zeepsoap

February 14, 2024

0.1 Notes

0.1.1 XML

More verbose than JSON. Allows mixing domains. Elements with attributes.

JSON can represent anything XML can represent, but XML is “cleaner” in complex cases. But still verbose.

Also XML has well standardized “XPath” mini-language for extracting from hierarchy, JSON alternatives more varied.

0.1.2 SOAP

Simple Object Access Protocol => “simple” in title, i.e. not simple. Part of an ecosystem:

WSDL Web Service Discovery Language(?)

RPC Remote Procedure Call

XSD XML Document Schema specification, defines types, “positive integers”, and complex types, “ParameterType” (string plus name plus optional type and encoding).

XSLT XML Document transformation language (not seen here)

Mostly replaced by REST HTTP / JSON APIs,

0.1.3 Issues

No R SOAP client? Can use templates from Python version. - R XML lib. to build / disassemble XML docs. - RCurl or similar to send / receive

One year / analyte per query, given XML overhead, kind of low performance.

Tokens last ~1hr, would need something else for distribution.

Also - 1hr long enough for all analytes / years? Probably.

[ ]:

```
# Python SOAP library
%pip install zeep
```

[29]:

```
import re
def show_response(response):
    """For debugging, shows request / response with redacted token."""
    print("\n".join(f"{k}: {v}" for k, v in response.request.headers.items()))
    print()
    print(re.sub("(?<=csm:.....)[-_A-Za-z0-9.]+", '...', response.request.body.
 ↪decode("utf8")))
    print(response.content.decode("utf8"))
```

```
[30]: from zeep import Client

      # client = Client('https://cdxnodengn.epa.gov/cdx-enws10/services/
       ↪NetworkNodePortType_V10?wsdl')
      client = Client("https://cdxnodengn.epa.gov/ngn-enws20/services/
       ↪NetworkNode2Service?wsdl")
```

```
[31]: # show functions / attributes of client
      [i for i in dir(client) if not i.startswith("_")]
```

```
[31]: ['bind',
       'create_message',
       'create_service',
       'get_element',
       'get_type',
       'namespaces',
       'plugins',
       'service',
       'set_default_soapheaders',
       'set_ns_prefix',
       'settings',
       'transport',
       'type_factory',
       'wsdl',
       'wsse']
```

```
[32]: client.namespaces
```

```
[32]: {'xsd': 'http://www.w3.org/2001/XMLSchema',
       'ns0': 'http://www.exchangenetwork.net/schema/node/2',
       'ns1': 'http://www.w3.org/2005/05/xmlmime'}
```

```
[33]: # show functions / attributes of client.service
      [i for i in dir(client.service) if not i.startswith("_")]
```

```
[33]: ['Authenticate',
       'Download',
       'Execute',
       'GetServices',
       'GetStatus',
       'NodePing',
       'Notify',
       'Query',
       'Solicit',
       'Submit']
```

[34]:
```python
# read password without displaying it
import getpass
cred = getpass.getpass()
```

```
········
```

[42]:
```python
# Valid authenticationMethod values are: [password, digest, certificate, xkms,␣
 ↪hmac]
userId = "TERRYNBROWN"
# with client.settings(raw_response=True):
#     pass  # this code used to extract raw result for template
# This is a Remote Procedure Call (RPC) calling a function on the remote system
token = client.service.Authenticate(userId=userId, credential=cred,␣
 ↪authenticationMethod='password')
token
```

[42]:
```
'csm:eyJjdHkiOiJKV1QiLCJlbmMiOiJBMTI4Q0JDLUhTMjU2IiwiYWxnIjoiZGlyIn0..Oig7zPfECi
xFCgwRco_kxw.6hbXfmPieXWYzco0lvw-PkdL9BAtcEFdD5DwT_0Gdmk5IsY9l3CuVxirwdSQyY3Quo8
3cAcH9nIJjSCJPJE8iWqMP6p7s3xFnEarrsa5VVIJpvtnM4aAXw_Gq8I_-61OG-cDkQmHzeJsWnEoMK4
cbib6Sabtw7q2RGvV3SNXOAif05B2m1hi3Kfm8SgjW8TT8iJ4it_SNoytebPD2h2EC98zkXd9nh6vtTD
hbuepfDWu6KECDZNCv8Yh8QXPDWV5lc9bf_pnCGlXgGYq13stdlCjGZyu_6p9eFFL3fv6RPngOeD2pfb
gePFc-
1TY5TJjayLKRFxEMoV1TKvcwHIM0YTF3rkKM1iToZ4Jivm9DcQizJWmBcKLO_lGRIW48N4e.i4LqPZMY
71HCm0OCQvYAcg'
```

[ ]:
```python
show_response(token)
```

Request, note `SOAPAction` in header, maybe required.

```
User-Agent: Zeep/4.2.1 (www.python-zeep.org)
Accept-Encoding: gzip, deflate, br, zstd
Accept: */*
Connection: keep-alive
SOAPAction: ""
Content-Type: application/soap+xml; charset=utf-8; action=""
Cookie: JSESSIONID=63751690DC5CDF49F835014AE28C004E; cdx-prod-coreservices=1707921987.09.32.28
Content-Length: 495

<?xml version='1.0' encoding='utf-8'?>
<soap-env:Envelope xmlns:soap-env="http://www.w3.org/2003/05/soap-envelope">
  <soap-env:Body>
    <ns0:Authenticate xmlns:ns0="http://www.exchangenetwork.net/schema/node/2">
      <ns0:userId>{{USERNAME}}</ns0:userId>
      <ns0:credential>{{PASSWORD}}</ns0:credential>
      <ns0:domain xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:nil="true"/>
      <ns0:authenticationMethod>password</ns0:authenticationMethod>
    </ns0:Authenticate>
  </soap-env:Body>
</soap-env:Envelope>
```

Response - “multi-part/mime” (Multipurpose Internet Mail Extensions).

```
--uuid:8ce7a3c2-f555-4561-aa2f-7d9d439a4a5c
Content-Type: application/xop+xml; charset=UTF-8; type="application/soap+xml"
Content-Transfer-Encoding: binary
Content-ID: <root.message@cxf.apache.org>

<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">
  <soap:Body>
    <AuthenticateResponse xmlns="http://www.exchangenetwork.net/schema/node/2" xmlns:xmime="ht
      <securityToken>      csm:eyJjdHkiOiJKV1QiLCJlbmMiOiJBMTI4Q0JDLUhTMjU2IiwiYWxnIjoiZGlyIn0
g58rusxSJyKIVUFHOIP9AlbjBmQnHYN34fqxg7_7miEDFgJb-PwN3RFW4-u7ckexdvWG7XXQrP8e0Od72NHnZ8NIalf-ze
AoetFtoVMcbnHpwKtv-3SlkBAlwO2WqAcQFdNxjKOo03CvDvhLvhEZvBn6P39OPAjWuXEE3bBbQqs8BzUY1p-562FWY.Ub
      </securityToken>
    </AuthenticateResponse>
  </soap:Body>
</soap:Envelope>
--uuid:8ce7a3c2-f555-4561-aa2f-7d9d439a4a5c--
```

[36]:

```
client.service.Query.__doc__
```

[36]:

```
'Query(securityToken: xsd:string, dataflow: xsd:NCName, request: xsd:string,
rowId: xsd:integer, maxRows: xsd:integer, parameters: ns0:ParameterType[]) ->
rowId: xsd:integer, rowCount: xsd:integer, lastSet: xsd:boolean, results:
ns0:GenericXmlType'
```

[37]:

```
ParamType = client.get_type("ns0:ParameterType")
ParamType(2010, "Year", "integer", "utf8")  # not using last two
```

[37]:

```
{
    '_value_1': 2010,
    'parameterName': 'Year',
    'parameterType': 'integer',
    'parameterEncoding': 'utf8'
}
```

[54]:

```
# with client.settings(raw_response=True):
#     pass

response = client.service.Query(
    securityToken=token,
    request='GetWaterQualityResults_v1',
    rowId=0,  # not used
    maxRows=500,  # not used
    parameters=[
        ParamType(2010, "Year"),
        ParamType("Cond", "AnalyteCode"),
        ParamType("Temp", "AnalyteCode"),
        # alternative:
        # {'_value_1': "2010", 'parameterName': 'Year', },
        # {'_value_1': "Cond", 'parameterName': 'AnalyteCode', },
    ],
    dataflow="GLENDA",
    )
```

[ ]:
```
# show_response(response)
```

```
User-Agent: Zeep/4.2.1 (www.python-zeep.org)
Accept-Encoding: gzip, deflate, br, zstd
Accept: */*
Connection: keep-alive
SOAPAction: ""
Content-Type: application/soap+xml; charset=utf-8; action=""
Cookie: JSESSIONID=63751690DC5CDF49F835014AE28C004E; cdx-prod-coreservices=1707921987.09.32.283
Content-Length: 1098

<?xml version='1.0' encoding='utf-8'?>
<soap-env:Envelope xmlns:soap-env="http://www.w3.org/2003/05/soap-envelope">
  <soap-env:Body>
    <ns0:Query xmlns:ns0="http://www.exchangenetwork.net/schema/node/2">
      <ns0:securityToken>csm:eyJjd...</ns0:securityToken>
      <ns0:dataflow>GLENDA</ns0:dataflow>
      <ns0:request>GetWaterQualityResults_v1</ns0:request>
      <ns0:rowId>0</ns0:rowId>
      <ns0:maxRows>500</ns0:maxRows>
      <ns0:parameters parameterName="Year" parameterEncoding="None">2010</ns0:parameters>
      <ns0:parameters parameterName="AnalyteCode" parameterEncoding="None">Cond</ns0:parameters
    </ns0:Query>
  </soap-env:Body>
</soap-env:Envelope>

--uuid:d7b64491-ce50-4d34-a0d7-03086c14ff82
Content-Type: application/xop+xml; charset=UTF-8; type="application/soap+xml"
Content-Transfer-Encoding: binary
Content-ID: <root.message@cxf.apache.org>

<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">
  <soap:Body>
    <QueryResponse xmlns="http://www.exchangenetwork.net/schema/node/2" xmlns:xmime="http://www
      <rowId>0</rowId>
      <rowCount>0</rowCount>
      <lastSet>true</lastSet>
      <results format="XML">
        <node2:GLENDA xmlns:node2="http://www.exchangenetwork.net/schema/glenda/1" xmlns:xsi="
          <node2:Result>
            <node2:ProgramTypeCode>WQSF</node2:ProgramTypeCode>
            <node2:GreatLakeCode>Erie</node2:GreatLakeCode>
            <node2:CruiseIdentifier>ER1011</node2:CruiseIdentifier>
            <node2:VisitIdentifier>E009M10</node2:VisitIdentifier>
            <node2:StationIdentifier>ER09</node2:StationIdentifier>
            <node2:ResultLocation>
              <node2:ResultLatitudeMeasure>42.538283</node2:ResultLatitudeMeasure>
              <node2:ResultLongitudeMeasure>-79.616800</node2:ResultLongitudeMeasure>
            </node2:ResultLocation><node2:ResultSamplingDate>2010-04-10</node2:ResultSamplingDa
            <node2:ResultStationDepthMeasure>
              <node2:MeasureValue>50.1000</node2:MeasureValue>
              <node2:MeasureUnitCode>meters</node2:MeasureUnitCode>
            </node2:ResultStationDepthMeasure><node2:ResultSampleDepthMeasure>
              <node2:MeasureValue/>
              <node2:MeasureUnitCode>meters</node2:MeasureUnitCode>
            </node2:ResultSampleDepthMeasure>
            <node2:ResultDepthCode>Synthetic Sample</node2:ResultDepthCode>
            <node2:ResultMediumName>surface water</node2:ResultMediumName>
            <node2:ResultSampleTypeName>Individual QC</node2:ResultSampleTypeName>
            <node2:ResultQualityControlTypeName>field blank</node2:ResultQualityControlTypeNam
            <node2:ResultSampleIdentifier>10GC20R80</node2:ResultSampleIdentifier>
            <node2:ResultAnalyteCode>Cond</node2:ResultAnalyteCode>
            <node2:ResultAnalyteText>Conductivity</node2:ResultAnalyteText>
            <node2:ResultMeasure>
              <node2:MeasureValue>2.095</node2:MeasureValue>
              <node2:MeasureUnitCode>umho/cm</node2:MeasureUnitCode>
            </node2:ResultMeasure>
            <node2:ResultSampleFractionText>Total/Bulk</node2:ResultSampleFractionText>
            <node2:ResultMethodIdentifier>LG500</node2:ResultMethodIdentifier>
            <node2:ResultRemarkText>Field Reagent Blank, failed</node2:ResultRemarkText>
          </node2:Result>
          <node2:Result>
             ...
```

[55]:

```python
root = response['results']['_value_1']  # an ElementTree element (XML node)
```

[ ]:

```python
# etree.dump(root)
```

[56]:

```python
# Make data frame
from lxml import etree
NS = {"n2": "http://www.exchangenetwork.net/schema/glenda/1"}
rows = root.xpath("//n2:Result", namespaces=NS)
# Flatten nested elements
```

```
fields = []
for elem in rows[0]:
    if len(elem) > 0:  # has children
        fields.extend(
            etree.QName(elem).localname + "_" + etree.QName(child).localname
            for child in elem
        )
    else:
        fields.append(etree.QName(elem).localname)

print("\n".join(fields))

# .//* recurses into nested lat/lon and value/unit elements, len(elem) == 0␣
 ↪selects the leaf nodes
import pandas as pd
df = pd.DataFrame(
    [
        [elem.text for elem in row.xpath(".//*") if len(elem) == 0]
        for row in rows
    ],
    columns=fields,
)
df
```

```
ProgramTypeCode
GreatLakeCode
CruiseIdentifier
VisitIdentifier
StationIdentifier
ResultLocation_ResultLatitudeMeasure
ResultLocation_ResultLongitudeMeasure
ResultSamplingDate
ResultStationDepthMeasure_MeasureValue
ResultStationDepthMeasure_MeasureUnitCode
ResultSampleDepthMeasure_MeasureValue
ResultSampleDepthMeasure_MeasureUnitCode
ResultDepthCode
ResultMediumName
ResultSampleTypeName
ResultQualityControlTypeName
ResultSampleIdentifier
ResultAnalyteCode
ResultAnalyteText
ResultMeasure_MeasureValue
ResultMeasure_MeasureUnitCode
ResultSampleFractionText
ResultMethodIdentifier
```

```
     ResultRemarkText

[56]:      ProgramTypeCode GreatLakeCode CruiseIdentifier VisitIdentifier  \
      0               WQSF          Erie           ER1011         E009M10
      1               WQSF          Erie           ER1011         E009M10
      2               WQSF          Erie           ER1011         E009M10
      3               WQSF          Erie           ER1011         E009M10
      4               WQSF          Erie           ER1011         E009M10
      …              …            …              …             …
      1632            WQSF      Superior           SU1021         SOFEG10
      1633            WQSF      Superior           SU1021         SOFEG10
      1634            WQSF      Superior           SU1021         SOFEG10
      1635            WQSF      Superior           SU1021         SOFEG10
      1636            WQSF      Superior           SU1021         SOFEG10

           StationIdentifier ResultLocation_ResultLatitudeMeasure  \
      0                 ER09                            42.538283
      1                 ER09                            42.538283
      2                 ER09                            42.538283
      3                 ER09                            42.538283
      4                 ER09                            42.538283
      …                …                                  …
      1632              SUFE                            46.916667
      1633              SUFE                            46.916667
      1634              SUFE                            46.916667
      1635              SUFE                            46.916667
      1636              SUFE                            46.916667

           ResultLocation_ResultLongitudeMeasure ResultSamplingDate  \
      0                               -79.616800         2010-04-10
      1                               -79.616800         2010-04-10
      2                               -79.616800         2010-04-10
      3                               -79.616800         2010-04-10
      4                               -79.616800         2010-04-10
      …                                    …                 …
      1632                            -90.416617         2010-08-23
      1633                            -90.416617         2010-08-23
      1634                            -90.416617         2010-08-23
      1635                            -90.416617         2010-08-23
      1636                            -90.416617         2010-08-23

           ResultStationDepthMeasure_MeasureValue  \
      0                                   50.1000
      1                                   50.1000
      2                                   50.1000
      3                                   50.1000
      4                                   50.1000
```

```
…                                     …
1632                                13.2000
1633                                13.2000
1634                                13.2000
1635                                13.2000
1636                                13.2000

     ResultStationDepthMeasure_MeasureUnitCode  … ResultSampleTypeName  \
0                                       meters  …          INSITU_MEAS
1                                       meters  …           Individual
2                                       meters  …           Individual
3                                       meters  …           Individual
4                                       meters  …           Individual
…                                        …  …                  …
1632                                    meters  …           Individual
1633                                    meters  …           Individual
1634                                    meters  …           Individual
1635                                    meters  …            Composite
1636                                    meters  …          INSITU_MEAS

     ResultQualityControlTypeName ResultSampleIdentifier ResultAnalyteCode  \
0                            None           999994404823              Temp
1            routine field sample              10GC20S81              Temp
2            routine field sample              10GC20S8A              Temp
3            routine field sample              10GC20S8B              Temp
4            routine field sample              10GC20S8C              Temp
…                           …                    …               …
1632         routine field sample              10GS51S61              Temp
1633         routine field sample              10GS51S62              Temp
1634         routine field sample              10GS51S71              Temp
1635         routine field sample              10GS51I72              Temp
1636                         None           999994407217              Temp

     ResultAnalyteText ResultMeasure_MeasureValue  \
0          Temperature                        4.8
1          Temperature                        1.7
2          Temperature                        1.7
3          Temperature                        1.7
4          Temperature                        1.7
…                …                        …
1632       Temperature                       18.8
1633       Temperature                       18.3
1634       Temperature                       15.5
1635       Temperature                       18.8
1636       Temperature                       19.1

     ResultMeasure_MeasureUnitCode ResultSampleFractionText  \
0                                C           Not applicable
1                                C           Not applicable
2                                C           Not applicable
3                                C           Not applicable
4                                C           Not applicable
…                            …                      …
1632                             C           Not applicable
1633                             C           Not applicable
1634                             C           Not applicable
1635                             C           Not applicable
1636                             C           Not applicable

     ResultMethodIdentifier ResultRemarkText
0                        NA             None
1                        NA             None
2                        NA             None
3                        NA             None
4                        NA             None
…                     …              …
1632                     NA             None
1633                     NA             None
1634                     NA             None
1635                     NA             None
1636                     NA             None

[1637 rows x 24 columns]
```

[ ]:
```
# show_response(response)
```