GML application schema made easy in GDAL/OGR and QGIS

GDAL/OGR GMLAS driver

Francois Prunayre, Even Rouault, Christian Ansorge, Sylvain Grellet, Hugo Mercier, Paul van Genuchten

GDAL/OGR GMLAS driver: intro

- **GDAL?** Geospatial Data Abstraction Library. The swiss army knife for geospatial.
- Read/write access to more than 200 (mainly) geospatial formats and protocols.
- Command line utilities, C/C++/Python/… API
- Widely used (FOSS & proprietary): GRASS, MapServer, Mapnik, **QGIS**, gvSIG, PostGIS, OTB, SAGA, FME, ArcGIS, Google Earth…
- **GMLAS** ? GML + AS = GML Application Schema
- New driver of GDAL 2.2 (released in May 2017)
- Funded through E.E.A (through Copernicus) and BRGM

GDAL/OGR GMLAS driver: intro

- Read XML/GML documents following (almost) arbitrary XML schemas (Inspire, GroundWaterML2, GeoSCIML, etc.. compatible, but no hard-coded knowledge of them)
- Analyze the schema to create a relational model, that can be consumed by target databases (PostGIS, Spatialite, etc…)
- Arbitrary big documents can be read and converted (potentially GB)
- Tunable behaviour

- Write side: can regenerate a GML/XML from an imported database that has been modified

GDAL/OGR GMLAS driver: GML driver vs GMLAS driver

- Before GDAL 2.2, there was already a GML driver
  - Only/mostly simple feature compliant
  - With various hacks added over time to better handle some particular complex feature schemas
  - Can work without an explicit schema
  - Still the default

- New GMLAS driver uses the “GMLAS:filename” syntax
- Documentation page: http://www.gdal.org/drv_gmlas

GDAL/OGR GMLAS driver: example of a complex feature doc 1/2

```
<wfs:FeatureCollection ...>
  <wfs:member>
    <ef:EnvironmentalMonitoringFacility gml:id="Piezometre.06512X0037.STREMY.2">
        <gml:description>Water well from national BSS (Banque du Sous-Sol) Data database. Piezometer monitoring ground water level</gml:description>
        <gml:identifier codeSpace="http://www.ietf.org/rfc/rfc2616">http://ressource.brgm-rec.fr/data/Piezometre/06512X0037/STREMY.2</gml:identifier>
        <ef:inspireId>
            <base:Identifier>
                <base:localId>Piezometre/06512X0037/STREMY.2</base:localId>
                <base:namespace>http://ressource.brgm-rec.fr/data</base:namespace>
                <base:versionId/>
            </base:Identifier>
        </ef:inspireId>
        <ef:name>Piézomètre de St-Rémy - 01</ef:name>
        <ef:additionalDescription/>
        <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codelist/MediaValue/water" xlink:title="water"/>
        <ef:legalBackground/>
        <ef:geometry>
            <gml:Point gml:id="Piezometre.geom.2.06512X0037-STREMY" srsDimension="2" srsName="urn:ogc:def:crs:EPSG::4326">
                <gml:pos>46.1909541655103 5.18713262971692</gml:pos>
            </gml:Point>
        </ef:geometry>
        <ef:onlineResource>http://fichebsseau.brgm.fr/bss_eau/fiche.jsf?code=06512X0037/STREMY</ef:onlineResource>
        <ef:purpose xlink:href="http://www.sandre.eaufrance.fr/?urn=urn:sandre:donnees:148::CdElement:2:::referentiel:3.1:xml" xlink:title="Ground water level measurement"/>
        <ef:broader/>
        <ef:supersedes/>
        <ef:supersededBy/>
        <ef:reportedTo xsi:nil="true" nilReason="http://inspire.ec.europa.eu/codelist/VoidReasonValue/Unpopulated"/>
        <ef:hasObservation
xlink:href="http://ressource.brgm-rec.fr/obs/RawOfferingPiezo/06512X0037/STREMY.2&amp;responseFormat=http://www.opengis.net/waterml/2.0&amp;temporalFilter=om%3AphenomenonTime%2Clatest"
xlink:title="Latest value (WaterML 2.0 format): Raw groundwater level measurement from piezometer 06512X0037/STREMY.2"/>
        <ef:hasObservation   xlink:href="http://ressource.brgm-rec.fr/obs/RawOfferingPiezo/06512X0037/STREMY.2&amp;responseFormat=application/json&amp;temporalFilter=om%3AphenomenonTime%2Clatest"
xlink:title="Latest value (JSON format): Raw groundwater level measurement from piezometer 06512X0037/STREMY.2"/>
        <ef:hasObservation  xlink:href="http://ressource.brgm-rec.fr/obs/RawOfferingPiezo/06512X0037/STREMY.2&amp;responseFormat=http://www.opengis.net/waterml/2.0"  xlink:title="All  observations  (WaterML
2.0 format): Raw groundwater level measurement from piezometer 06512X0037/STREMY.2"/>
        <ef:hasObservation xlink:href="http://ressource.brgm-rec.fr/obs/RawOfferingPiezo/06512X0037/STREMY.2&amp;MergeObservationsIntoDataArray=true" xlink:title="SWEArrayObservation: Raw groundwater
level measurement from piezometer 06512X0037/STREMY.2"/>
        <ef:involvedIn/>
        <ef:representativePoint>
            <gml:Point gml:id="Piezometre.reppoint.2.06512X0037-STREMY" srsDimension="2" srsName="urn:ogc:def:crs:EPSG::4326">
                <gml:pos>46.1909541655103 5.18713262971692</gml:pos>
            </gml:Point>
        </ef:representativePoint>
```

GDAL/OGR GMLAS driver: example of a complex feature doc 2/2

```
            <ef:measurementRegime/>
            <ef:mobile>false</ef:mobile>
            <ef:resultAcquisitionSource xlink:href="http://inspire.ec.europa.eu/codelist/ResultAcquisitionSourceValue/inSitu/" xlink:title="in-situ"/>
            <ef:specialisedEMFType xlink:href="http://www.sandre.eaufrance.fr/urn.php?urn=urn:sandre:dictionnaire:PTE::entite:Piezometre:ressource:2.1:::html" xlink:title="Piezometre"/>
            <ef:operationalActivityPeriod>
            <ef:OperationalActivityPeriod gml:id="Piezometre.OperationalActivityPeriod.2.06512X0037-STREMY">
                        <ef:activityTime>
                                    <gml:TimePeriod gml:id="TimePeriod.2.225196">
                                                <gml:beginPosition>1977-10-08T23:00:00Z</gml:beginPosition>
                                                <gml:endPosition>2014-10-14T06:00:00Z</gml:endPosition>
                                    </gml:TimePeriod>
                        </ef:activityTime>
            </ef:OperationalActivityPeriod>
            </ef:operationalActivityPeriod>
            <ef:relatedTo/>
            <ef:belongsTo>
                        <ef:NetworkFacility gml:id="ef_networkfacility_2.fid--28738dc3_15d17368ff8_-2e65">
                                    <ef:linkingTime>
                                                <gml:TimePeriod gml:id="TimePeriod.0000000029.06512X0037-STREMY">
                                                            <gml:beginPosition>1977-09-30T23:00:00Z</gml:beginPosition>
                                                            <gml:endPosition/>
                                                </gml:TimePeriod>
                                    </ef:linkingTime>
                                    <ef:belongsTo xlink:href="http://ressource.brgm-rec.fr/data/DispositifCollecte/0000000029" xlink:title="Réseau patrimonial national de suivi quantitatif des eaux souterraines(RNESP)"/>
                                    <ef:contains/>
                        </ef:NetworkFacility>
            </ef:belongsTo>
            <ef:belongsTo>
                        <ef:NetworkFacility gml:id="ef_networkfacility_2.fid--28738dc3_15d17368ff8_-2e64">
                                    <ef:linkingTime>
                                                <gml:TimePeriod gml:id="TimePeriod.0600000004.06512X0037-STREMY">
                                                            <gml:beginPosition>1977-09-30T23:00:00Z</gml:beginPosition>
                                                            <gml:endPosition/>
                                                </gml:TimePeriod>
                                    </ef:linkingTime>
                                    <ef:belongsTo xlink:href="http://ressource.brgm-rec.fr/data/DispositifCollecte/0600000004" xlink:title="Réseau de suivi quantitatif des eaux souterraines du bassin Rhône-Méditerranée(RBESOUPRMC)"/>
                                    <ef:contains/>
                        </ef:NetworkFacility>
            </ef:belongsTo>
      </ef:EnvironmentalMonitoringFacility>
   </wfs:member>
</wfs:FeatureCollection>
```

GDAL/OGR GMLAS driver: ogrinfo

$ ogrinfo -ro **GMLAS:**Piezometre.06512X0037.STREMY.2.gml

```
INFO: Open of `GMLAS:Piezometre.06512X0037.STREMY.2.gml'
      using driver `GMLAS' successful.
1: observingcapability1 (None)
2: observingcapability_metadataproperty (None)
3: observingcapability_name (None)
4: environmentalmonitoringactivity (None)
5: environmentalmonitoringactivity_metadataproperty (None)
6: environmentalmonitoringactivity_name (None)
7: environmentalmonitoringactivity_onlineresource (None
[...]
611: ds_productionseries_seriesmetadata_seriesmetadata (None)
612: ds_initiative_composedof_composedof (None)
613: ds_initiative_seriesmetadata_seriesmetadata (None)
614: dynamifeaturcollec_dynamimember_dynamifeature_dynamicfeature (None)
615: dynamfeatucolle_dynammembe_dynamfeatu_dynamifeaturcollection (None)
```

GDAL/OGR GMLAS driver: ogrinfo

```
$ ogrinfo -ro GMLAS:Piezometre.06512X0037.STREMY.2.gml -oo REMOVE_UNUSED_LAYERS=YES
1: environmentalmonitoringfacility (Unknown (any), Point)
2: environmentalmonitoringfacility_ef_name (None)
3: environmentalmonitoringfacility_mediamonitored (None)
4: environmentalmonitoringfacility_onlineresource (None)
5: environmentalmonitoringfacility_purpose (None)
6: environmentalmonitoringfacility_supersedes (None)
7: environmentalmonitoringfacility_supersededby (None)
8: environmentalmonitoringfacility_resultacquisitionsource (None)
9: environmentalmonitoringfacility_operationalactivityperiod (None)
10: environmentalmonitoringfacility_relatedto (None)
11: environmentalmonitoringfacility_belongsto (None)
12: operationalactivityperiod (None)
13: timeperiod (None)
14: ef_inspireid (None)
15: legalbackground (None)
16: broader (None)
17: reportedto (None)
18: hasobservation (None)
19: involvedin (None)
20: networkfacility (None)
21: environmenmonitoringfacility_legalbackground_legalbackground (None)
22: environmentalmonitoringfacility_reportedto_reportedto (None)
23: environmentamonitoringfacility_hasobservation_hasobservation (None)
24: environmentalmonitoringfacility_involvedin_involvedin (None)
```

GDAL/OGR GMLAS driver: convert from GML to PostGIS

```
$ ogrinfo “PG:dbname=mydb” -sql “CREATE SCHEMA piezometre”
$ ogr2ogr -f PostgreSQL \
                                “PG:dbname=mydb      active_schema=piezometre”      \
    GMLAS:Piezometre.06512X0037.STREMY.2.gml \
    -oo EXPOSE_METADATA_LAYERS=YES \
    -oo REMOVE_UNUSED_LAYERS=YES \
    -nlt  CONVERT_TO_LINEAR -forceNullable
```

All other ogr2ogr options are possible, among:

- -overwrite, -append
- -spat xmin ymin xmax ymax
- reprojection: -t_srs EPSG:XXXX

GDAL/OGR GMLAS driver: convert from GML to Spatialite

```
$ ogr2ogr -f SQLite -dsco SPATIALITE=YES \
        my.db \
        GMLAS:Piezometre.06512X0037.STREMY.2.gml \
        -oo EXPOSE_METADATA_LAYERS=YES \
       -nlt  CONVERT_TO_LINEAR -forceNullable
```

GDAL/OGR GMLAS driver: convert from GML to Spatialite with Python GDAL

```
$ python
from osgeo import gdal
gdal.VectorTranslate('my.db', 'GMLAS:Piezometre.06512X0037.STREMY.2.gml',
        options = '-f SQLite -dsco SPATIALITE=YES ' +
                '-oo EXPOSE_METADATA_LAYERS=YES ' +
                '-nlt  CONVERT_TO_LINEAR -forceNullable'
```

GDAL/OGR GMLAS driver: convert to GML from PostgreSQL/Spatialite

$ ogr2ogr -f GMLAS out.xml my.db

$ ogr2ogr -f GMLAS out.xml “PG:dbname=mydb active_schema=piezometre”

Requires the OGR metadata layers to be stored in the database

Options limited to: -t_srs, -spat, -where, -dsco, subset of layers

GDAL/OGR GMLAS driver: configuration

Default configuration in GDAL configuration file:
https://svn.osgeo.org/gdal/trunk/gdal/data/gmlasconf.xml

What can be tuned (non exhaustive list):

- whether remote schemas should be downloaded. Enabled by default.
- whether the local cache of schemas is enabled. Enabled by default.
- the path of the local cache. By default, $HOME/.gdal/gmlas_xsd_cache
- whether validation of the document against the schemas should be enabled. Disabled by default.
- whether validation error should cause dataset opening to fail. Disabled by default.
- whether the metadata layers should be exposed by default. Disabled by default.
- whether layers and fields that are not used in the XML document should be removed. Disable by default.
- whether OGR array data types can be used. Enabled by default.
- whether the XML definition of the GML geometry should be reported as a OGR string field. Disabled by default.
- whether only XML elements that derive from gml:_Feature or gml:AbstractFeature should be considered in the initial pass of the schema building, when at least one element in the schemas derive from them. Enabled by default.
- several rules to configure if and how xlink:href should be resolved.
- a definition of XPaths of elements and attributes that must be ignored, so as to lighten the number of OGR layers and fields.

GDAL/OGR GMLAS driver: vocabulary resolution

Typically http://inspire.ec.europa.eu/codelist

Requires uncommenting <URLSpecificResolution> in GMLAS configuration file

```
$ ogrinfo GMLAS:Piezometre.06512X0037.STREMY.2.gml -al \
    -oo CONFIG_FILE=gmlasconf_with_inspire_codelist.xml  \
    environmentalmonitoringfacility_resultacquisitionsource
```

OGRFeature(environmentalmonitoringfacility_resultacquisitionsource):1

ogr_pkid (String) = Piezometre.06512X0037.STREMY.2_resultAcquisitionSource_1

parent_id (String) = Piezometre.06512X0037.STREMY.2

owns (Integer(Boolean)) = 0

href (String) = http://inspire.ec.europa.eu/codelist/ResultAcquisitionSourceValue/inSitu/

href_name (String) = in-situ

href_definition (String) = The FeatureOfInterest is a sampling feature which is co-located with the ultimate FeatureOfInterest (i.e. the sampledFeature).

title (String) = in-situ

GDAL/OGR GMLAS driver: strengths

- The GMLAS reader is robust to various non-conformities in the input XML/GML document
  - Missing requirement elements/attributes
  - Extra/unexpected elements/attributes (→ ignored)

  But the input schemas must be valid, and the document at least well formed as XML

- Designed to process arbitrarily large datasets with few memory requirements

GDAL/OGR GMLAS driver: limitations

- Some models are loosely typed (xs:any elements) and require addition of manual constraints. No use of schematrons
- Foreign keys / relational model not yet in OGR abstraction. No automatic creation of databases relationships (created by the QGIS plugin for the PostgreSQL case)
- Some issues (non schema conformity) on the write side of the driver depending on models / documents regarding optional / null elements
- Potential performance issues on the write side on large input databases