators, through a data chain and into an end user system designed specifically for particular use cases with a defined stage of "compilation" where real-world data are classified according to IHO definitions as features.

The essential characteristics, therefore, of the current IHO paradigm are:

Characteristics	IHO S-57 – ENC	IHO S-100, ENC + others…
Data Capture	Survey, Imagery, Previous editions	Survey, Imagery, Previous editions
Compilation	UOC, member state cartographic practice	DCEG, Feature Catalogue
"Registry"	S-57 Appendix 1 Objects and Attribute Catalogues, S-57 AppB1	IHO Geospatial Registry
Format	ISO8211	ISO8211, GML, HDF etc…
Metadata	DSID,DSPM,CATALOG.031-ISO8211	S-100 Part 4a XML
Exchange Set	ENC_ROOT	Multiple products, expressed within XML metadata.
Service Levels	PRODUCTS.TXT SERIAL.ENC	Part 15, service.xml
Issuing timestamp	ISDT,UADT	Within XML metadata
Distribution	S-63, file based, through intermediaries	S-100 Part 15, file based, through intermediaries

This model also reflects some of the key characteristics of the marine environment in which S-100 data (ENC and other product specifications) is used, e.g.

1. Restricted bandwidth (disconnected). Recipients and distributors frequently need to use hard media for transport and import to a system. The end user system does not necessarily have access to the Internet for e.g. verification or update.
2. Data integrity (liability). A huge amount of liability is attached to navigational (and auxiliary) processes. A great degree of safety-critical thinking is therefore applied to the compilation and transport of data to users under the IHOs data protection scheme. This process digitally signs actual data rather than its transport to the end user.
3. Geographic extent. Uses for IHO standards based data are global and require consistency across all holdings which has led to an extensive standardisation organisation to provide this consistency
4. Encodings (AIS) – The nature of the marine environment also leads to communications mediums not generally available on land. IN particular are VHF, still used for MSI broadcasts and AIS which can be used for messaging – the current Part 14 of S-100 includes descriptions of such protocols and these should be accommodated in any proposal
5. Regulatory scope (SOLAS) – there is a large influence from the SOLAS framework which defines the requirement for carriage of navigational data.

The IHO S-100 standards (and its predecessors) are mature, time-tested and widely used mechanisms which work very well in their environment and are eminently suited for their use cases. However, the model is generally not suitable for some emerging data products which require data to be exchanged in real-time or close to real-time.

Real-time data is characterised by its inherently "streamed" nature where there may be no "compilation" stage (as per the description at the start of this section) at all, nor an overall "exchange set" aggregating multiple registry entity instances together in an overall "package".

In this sense real-time data modelling and distribution can resemble a data stream of a particular kind of timestamped observations or data points. With a streamed data service the key difference being is that it flows through the system as it is generated, rather than in a packaged datasets. With real-time data a temporal element is an integral part of the structure (discussed in detail later) and although a streamed service is often sufficient there is a need (as seen in the use case example in this document) to also provide "cached" services in multiple encodings (tidal/water level data being a key case in point).

Terrestrial Web Services

Terrestrial data exchange has evolved in a more highly connected environment, on the internet, and the predominant model is one of web services, where a "connection" is initiated between the data provider and the end user and data is then fed through the connection in an agreed form using an encoding embedded within a protocol, such as http. Data integrity is assured, most usually, through https[1] and an online system of assured identity where the system's online connection to a global network of certification authorities (CA) reference the identity of the issuing authority. (The IHO system (the data protection scheme) is not unlike it in structure (and it shares many of the same protocols under the new IHO Part 15).)

The task, therefore, of supporting real-time data comprehensively in S-100 is, therefore, one of integrating those existing S-100 mechanisms and structures which are useful (and which reflect the underlying marine character of the data itself) and enhancing the S-100 framework to support the essential different character of such real-time streams. This goes far beyond establishing new data encodings as it challenges the entire nature of the "exchange set" model of distribution. Instead, a new mechanism needs to be established covering all machine-to-machine communications, incorporating data encoding, presentation, transmission protocols and service level elements.

This technical brief aims to outline various options how this could be achieved and serves as a positioning paper for selecting the real-time data path to pursue in S-100.

S-100, Data Streams

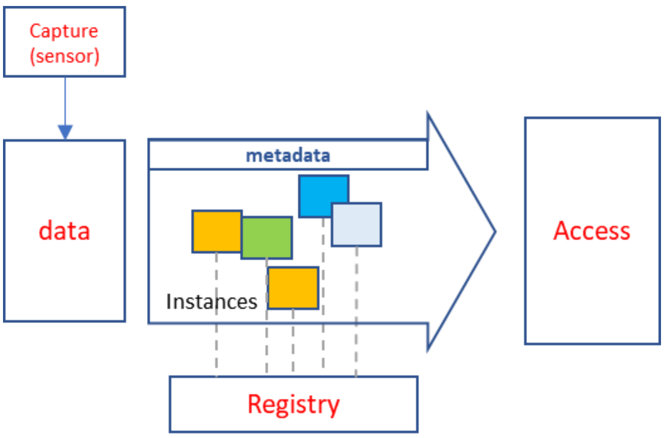

Figure 2: An S-100 realtime data stream

In the above diagram a hypothetical S-100 data stream is described. In it a set of S-100 features are streamed from a sensor (or sensors) along with service-level metadata through to a set of disparate end users via a standardised API. Features continue to be conformant with the relevant parts of S-100 and align with the definitions and bindings specified in a particular product specification and their definitions within the IHO geospatial registry.

[1] A key difference. In https the data stream itself is encrypted and digitally signed. In an environment which is predominantly disconnected such options are difficult to implement so data is pre-encrypted and signed so it can be distributed either over (possibly unencrypted) data streams (like http or ftp downloads) or via hard media. The connected nature of modern data services merges the two processes (the encryption/signing and the data transfer) together.

In order for S-100 to be adapted to encompass such a model the existing exchange set (and Part 14) and delivery mechanisms will need some adaptation. The following sections describe these in brief – in reality a new part of S-100 is likely to be required to draw the necessary pre-existing and new elements together to provide a normative method of distribution for such data.

Creation of Real-Time Data "Pipelines"

The traditional data exchange mechanism outlined in the initial section of this paper relies in practice on four distinct stages:

1. Data Capture by one or more of a set of sensors – data in its "raw form"
2. Data Compilation into a set of IHO Registry-defined features according to a model-driven application schema defined within the product specification.
3. Packaging into a standardised exchange set including any encryption necessary and digital signing of data to provide an authenticated, traceable data integrity mechanism for the collection of features in its entirety.
4. Distribution to end users through (currently file-based[2] mechanisms)

Whilst these elements address complex sets of the underlying requirements, they are relatively straightforward to implement especially given the mature status of existing implementations and established distribution structures in the end user community. The application scope is relatively narrow, the main concerns being reliability packaging, integrity and making the data available to end users.

In contrast, real-time data dissemination is a more complex undertaking which entails the construction of a "pipeline" rather than a compiled dataset where data is automatically transformed into instances of feature types at particular points in time. Given its nature as a pipeline there is a far greater emphasis on the access mechanisms by which end users acquire data rather than the packaging and arrangement of multiple features within a "container" dataset.

There is also a far greater potential application scope and so the pipeline not only needs to address the same data packaging and protection aspects but also it must cover the service aspects of data transmission protocols, connectivity, service operation including managing service levels and fail-safe provisions. Such a structure likely needs to establish new provisions for data portrayal, updating, and data protection as their traditional equivalents are not likely to be feasible for implementation in the limited data bandwidth of the marine environment.

Many of these aspects are broadly described in the existing section 14 of S-100. As such this paper does not aim to redevelop them but rather build on the existing baseline while adding new information related to the more recent developments in this field. Some proposed groupings follow which explain the scope of the task of definition for real-time data support.

Real-Time Data Characteristics in S-100

The nature of real-time data should be considered primarily in order to define the core of the real-time support in S-100. Real-time data:

- Tends to be multiple measurements of (normally) physical phenomena with (largely) repeated locations (and, therefore, a smaller set of distinct feature types) and with each measurement attributed with a temporal value marking the time of measurement and/or access. This is complemented by metadata for the collecting sensor/device/station and other contextual data (probably of a more traditional static form)

[2] "file-based" in this sense does not mean that it is always file protocols (like hard media or FTP) which are used but the data features remain embedded in a file which is then compressed (S-100/S-63 use zip algorithms) split into parts and reconstituted by protocols for transmission (like FTP/HTTP).

- Can be very scientific in nature with large numbers of attributes and values[3].
- Can be either a pure pipeline where features are available "at an instant" or on request for a (possibly limited) window of requests. For instance a real-time water level sensor might be interrogated to return only the current value or all values in the last (e.g.) 12 hours.
- Can be potentially encoded in a number of different forms depending on user preference and its nature or volume. These forms/encodings are often required dynamically on demand.

The main similarity is that the actual data content itself can be modelled and expressed through S-100's existing mechanisms, using a defined application schema and encoding such as GML, ISO8211 or HDF5. These encodings can all be used to distribute real-time data according to the content within the model. The important thing to note is that features can be conformant with the IHO geospatial registry and expressed in an S-100 encoding. This suggests that S100 requires:

- Mechanisms for packaging of arbitrary quantities of individual features together (from single features to multiple features located in different points of space and time). This means a reformulation of the feature aggregation process and construction of exchange sets to something more at a feature level so that individual features may be distributed without the overhead of the entire dataset "packaging"
- The emphasis on real-time implies that a temporal "stamp" is an integral part of such features and therefore these data require such a timestamp at the GFM level rather than at (possibly optional) feature attribute level.
- The requirement for support of a limited bandwidth environment and the potential for very large numbers of individual features to be accessed implies the need to keep data volumes as low as possible and also to avoid the repeated issuing of complex metadata and catalogues within individual transactions.
- The requirement for reliable data integrity and protection mechanisms implies means of doing so either within the "streaming" mechanism (as per traditional web services) as well as at a feature level within the aggregated response to the access request. This could be accomplished by extending S-100 Part 15 to individual features in the encoding adopted by individual product specifications.

Data Access Mechanisms and APIs.

Data Access for real-time data is likely to be predominantly API driven. A defined API provides a specification for formatting a request for data. This request is then processed by the data source and the content returned accordingly. This is a reasonably standard web services model with a long history and many implementations where an API is defined and data can be requested using a set of predefined parameters.

Using the S-100 framework features are defined within a feature catalogue and portrayal defined according to a portrayal catalogue. For ECDIS-based data additional catalogues can define interoperability with other data layers, feature interrogation formatting and alert/indication behaviour. All data conforming with the product specification is thus covered by such catalogues. There is a need to establish compact API responses for data so that catalogues are not continually re-sent to requesting clients. These elements are all reusable but could be "abbreviated" by distributing links to online versions of such catalogues for later onward use (and using the inherent online nature of API services). Additionally different categories of API calls can request individual catalogues or actual data itself (which conforms to those catalogues).

The other observation to make at this stage is that instead of providing standards relating to the definition of file-based exchange sets IHO S-100 should consider the specification of standardised API calls to ensure uniformity of data service components within a real-time pipeline. There are numerous candidates for this kind of approach,

[3] Consideration should be given to the establishment of domains specific to real-time data. Interoperability of sensors with their representation in other product specifications is important but due to their temporal nature and scientific use it may be logical to group their features in their own domain?

most notably from the OGC family of web services and consideration should be given to the lessons learned from the OGC in its latest form (described later).

Data Portrayal

In the real-time case where data is presented through an API (conceptually similar to the S-100 Part 14 model) the point at which data has its portrayal "applied" could be either before or after the actual API request is generated. For instance, a request for a time series of data points could return either a vector array of individual data values (in a single location with different timestamps) or a rendered graph of such an array (as shown in the example). Tidal levels could superimpose predicted on top of actual values as well as allowing access to the raw data as well.

There is an opportunity to reuse the existing portrayal mechanism or to reuse some similar structures found in the existing standards base for geospatial data in other domains. Portrayal mechanisms which have a high degree of simplicity are also better oriented towards simple features which may be distributed in large quantities.

Existing S-100 Part 14 and Data Exchange

Transmission of real-time data through AIS should certainly be considered – this is already included in S-100 Part 14 and S-100 should continue to support this mode of operation for product specifications where it is required. The content of Part 14 of S-100 could be enhanced to include a more abstract definition of APIs and web services to provide a way of defining APIs for access to real-time data.

The existing S-100 part contains much useful information including a model for service definition, the layered model of data streams and REST interface definitions for data. If the data content of such services is enhanced to include real-time data features (and supporting structures like Part 15 (data integrity) are combined with multiple encoding mechanisms it is clear that a powerful framework could be constructed which would integrate static S-100 data from product specifications with real-time data accessed through APIs and preserving the many marine-domain specific developments currently within IHO S-100.

An Example: NOAA water level sensors

An example which illustrates the kind of mechanisms required under S-100 is NOAAs' water level sensors, from the website at https://tidesandcurrents.noaa.gov/ports.html

Within this portal a number of individual stations are accessible along with time-series data rendered into graphs of predicted/actual water levels and currents (amongst many other parameters as well as station information, tidal harmonics, datums etc…) This provides a good use case for the implementation of an S-100 based version of this data accessed through a (to be defined) standardised API and illustrated many of the concepts and observations made within this paper.

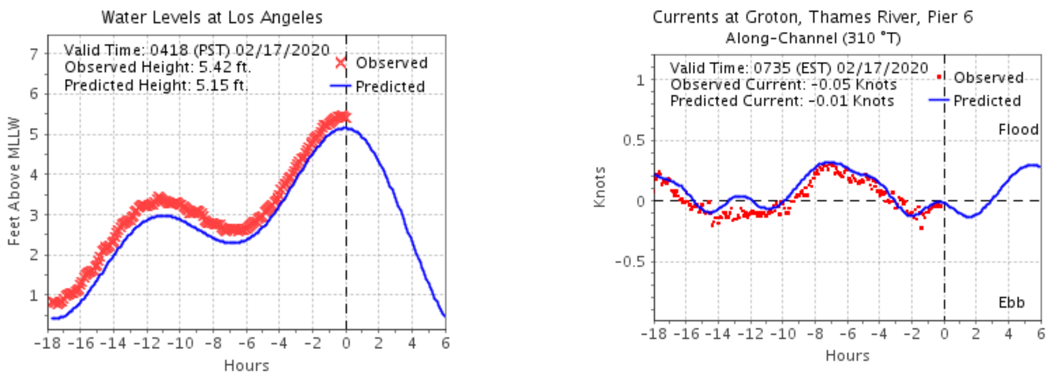

Figure 3: Example Water Level and Current images

- Data is simple in nature (often a single floating point measurement of a defined physical quantity) but potentially large in volume (e.g. the predicted values shown in the graphs). In this case the defined S-100 "feature" would be observed water level or current. Other features within the product specification would cover station information including position and name/timezone information.

- Although we show rendered images of data it would be useful for some clients to be able to access the raw data as well (in standardised form) as unrendered arrays of vector data (encoded in, e.g GML, JSON, CSV or ISO8211)

OGC Web services and latest developments

The development of OGC web services has progressed over many years from initial WFS/WMS to the latest OGC API (formerly known as WFS3) and there is much to learn from their perspective on development of web services to support a very broad variety of data content. The new generation of web services have a number of aspects which are relevant to the discussion here, namely:

1. Preservation of the most valuable elements of the web services model, the request-response action in order to obtain data for a client from a "service provider"
2. A "content neutral" concept of the web service where the encoding of the returned data can be defaulted or be determined by the content of the API call. This extends to many of the mandatory query parameters in the API call allowing the source domain (i.e. IHO) to specify the API calls.
3. Families of API calls for accessing metadata, schemas and other elements as well as content itself.

https://www.opengeospatial.org/standards/ogcapi-features: The newly published (2019) API specification contains much of value to the definition of support for real-time data within S-100 and much of its core ideas and mechanisms could be implemented or reused within S-100WG for engineering real-time support.

OGC also have experience within the real-time domain through the Sensor Web Enablement (SWE) framework – this long-standing framework of interconnecting standards builds on the existing WFS/WMS (primarily XML/GML based) structure to define a family of methods for modelling, access and encoding of data from a wide variety of sensors in different domains. Of interest is the SWE Common data model encoding which contains a lot of relevant structures around modelling of sensor observations and their transfer and the SOS (Sensor Observation Service) which models how observations are recorded and uses in onward services.

- https://www.opengeospatial.org/node/698
- https://www.opengeospatial.org/standards/sensorml
- https://www.opengeospatial.org/standards/swecommon

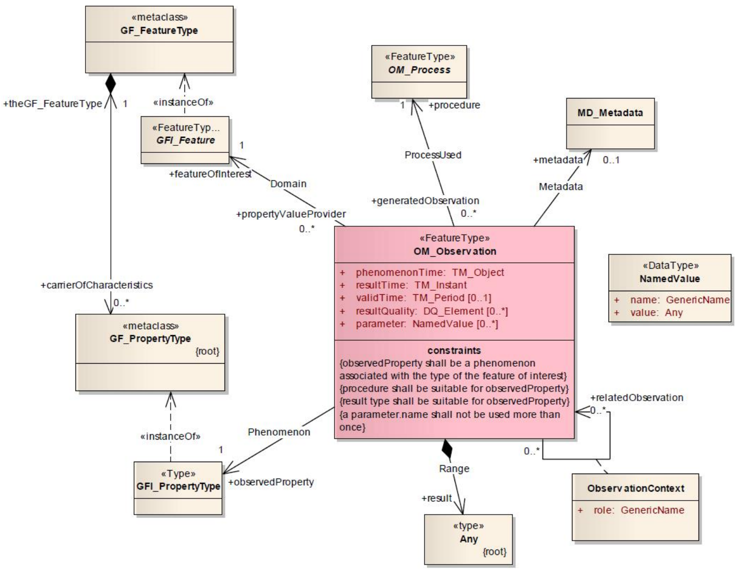

Figure 4: OGC Observation Model Overview.

There is much to gain from a consideration of the OGC model of both API development for web services (which may be useful in redefining some of the content in S-100 Part 14) and the SWE framework (specifically for enhancement to S-100 in respect of real-time data services). Whether it is possible to wholly reuse (or profile) such standards remains to be seen and requires more analysis and input from S-100WG experts but their existence should be noted and lessons learnt from organisations who have implemented such systems in the past. Certainly the IHO domain has some unique challenges (listed here) but the dynamic and content-neutral nature of the newest OGC standards in an API context is a model S-100 could usefully learn from and implement.

Conclusion and Recommendations

Conceptually there is little stopping enhancements to be made to S-100 for support of real-time data. This paper has defined areas of S-100 which are well suited to production of such data and suggested areas where some enhancements are required. There are also many good use cases within the IHO domain which can be used to guide such enhancements.

There is also much to be learned from the existing OGC standards in the area of web services API definition and the definition of sensor networks to produce a “best of both worlds” addition to S-100 which will maximise interoperability with such existing standards (easing implementation) and enable straightforward implementation in open software.

It is recommended that additions to S-100 be defined within a new part with some impacts (and modernisation of) the existing Part 14. New content would contain:

1. Recommendations for modelling of features for use in promulgation of real-time data feeds. This would provide a minimum for such features and guarantee their interoperability with other S-100 product specifications
2. An abstract definition of APIs to support the request and retrieval of real-time data conforming to S-100 product specifications. This should include prescriptive method specifications to ensure retrieval of relevant S-100 catalogue information and metadata as well as data content itself.
3. Data content formats independent of encoding as per S-100 product specifications. Consideration should be given to support for encoding in AIS messaging as well as the more traditional encodings. Web-based encodings such as geoJSON and GML can sit alongside more complex binary encodings such as ISO8211 and HDF as per existing S-100
4. Mechanisms for adapting S-100 Part 15 to provide dynamic and feature level data protection and authenticity mechanisms, recognising the importance of data integrity to end users of such data operating within strict regulatory frameworks.

Action required of S-100WG5

The S-100WG is invited to:

- Note the contents of this paper as a summary of the current situation in regard to support for real-time data within S-100
- Consider the definition of a new part of S-100 specifically aimed at support for such data and the elements it should contain, noting the differentiating factors listed in this paper
- Define additional key use cases and domains in which real-time data plays a large part. Drawing adequate scope around such definitions is key to its success
- Acknowledge the useful work already in the public domain by the OGC and explore how this can be used, profiled and potentially referenced in order to avoid its re-invention for the IHO use case.