19th Meeting of the Data Quality Working Group

Automated cross-check between S-101 FC and DCEG

Agenda Item 05.1B

DQWG-19, VTC Event, 25 - 26 March 2024

International
Hydrographic
Organization

- Cross-check between S-101 FC and S-101 DCEG (version 1.1.0)

- Python script developed to facilitate cross-check

- Works as follows :

- Read and parse the Feature Catalogue (.xml file)

- Read and parse DCEG (.docx file) and selection of information to be compared

- Comparison of the presence/consistency of information extracted from the DCEG and the FC

```
cross_check_FC_S101.py  X    compare_files.py  X    read_files.py  X

1 import read_files
2 import compare_files
3
4
5 S101_FC_filepath = "..\\..\\documents\\S-101_FC_1.1.0_20230202.xml"
6 S101_DCEG_filepath = "..\\..\\documents\\S-101 Annex A_DCEG Edition 1.1.0_Final_Clean.docx"
7
8
9 dict_FC_S101 = read_files.read_FC_S101(S101_FC_filepath)
10 dict_DCEG_S101 = read_files.read_DCEG(S101_DCEG_filepath)
11
12 (omission_SimpleAttributes_DCEG,
13 omission_SimpleAttributes_definition_DCEG,
14 inconsistent_SimpleAttributes_definition,
15 omission_SimpleAttributes_alias_DCEG,
16 inconsistent_SimpleAttributes_alias,
17 inconsistent_SimpleAttributes_type,
18 ommission_SimpleAttributes_enum_DCEG,
19 ommission_SimpleAttributes_enum_FC,
20 inconsitent_SimpleAttributes_enum_label,
21 inconsitent_SimpleAttributes_enum_def,
22 inconsistent_SimpleAttributes_code_FC) = compare_files.compare_FC_DCEG_S101_SimpleAttributes(dict_FC_S101, dict_DCEG_S101)
```

- Compared items between FC and DCEG are the following :

- Associations
- SimpleAttributes
- ComplexAttributes
- FeatureTypes
- InformationTypes
- Roles

Category of Obstruction	Fish Haven	areas established by private interests, usually sport fishermen, to simulate natural reefs and wrecks that attract fish. the reefs are constructed by dumping assorted junk in areas which may be of very small extent or may stretch a considerable distance along a depth contour.	areas established by private interests, usually sport fishermen, to simulate natural reefs and wrecks that attract fish. the reefs are constructed by dumping assorted junk in areas which may be of very small extent or may stretch a considerable distance along a depth contour. also called fishery reefs.	For DCEG Edition 1.2.0: Have applied the change to the DCEG.
	Mangrove	one of several genera of tropical trees or shrubs which produce many prop roots and grow along low-lying coasts into shallow water.	one of several genera of tropical trees or shrubs which produce many prop roots and grow along low-lying coasts into shallow water.	For DCEG Edition 1.2.0: Have applied the change to the Registry. This should then translate through to the next version of the FC.
	Floating Production, Storage and Off-Loading Vessel	an offshore oil/gas facility consisting of a moored tanker/barge by which the product is extracted, stored and exported.	an offshore facility consisting of a moored tanker/barge by which the product is extracted, stored or exported.	For DCEG Edition 1.2.0: Have applied the change to the Registry. This should then translate through to the next version of the FC.
Category of Pilot Boarding Place	Pilot Comes Out from Shore	pilot embarks from a vessel or disembarks on a vessel which comes out from the shore on request.	pilot boards from a vessel or disembarks to a vessel which comes out from the shore on request.	For DCEG Edition 1.2.0: Have applied the change to the DCEG.
Category of Pipeline/Pipe	Outfall Pipe	a pipe (generally a sewer or drainage pipe) discharging into the sea or a river.	a pipe (generally a sewer or drainage pipe) discharging in to the sea or a river.	For DCEG Edition 1.2.0: Have applied the change to the DCEG.

- Most of the differences identified are "typo".

- Taken into account in version 1.2.0 of the S-101.

ADVANTAGES TO AUTOMATED CROSS-CHECKS

- Less chance to miss something

- Acceleration of a tedious work to do in "manual" mode

- Once the script has been developed, it might be possible with few changes to replay all the cross-checks for a new version of the FC or DCEG

- The script is probably adaptable for FC/DCEG cross-checks of other S-100 products
(no problem to share the script)

Questions ?