![JNCC - Joint Nature Conservation Committee]

JNCC Report
No: 516

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

Alberto Telletxea

June 2014

Summary

This document will comprise the feasibility study and recommendations for the integration of the Web Feature Services (WFS) network services into the Marine Protected Area (MPA) database (taking into account INSPIRE implementation. The source WFS have been derived from the Environment Agency's portal known as "Datashare" http://www.geostore.com/environment-agency/. We have focussed mainly on Marine Conservation Zones (MCZs)(one of the source layers available from Datashare). Various methods of linking to these WFS are evaluated. The results obtained will serve as a feasibility study for the potential future use of WFS to feed data to databases owned by JNCC.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

Contents

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

1   Introduction

This document will describe the technical procedures for the population of a SQL Server Spatial database with data provided by third party Web Feature Services (WFS), taking into account INSPIRE specifications. The results obtained will serve as a feasibility study for future use of WFS by JNCC to provide data to databases at JNCC. The WFS used for this report are published by the Environment Agency[1] on behalf of the Defra Network as part of the Datashare portal http://www.geostore.com/environment-agency/.

1.1   Data sources

The following information sources were considered when generating the information presented in the current report:

- Catalogue of WFS services available at the Environment Agency. The WFS can be found here:
  http://www.geostore.com/environment-agency/WebStore?xml=environment-agency/xml/inspireDataWfs.xml

- The WFS for Marine Conservation Zones. .
  Marine Conservation Zones (MCZs) are designated under the Marine and Coastal Access Act 2009. MCZs protect nationally important marine wildlife, habitats, geology and geomorphology in English inshore waters and offshore waters next to England, Wales and Northern Ireland. The lead organisation for this WFS is Natural England. The WFS can be found here:
  http://www.geostore.com/OGC/OGCInterface?SERVICE=WFS&UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS&VERSION=1.1.0&LC=8000000000000

- WFS: Marine Plan Areas
  There are 10 sub-areas corresponding to the 'English inshore and offshore marine planning regions'. The marine plan authority (the Secretary of State ifor the Environment, Food and Rural Affairs) is responsible for preparing marine plans for these 10 plan areas with responsibility for plan development delegated to The Marine Management Organisation. This dataset provides details of the spatial extent of each marine plan area as defined in the Marine and Coastal Access Act 2009. The WFS can be found here:
  http://www.geostore.com/OGC/OGCInterface?SERVICE=WFS&UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS&VERSION=1.1.0&LC=8000000000000

- WFS: Special Protected Areas (SPAs)
  A Special Protection Area (SPA) is classified under Directive 79/409 on the Conservation of Wild Birds. The WFS can be found here:
  http://www.geostore.com/OGC/OGCInterface?SERVICE=WFS&UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS&VERSION=1.1.0&LC=8000000000000

- WFS: RAMSAR England

---

[1] http://www.environment-agency.gov.uk

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

> A Ramsar site is the land listed as a Wetland of International Importance under the Convention on Wetlands of International Importance Especially as Waterfowl Habitat (the Ramsar Convention) 1973.
> http://www.geostore.com/OGC/OGCInterface?SERVICE=WFS&UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS&VERSION=1.1.0&LC=8000000000000

2   Software

This section will describe the different software packages that can be used for coupling WFS with an SQL Server Spatial (2008 R2-2012) database. WFS provides an HTTP hyperlink that allows remote access to geo-registered vector source data (points, lines and polygons) that can be displayed directly in your GIS software or other compatible applications.

The technologies that can be used are:

- QGIS (no license required): This can be used for data visualisation and conversion to an end user data format.
- ArcGIS Desktop – Data Interoperability Extension (license required): This can be used for data visualisation and conversion to an end user data format.
- OGC/OGR library (no license required): This can be used for WFS conversion into a SQL Server database.

Further information on these software is provided below:

2.1   QGIS (2.0.1 Dufour)

QGIS (previously known as "Quantum GIS") is a cross-platform free and open source desktop geographic information systems (GIS) application that provides data viewing, editing, and analysis capabilities. QGIS is distributed as a software package for installation purposes in MS Windows systems by the OSGeo4W project. OSGeo4W includes GDAL/OGR, GRASS, MapServer, OpenEV, uDig, and QGIS as well as many other packages (over 150 in number). This project is under the umbrella of the Open Source Geospatial Foundation[2].

2.1.1   QGIS visualisation of WFS and conversion to end user format (e.g. shapefiles):

To trial the use of QGIS to visualise WFS and convert the data to an end user format, the WFS for Marine Conservation Zones and Marine Plan Areas (MPAs) provided by the Environment Agency (see Data Sources above). The actions that should be taken are described below:

i. The Environment Agency uses the schema version 2.0.0 for the WFS, and so a plugin must be installed on QGIS from the menu bar by selecting **Plugins\Manage and install plugins**

The plugin is called **WFS 2.0 Client**

ii. From the QGIS menu bar go to **Web\ WFS 2.0 Client** (this is illustrated in the screen-grab below):

---

[2] http:/www.osgeo.org/

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)
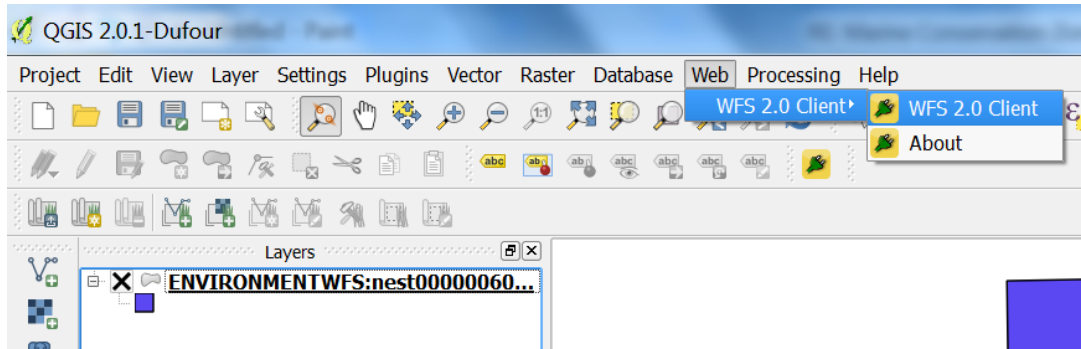

iii. Insert the connection properties required to access the WFS services catalogue and select the specific GML/feature of interest (e.g. marine related data) to be loaded for the visualisation.

This is the URL for openly available WFS provided by the Environment Agency:

http://www.geostore.com/OGC/OGCInterface?UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS&VERSION=2.0.0

Note: The WFS schema version to be used must to be 2.0.0.

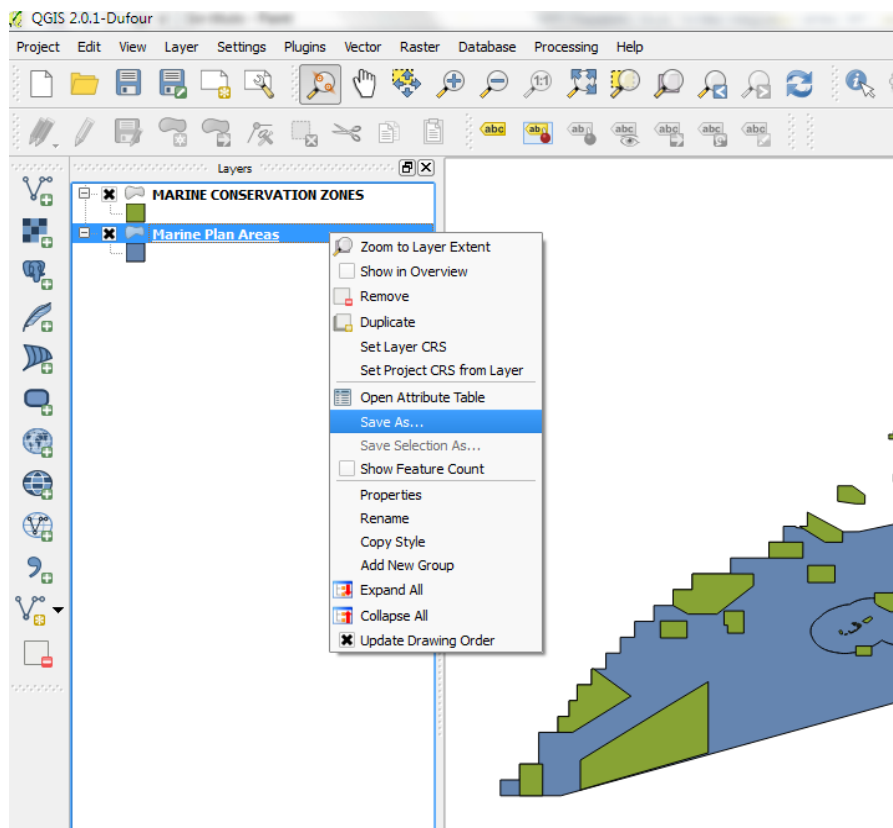The screen grab below shows the window for connecting to the WFS catalogue:
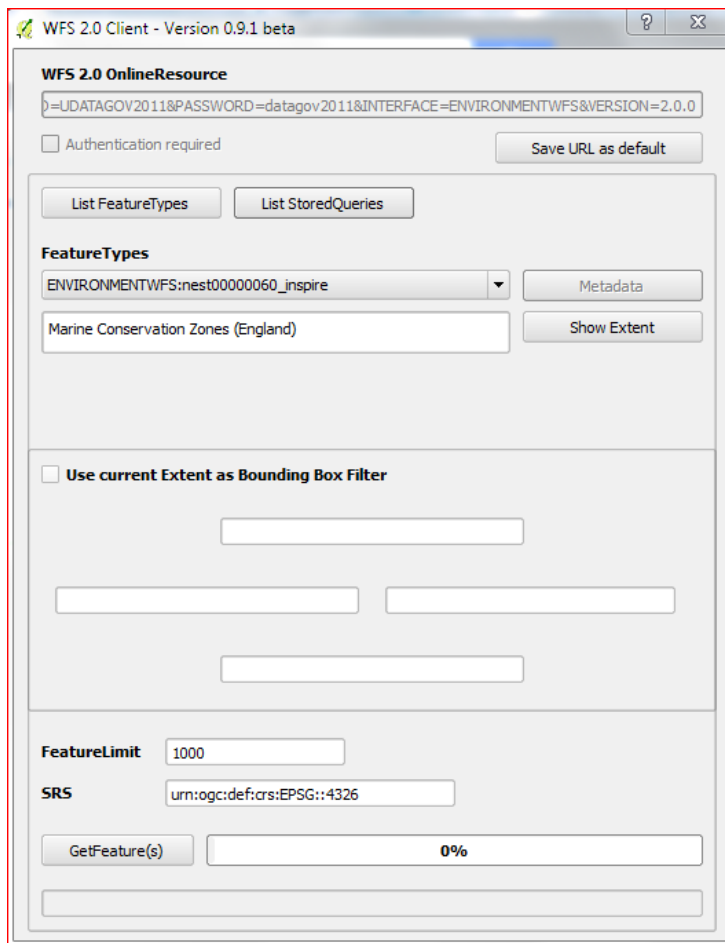
Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

iv. The GML (Geographic Mark up language – an open standard for holding spatial coordinates) features can then be loaded for visualisation of spatial and attributes information, as illustrated in the screen grab below
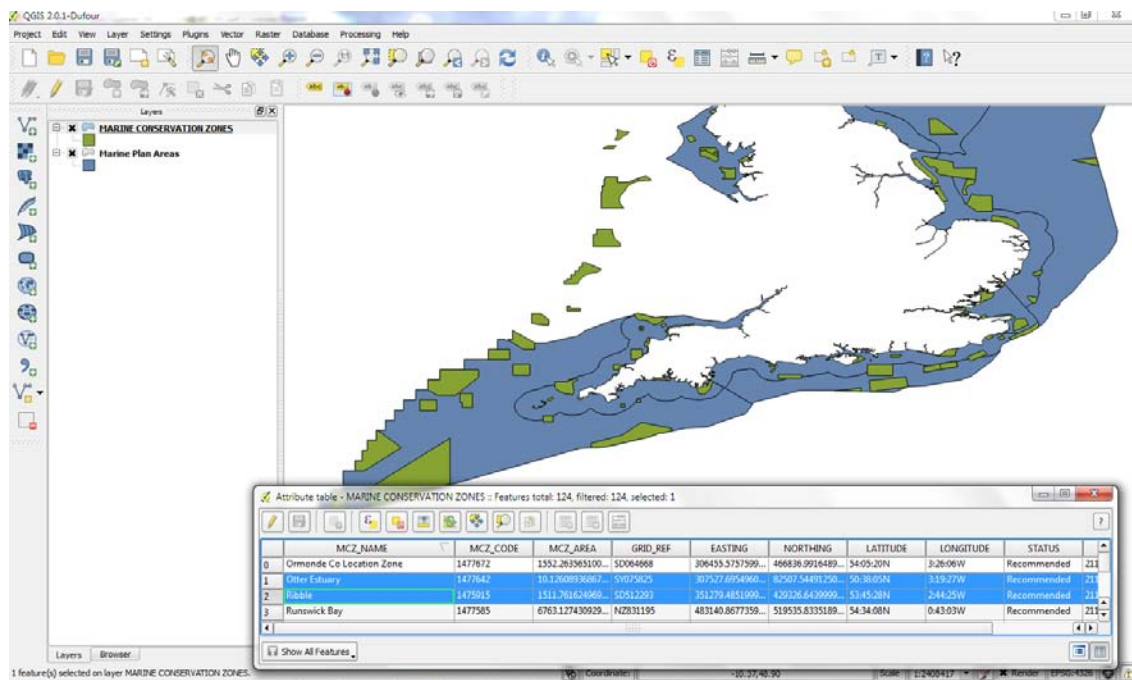
v. The features imported from WFS visualised can be exported as Shapefiles by right clicking on the feature and selecting "Save as", as illustrated below.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

The screen grab below shows the window for exporting the shapefile:
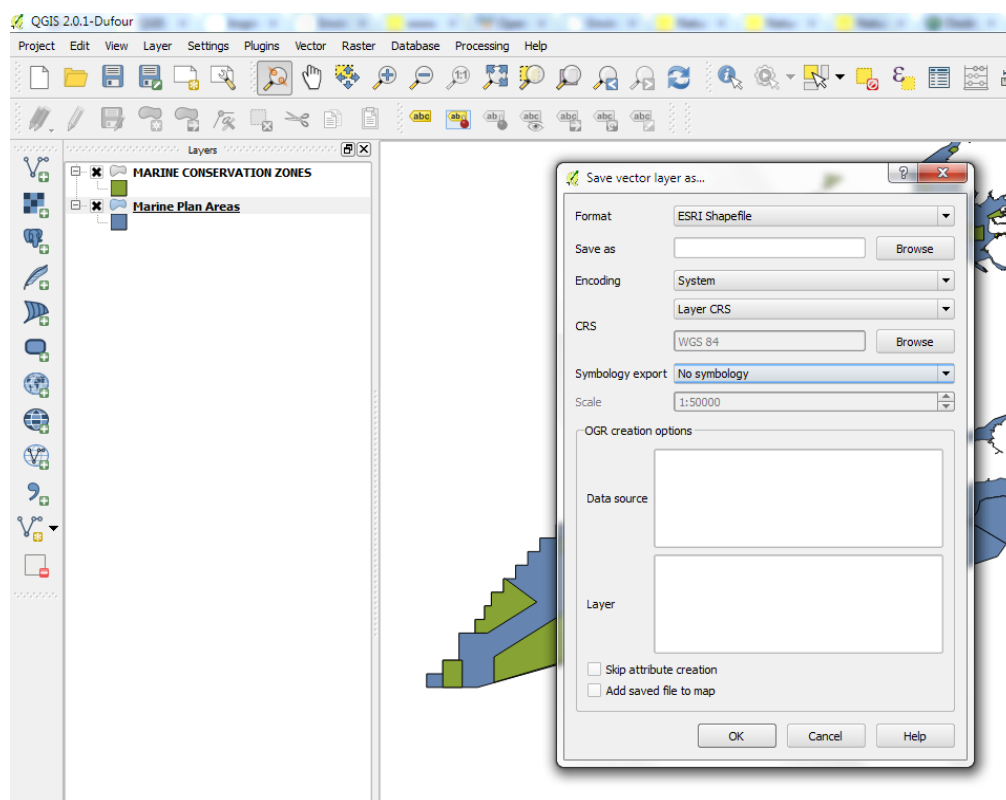
2.2   ESRI ArcGIS (ArcCatalog+Data Interoperability extension)

ESRI ArcGIS can also be used to access WFS services and export them to an end user format (e.g. Shapefiles, ESRI Geodatabases or SQL Server Spatial). For this feasibility study, the export to SQL Server Spatial was the format of interest, as JNCC wishes to explore storing spatial datasets in SQL Server 2008 or 2012 relational databases.

The advantage of these capabilities is that these ETL (Extract Transform and Load) processes can be automated using the ArcGIS Model Builder or Python script.

Note that the main issue encountered when using WFS was that the Environment Agency is currently using version 2.0 (version 1.0 is not currently supported) as the WFS schema. This version is only supported by the ArcGIS 10.2 Data Interoperability Tool.

2.2.1 Configure an access point from an ArcCatalog for visualisation in ArcMap

To configure an access point from an ArcCatalog for visualisation of WFS data in ArcMap, the following steps need to be undertaken:
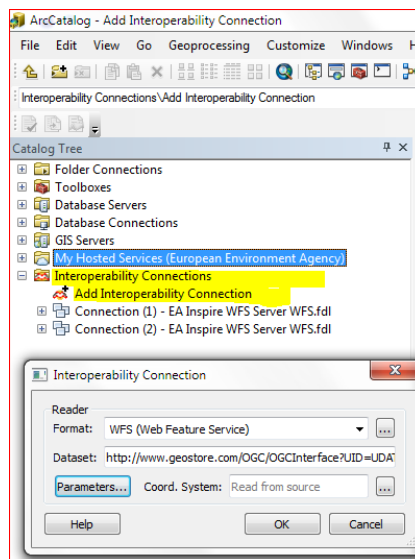
i.   Enable the use of the Data Interoperability extension, through the menu toolbar by selecting: Customize/Extensions…: This is illustrated in the screen grab, below.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)
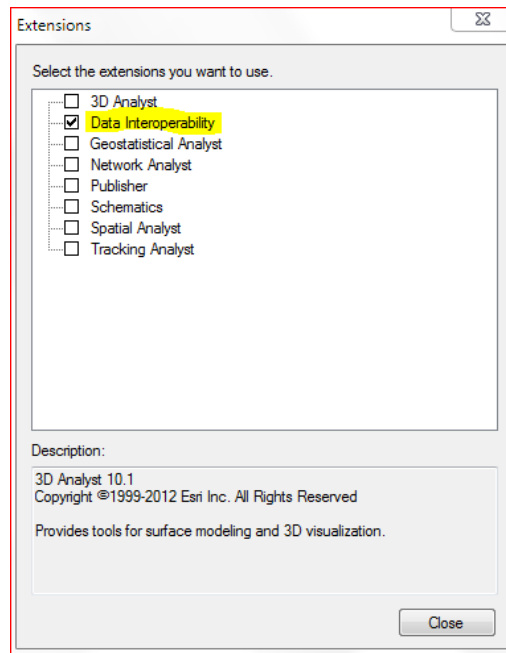
ii. Add an Interoperability connection and provide the following WFS URL (provided by the Environment Agency) as the dataset parameter:

http://www.geostore.com/OGC/OGCInterface?UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS&VERSION=2.0.0

This is illustrated in the screen grab below:

iii. Select the features of interest of the WFS (e.g. marine related features) as illustrated in the screen grab below.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)
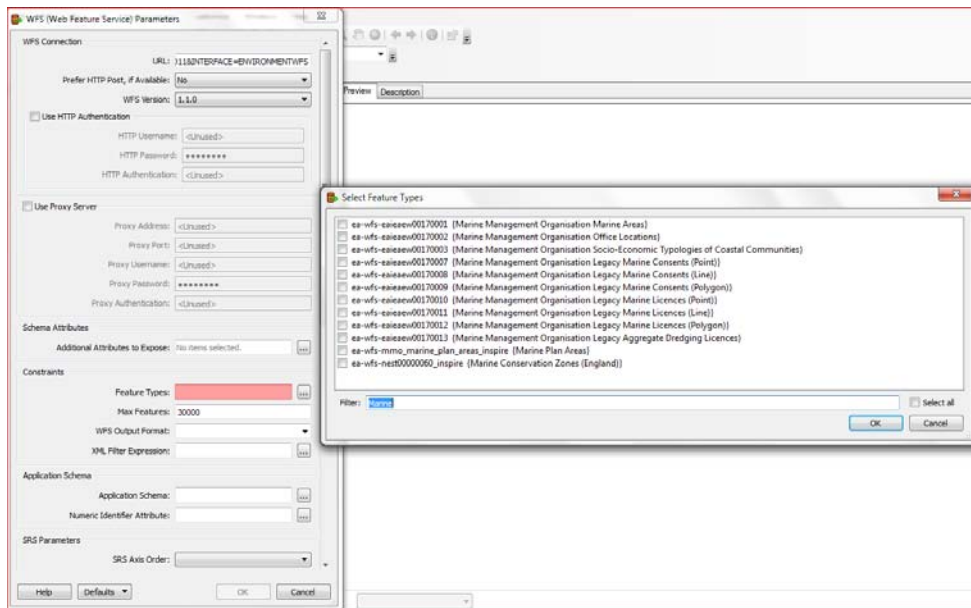

2.2.2 Export from WFS to SQL Server Spatial

As mentioned above, the main issue underpinning the use of WFS is the version of the WFS used. This is because only ArcGIS 10.2 can read WFS version 2.0 (previous versions cannot read WFS). The ArcGIS Data Interoperability tool allows you to export WFS data into an intermediate format (a File Geodatabase) and then export it to an SLQ Server database. See the screen grab below which illustrates the menu for accessing the import function in the Data Interoperability tools:
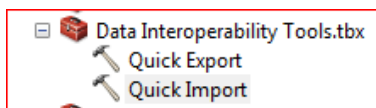
The recommendation is to use the ArcGIS Model Builder to establish the work flow. This has the benefit of allowing the process to be automated by a script in Python.

The diagram below shows the required dataflow in the Model Builder, using the Data Interoperability extension.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

The dataflow can be exported into a Python script for further automatisation. See below an extract of it:

```
# -*- coding: utf-8 -*-
# ---------------------------------------------------------------------
# export_import.py
# Created on: 2014-02-13 13:44:26.00000
#   (generated by ArcGIS/ModelBuilder)
# Description:
# ---------------------------------------------------------------------

# Import arcpy module
import arcpy
# Local variables:
Input_Dataset_Marine_Conservation_Zones = "WFS,http://www.geostore.com/OGC/OGCInterface?UID=UDATAGOV2011&PASSWORD=datagov2011&INTERFACE=ENVIRONMENTWFS,
\"RUNTIME_MACROS,\"\"PREFER_POST,No,PREFERRED_VERSION,1.1.0,USE_HTTP_AUTH,NO,HTTP_AUTH_USER,,HTTP_AUTH_PASSWORD,
,HTTP_AUTH_METHOD,Basic,USE_PROXY_SERVER,NO,HTTP_PROXY,null,HTTP_PROXY_PORT,null,HTTP_PROXY_USER,,HTTP_PROXY_PASSWORD,
,HTTP_PROXY_AUTH_METHOD,Basic,WFS_EXPOSE_FORMAT_ATTRS,,TABLELIST,\"\"\"\"\"\"\"\"\"\"ea-wfs-nest00000060_inspire
{Marine Conservation Zones (England)}\"\"\"\"\"\"\"\"\"\"\"\",MAX_RESULT_FEATURES,30000,OUTPUT_FORMAT,\"\"\"\"text/xml;
subtype=gml/3.1.1\"\"\"\",FILTER_EXPRESSION,,XSD_DOC,,FME_FEATURE_IDENTIFIER,,SRS_AXIS_ORDER,,USE_OLD_READER,NO,DISABLE_XML_NAMESPACE_PROCESSING,
NO,MAP_EMBEDDED_OBJECTS_AS,ATTRIBUTES,READ_PREDEFINED_GML_PROPERTIES,NO,MAP_GEOMETRY_COLUMNS,YES,MAP_COMPLEX_PROPERTIES_AS,\"\"\"\"Nested Attributes\"\"\"\"
,MAX_MULTI_LIST_LEVEL,,XML_FRAGMENTS_AS_DOCUMENTS,YES,FLATTEN_XML_FRAGMENTS,NO,FLATTEN_XML_FRAGMENTS_OPEN_LIST_BRACE,,FLATTEN_XML_FRAGMENTS_CLOSE_LIST_BRACE,
,FLATTEN_XML_FRAGMENTS_SEPARATOR,,USE_SEARCH_ENVELOPE,NO,SEARCH_ENVELOPE_MINX,0,SEARCH_ENVELOPE_MINY,0,SEARCH_ENVELOPE_MAXX,0,SEARCH_ENVELOPE_MAXY,0
,CLIP_TO_ENVELOPE,NO,BBOX_COORDINATE_SYSTEM,<Unused>,WFS_SERVICE_TITLE,\"\"\"\"EA Inspire WFS Server\"\"\"\",_MERGE_SCHEMAS,YES\"\",META_MACROS,\
"\"SourcePREFER_POST,No,SourcePREFERRED_VERSION,1.1.0,SourceUSE_HTTP_AUTH,NO,SourceHTTP_AUTH_USER,,SourceHTTP_AUTH_PASSWORD,,
SourceHTTP_AUTH_METHOD,Basic,SourceUSE_PROXY_SERVER,NO,SourceHTTP_PROXY,null,SourceHTTP_PROXY_PORT,null,SourceHTTP_PROXY_USER,,
SourceHTTP_PROXY_PASSWORD,,SourceHTTP_PROXY_AUTH_METHOD,Basic,SourceWFS_EXPOSE_FORMAT_ATTRS,,SourceMAX_RESULT_FEATURES,30000,
SourceOUTPUT_FORMAT,\"\"\"\"text/xml; subtype=gml/3.1.1\"\"\"\",SourceFILTER_EXPRESSION,,SourceXSD_DOC,,SourceFME_FEATURE_IDENTIFIER,,
```

2.3   Use of GDAL/OGR for export to SQL Spatial

This section explores the use of geospatial data formats that are released under an X/MIT style Open Source license by the Open Source Geospatial Foundation. This library is continuously being extended and provides a development environment for the automatisation of data acquisition, integration and data transfer procedures that can be built using a variety of programming languages including:

- Python
- Java
- C# / .Net

See below the list of formats that can be managed by this library.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)


```
C:\Users\atelletxea\Desktop>ogr2ogr --formats
Supported Formats:
  -> "FileGDB" (read/write)
  -> "OCI" (read/write)
  -> "ESRI Shapefile" (read/write)
  -> "MapInfo File" (read/write)
  -> "UK .NTF" (readonly)
  -> "SDTS" (readonly)
  -> "TIGER" (read/write)
  -> "S57" (read/write)
  -> "DGN" (read/write)
  -> "VRT" (readonly)
  -> "REC" (readonly)
  -> "Memory" (read/write)
  -> "BNA" (read/write)
  -> "CSV" (read/write)
  -> "NAS" (readonly)
  -> "GML" (read/write)
  -> "GPX" (read/write)
  -> "KML" (read/write)
  -> "GeoJSON" (read/write)
  -> "GMT" (read/write)
  -> "SQLite" (read/write)
  -> "ODBC" (read/write)
  -> "PGeo" (readonly)
  -> "MSSQLSpatial" (read/write)
  -> "OGDI" (readonly)
  -> "PostgreSQL" (read/write)
  -> "MySQL" (read/write)
  -> "PCIDSK" (read/write)
  -> "XPlane" (readonly)
  -> "AVCBin" (readonly)
  -> "AVCE00" (readonly)
  -> "DXF" (read/write)
  -> "Geoconcept" (read/write)
  -> "GeoRSS" (read/write)
  -> "GPSTrackMaker" (read/write)
  -> "VFK" (readonly)
  -> "PGDump" (read/write)
  -> "OSM" (readonly)
  -> "GPSBabel" (read/write)
  -> "SUA" (readonly)
  -> "OpenAir" (readonly)
  -> "PDS" (readonly)
  -> "WFS" (readonly)
  -> "HTF" (readonly)
  -> "AeronavFAA" (readonly)
  -> "Geomedia" (readonly)
  -> "EDIGEO" (readonly)
  -> "GFT" (read/write)
  -> "SVG" (readonly)
  -> "CouchDB" (read/write)
  -> "Idrisi" (readonly)
  -> "ARCGEN" (readonly)
  -> "SEGUKOOA" (readonly)
  -> "SEGY" (readonly)
  -> "XLS" (readonly)
  -> "ODS" (read/write)
  -> "XLSX" (read/write)
  -> "ElasticSearch" (read/write)
  -> "PDF" (read/write)
```

GDAL - Geospatial Data Abstraction Library

The use of scripts using this library allows the user to transform data from WFS into MSSQLSpatial. A sample of transformation using this library is:
ogr2ogr -overwrite -f MSSQLSpatial
"MSSQL:server=oiz;database=EJN_TEST;UID=sa;PWD=hellfire" MarineConservationZones WFS:
"http://www.geostore.com/OGC/OGCInterface?UID=UDATAGOV2011&PASSWORD=datag ov2011&INTERFACE=ENVIRONMENTWFS&VERSION=2.0.0" ea-wfs-nest00000060_inspire

3   WFS compliancy with INSPIRE metadata specifications

We have looked at the metadata for Marine Conservation Zones.
The INSPIRE metadata can be requested through a web service call under OGL (Open Government License) licensing. See below an extract of the INSPIRE metadata XML provided for Marine Conservation Zones.

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

The metadata catalogue service for Natural England INSPIRE metadata is accessible through the web site at: http://data.gov.uk/data/search
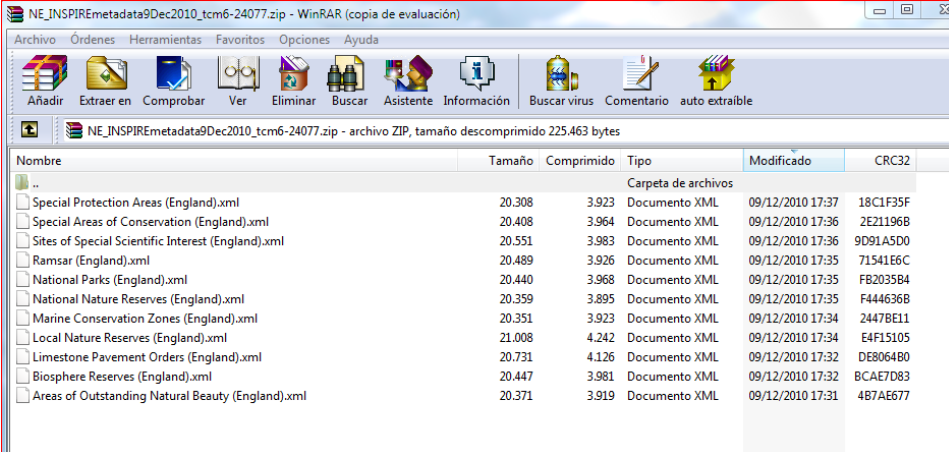
See below a screen grab of the website:
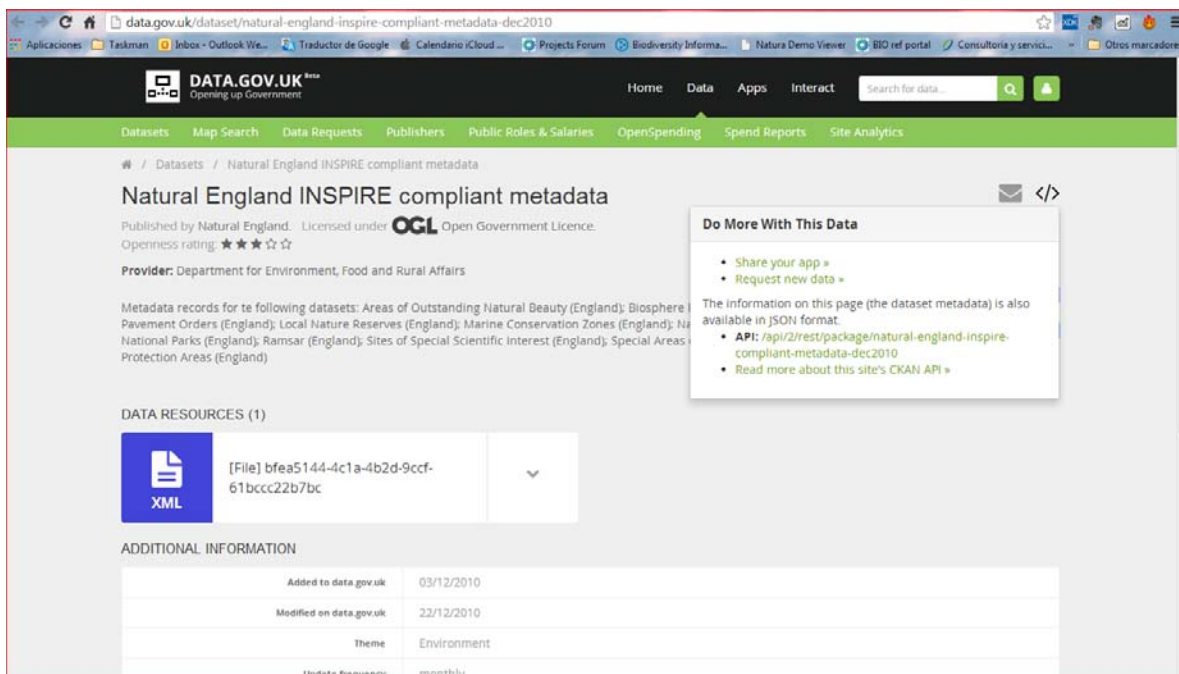
The metadata XML files can be downloaded under request of a web service, see below the URL request to obtain Natural England INSPIRE metadata:
http://data.gov.uk/api/2/rest/package/natural-england-inspire-compliant-metadata-dec2010

See below a capture of the files that can be downloaded:

Feasibility study and recommendations for the integration of the WFS network services into MPA database (taking into account INSPIRE implementation)

The metadata is not fully compatible with the INSPIRE profile defined for Protected Areas (The INSPIRE Guidelines on Data Specification on Protected Sites is available at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.0.pdf)

For example the XML for the MD_Keywords attribute is:
<gmd:descriptiveKeywords>
*   <gmd:MD_Keywords>*
*   <gmd:keyword>*
*   <gco:CharacterString>Protected sites</gco:CharacterString>*
*   </gmd:keyword>*

The Marine Conservation Zones XML metadata has been validated against the JRC INSPIRE validator (see: http://inspire-geoportal.ec.europa.eu/validator2/) and it has given just one inconformity. See below the response given by the JRC validator v2:

Inspire Validation Issues
The metadata element "Conformity" is missing, empty or incomplete but it is required (see documentation)
Relevant documentation:
Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata

There are some attributes specific to Protected Sites that are not included as the:
- Attribute: siteDesignation
    - Value type: DesignationType
    - Definition: The designation (type) of Protected Site.
    - Description: At least one designation is required, but designations may be available using a number of different designation schemes, all of which can be accommodated in the DesignationSchemeValue codelist.
    - Multiplicity: 1..*

A very important aspect to be mentioned is that the metadata has not been encountered together with the WFS. In principle the metadata should be delivered in the same resource as WFS.

4   Conclusions

The exercise of looking at the capabilities of integrating data from WFS published by the Environment Agency has proved that this way of providing information is still not mature enough to design robust routines of data management through Extract Transform and Load processes (ETL).

These are the main obstacles that we see in these technologies:
- The GIS software available to read and export from WFS is behind in terms of the different versions supported by the OGC for WFS (version 1.0 and 2.0).
- The IT server infrastructures of the orgnisations that publish WFS do not always provide high availability and in many cases are not functioning effectively or do not give the required performance the interchange of geometry features over the Internet needs.
- The metadata is not always together with the spatial data and other resources are required to be explored to acquire the metadata.
- Not all the WFS versions are compliant with INSPIRE.

These are the benefits that have been identified with these technologies:
- WFS is a good format to interchange geospatial data on the Internet through machine to machine routines, avoiding manual jobs.