Paper for Consideration by S-100WG

Management of Supporting Resources in S-100 Products

Submitted by:	IIC Technologies
Executive Summary:	Description of how Supplementary Files are managed in S-100 Products
Related Documents:	S-100 Edition 5.0.0
Related Projects:	S-100 Part 4a Exchange Set Metadata

Introduction / Background

The amendments to S-100 Metadata in support of exchange catalogue creation and maintenance are crucial to the implementing system and how it imports S-100 data. S-100 product specifications are free to include, as part of their model the ability to include external resources associated with instances of features or information types. In S-57 these were referred to as "Supplementary Files".

The objectives of support file management in S-100's metadata part is to ensure that S-100 is consistent in relation to:

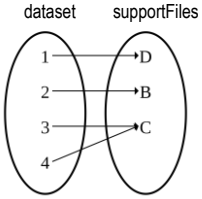
1. The ability to store external resources in exchange sets and represent them in exchange catalogues.
2. The provision of "referencing" between datasets and external resources on which they are dependent. This should allow for aggregators and exchange set creators to arrange such content independently for more efficient management and align to the defined exchange set folder structure defined in S-100
3. What is put into S-100 datasets (for all product specifications), including their primitive type, to reference external content and how such content is located from the value placed in the (sub-)attribute field.
4. How metadata content (files/"resources"[1]) are able to implement capabilities for referencing, signature, compression, and encryption

Analysis/Discussion

The following points summarise the approach taken in the metadata part of S-100 and what is described in the text in Section 4a-4.3

1. There does not have to be a single fixed "format" for the reference between the dataset and the support files in the metadata catalogue. However, such references must be surjective[2] and it must be calculable using the information which will be in the supportFileDiscoveryMetadata. This allows resources to be shared between multiple datasets.
2. The reference from the dataset to an external resource (e.g. a support file) will be in the form of a URI. If the resource is contained within the metadata (and in file based exchange sets they always will be) the URI must be resolvable without any additions to the metadata. Different types of URI are possible.

[1] Files are really "resources" which are external to the datasets. It is recommended that the term "resources" is used to describe data which is not part of the S-100 dataset for neutrality.



[2] It's ok for multiple datasets to map to a single support file but not the other way round (a single dataset reference can only map to a single supportFile.

3. References within datasets should always be defined as URI types in the Feature Catalogue so that implementing systems know that attribute values refer to an external resource and that the value needs to be parsed to establish the location of that resource.
4. By using URIs the existing (S-57) system can be closely mimicked (by using file URIs). This provides a persistent (i.e .latest version) identifier for the resource but (if it is a file) it would have to be unique. A file URI is formatted as "file::<File Name>". This also means the support file contents can be updated without an explicit dataset update. If producers want to ensure a dataset update is required when support file contents change then digital signature or hash URIs will satisfy this use case.
5. The URI to locate an external resource can also be an MRN, i.e urn:mrn:[MRN…]. These can be maintained either by the producer for long-term, persistent unique identifiers or they can be auto-generated from the file content using either a digital signature or a calculated hash value. MRN address space is reserved to define these, e.g. urn:mrn:iho:dsig:[digitalSignatureValue] or urn:mrn:iho:sha256:[hash value]. These are defined in Part 15. Alternatively (longer term), the URI could use other URN schemes (e.g. urn:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6) to identify content and, ultimately, URLs could even be introduced for web resources.
6. Digital Signatures can be used for the identifier in the CATALOG.XML reference but may complicate data production for internal exchange in terms of key management. Digitally Signed datasets can be subsequently re-signed, preserving the unique maps without loss of information, providing for a use case of internal exchange followed by a final "publication" of the dataset with all its internal resources. Under this system, use of "internal" keys by producers would be consistent and all signature certificates are always contained in the CATALOG.XML.
7. Use of hash algorithms to reconcile references is dependent on support files being unencrypted which is the case in the current metadata/Part 15 specification.
8. Exact producer best practices will vary. There are probably use cases for both a persistent identifier (as today with file names) and one which changes as content changes (precipitating updates to the dataset)
9. Producers have to make sure the references are well-formed and complete in the metadata. If using MRN they must be unique and, by implication, if using the file name for a URI, the filenames will need to be unique within the dataset.

Recommendations

This paper recommends accepting the text and associated changes to exchange set specifications in S-100. It provides a scalable, backwards compatible system which is consistent with the S-100 model for management of exchange set external resources and addresses many issues and ambiguities unspecified in S-57

Justification and Impacts

Management of supplementary files under S-57 caused difficulties for implementers and dataset aggregators even with static feature/attribute catalogues. S-100 has the opportunity to make specific the methods by which all external resources are managed which is scalable to all product specifications.

Action Required of S-100WG

The S-100 working group is asked to:

1. Note the additions to S-100 in respect of supplementary file management.