

Lukáš HERMAN

Zpracování XML dokumentů v FME

Zdroje I.

- KEMP, K. K.: *Encyclopedia of geographic information science*. Los Angeles, SAGE Publications, 2008. 558 s. ISBN 9781412913133.
- KOSEK, J.: *XML pro každého: Podrobný průvodce*. 1. vyd., Grada Publishing a.s., Praha, 2000. 164 s. ISBN 80-7169-860-1.
- POKORNÝ, J. a kol.: *XML technologie principy a aplikace v praxi*. 1. vyd., Grada Publishing a.s., Praha, 2008. 272 s. ISBN 978-80-247-2725-7.
- *XSLT Tutorial* [on-line]. 2009, [cit.2012]. Dostupné z WWW: <http://www.w3schools.com/Xsl/default.asp>
- ČERBA, O.: *XSLT eXtensible Stylesheet Language Transformation* [on-line]. 2007, [cit. 2012]. Dostupné z WWW: http://gis.zcu.cz/studium/pok/Materialy/o6_XSLT.pdf
- MURRAY, D.: *Battle the Forces of XML with FME 2011: Top 5 Ways* [on-line]. 2007, [cit. 2012]. Dostupné z WWW: <http://blog.safe.com/2011/01/battle-the-forces-of-xml-with-fme-2011-top-5-ways/>
- *KML Documentation Introduction* [on-line]. 2012, [cit. 2012]. Dostupné z WWW: <https://developers.google.com/kml/documentation/?hl=cs>
- RŮŽIČKA, J.: *GML – Geography Markup Language - Rozšířená verze* [on-line]. 2007, [cit. 2012]. Dostupné z WWW: <http://gis.vsb.cz/ruzicka/Seminare/Inter2/prezentace/GMLextended.pdf>

Zdroje II.

- DAHMEN, C.: *FME and CityGML – Generating 3D City Models in a Variety of Applications* [on-line]. 2009 [cit. 2012]. Dostupné z WWW: http://www.citygml.org/fileadmin/citygml/docs/CityGML_FME2009.pdf
- HERMAN, L.: *Moderní kartografické metody modelování měst*. Diplomová práce, Masarykova univerzita, Přírodovědecká fakulta, Brno, 2011. Dostupné z WWW: http://is.muni.cz/th/222752/prif_m/
- SAFE SOFTWARE: *FME Readers and Writers* [on-line]. 2011 [cit. 2012]. Dostupné z WWW: http://docs.safe.com/fme/html/FME_ReadersWriters/Default.htm
- SAFE SOFTWARE: *FME Transformer Reference Guide* [on-line]. 2011 [cit. 2012]. Dostupné z WWW: <http://downloads.safe.com/fme/brochures/transformers.pdf>
- Návod k programu FME Desktop 2011
- <http://www.safe.com>
- <http://fmepedia.safe.com>
- <http://groups.google.com/group/fmetalk>

XML

- **eXtensible Markup Language**
- Formát definovalo konsorcium W3C jako formát pro přenos dat a obecných dokumentů
- XML je zkratka pro, tzn. rozšiřitelný značkovací jazyk. Návrh XML vychází ze staršího a obecnějšího standardu SGML (*Standard Generalized Markup Language*)
- Sada XML značek není pevná a může být definována pro různé sady dokumentů různě
- Definice značek může být součástí definice XML dokumentu, může být specifikována odkazem, nebo může být dohodnuta předem



XML - syntaxe

XML deklarace

kódování dokumentu

```
<?xml version="1.0" encoding="windows-1250"?>
```

```
<cenik>
```

```
<název>Počítačové komponenty</název>
```

prázdný element

```
<platnost od="1.1.2000" do="31.3.2000"/>
```

atribut

```
<dodavatel>
```

```
<název>První hardwarová, s.r.o.</název>
```

```
<adresa>
```

počáteční tagy

```
<ulice>Průmyslová 12</ulice>
```

```
<město>Praha 10</město>
```

```
<psč>100 00</psč>
```

```
<email>info@prhw.cz</email>
```

ukončovací tagy

```
</dodavatel>
```

obsah elementu nabídka

```
<nabídka>
```

```
<produkt kategorie="polohovací zařízení" kód="pxbd-21">
```

```
<název>Hyperoptická digitální myš</název>
```

```
<cena měna="CZK">368.30</cena>
```

```
</produkt>
```

název atributu

hodnota atributu

```
<produkt kategorie="pevné disky" kód="sdhd-99">
```

```
<název>Soft-slow disc 19,3 GB</název>
```

```
<cena měna="CZK">8500</cena>
```

```
</produkt>
```

obsah elementu cena

```
<produkt kategorie="polohovací zařízení" kód="pxbd-13">
```

```
<název>Tlakový tablet</název>
```

```
<cena měna="CZK">5635.20</cena>
```

```
</produkt>
```

```
</nabídka>
```

kořenový element

Dokument musí obsahovat PROLOG a právě JEDEN KOŘENOVÝ ELEMENT

Všechny ELEMENTY jsou zapsány pomocí PÁROVÝCH a NEPÁROVÝCH TAGŮ

TAGY musí být řádně UKONČENY a nesmí se KŘÍŽIT

Elementy mohou obsahovat ATRIBUTY

Lze vkládat i KOMENTÁŘE

Blok textu, který by mohl být mylně považován za značku, vyjadřujeme sekcí CDATA

Tagy lze definovat pomocí DTD (definice typu dokumentu)

XML – obecné zásady tvorby

- Formát XML musí být kompatibilní s formátem SGML
- Formát XML musí být použitelný v rámci internetu
- Formát XML by měl podporovat širokou škálu aplikací
- Musí být snadné vytvářet programy, které manipulují s dokumenty v XML
- Množství variant XML by mělo být minimální – nejlépe žádné
- XML dokumenty by měly být čitelné a pochopitelné i pro člověka
- Na základě těchto principů byla navržena definice XML, které zahrnuje dvě části:
 - Definici, co to je XML dokument,
 - Definici programů, které zpracovávají XML dokumenty – XML procesorů

Aplikace XML

- Široké využití – geodata (2D, 3D), metadata, vektorová grafika (2D, 3D), multimédia, kancelářské aplikace, webové aplikace, ...

```
<dc:metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:subject>Utah</dc:subject>
  <dc:subject>boundaries</dc:subject>
  <dc:subject>County</dc:subject>
  <dc:subject>Administrative</dc:subject>
  <dc:subject>geoscientificInformation</dc:subject>
  <dc:description>This data set represents county boundaries in Utah at 1:24,000 scale.</dc:description>
  <dc:date>2004-04-20T00:00:00.000</dc:date>
  <dc:type>dataset</dc:type>
  <dc:identifier xmlns:dc="http://purl.org/dc/elements/1.1/">{42AE2814-FCC1-4BC2-BAF4-CA3E55514997}</dc:identifier>
  <dc:language>en</dc:language>
  <dc:spatial>
    <dcmiBox:Box name="Geographic" projection="EPSG:4326" xmlns:dcmiBox="http://dublincore.org/documents/2000/07/11
      <dcmiBox:northlimit units="decimal degrees">42.01</dcmiBox:northlimit>
      <dcmiBox:eastlimit units="decimal degrees">-109.21</dcmiBox:eastlimit>
      <dcmiBox:southlimit units="decimal degrees">36.98</dcmiBox:southlimit>
      <dcmiBox:westlimit units="decimal degrees">-114.1</dcmiBox:westlimit>
    </dcmiBox:Box>
  </dc:spatial>
  <dc:rights></dc:rights>
</dc:metadata>
```

XML v FME I.

- **XMLFeatureMapper** vytváří prvky z XML dokumentů prostřednictvím xfMaps
- **XMLFormatter** provádí formátování XML dokumentu
- **XMLFragmenter** mapuje elementy z XML dokumentu do XML fragmentů

* "dc:subject"	Extracts every <dc:subject> into an XML fragment.
* "dc:spatial/dcmiBox:Box"	Extracts the <dcmiBox:Box> fragment, but <dc:spatial> must be the parent.
* "dcmiBox:Box/*"	Extracts every child of <dcmiBox:Box> into fragments, 4 fragments features corresponding to <dcmiBox:northlimit>, <dcmiBox:eastlimit>, <dcmiBox:southlimit> and <dcmiBox:westlimit> are output.
* "dc:subject dc:spatial/dcmiBox:Box dcmiBox:Box/*"	The three previous matched expressions combined, each separated by whitespace.

XQuery

- Původně se mělo jednat o dotazovací jazyk nad daty ve formátu XML (jako SQL)
- Dnes je využití mnohem širší: „XQuery je standardní jazyk pro kombinování dokumentů, databází, webových stránek a skoro všeho ostatního.“
- Formálně složen ze tří částí: Xquery 1.0, Xpath 2.0 a XPath 1.0
- Jádro tvoří tzv. výrazy FLWOR
- Obsahuje proměnné, cykly, funkce, ...

XML v FME II.

- **XQueryExploder** používá jazyk XQuery k extrakci dat z XML a jejich uložení do nových FME prvků
- **XQueryExtractor** používá jazyk XQuery k extrakci částí XML struktury a jejich uložení do atributů
- **XQueryUpdater** provádí změny v XML dokumentu prostřednictvím jazyka XQuery
- **XMLTemplater** z XML dokumentu vytváří šablony (templates) s hodnotami podle atributů

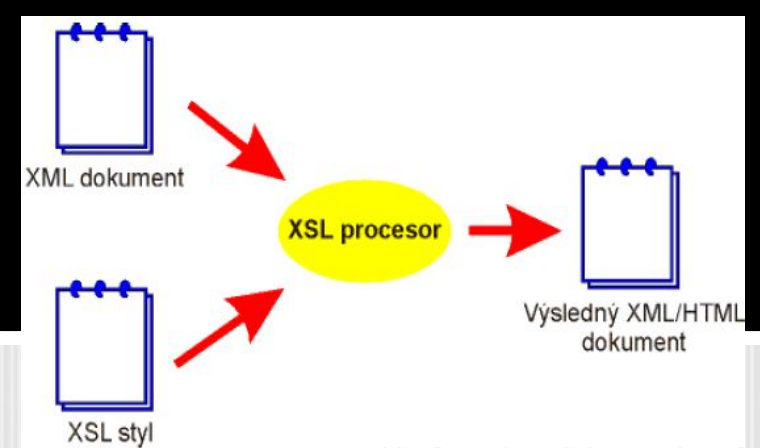
```
<ogc:PropertyIsLike xmlns:ogc="http://www.opengis.net/ogc"
  wildcard="{fme:get-attribute("_wildcard")}"
  singleChar="{fme:get-attribute("_singlechar")}"
  escape="{fme:get-attribute("_escapechar")}">

  <ogc:PropertyName>{fme:get-attribute("_property_name")}</ogc:PropertyName>
  <ogc:Literal>{fme:get-attribute("_literal")}</ogc:Literal>
</ogc:PropertyIsLike>
```

Jmenný prostor (Namespace)

- Jmenný prostor je sadou jmen, které lze v dokumentu použít jako elementy a atributy, tato sada jmen je identifikována pomocí URI
- Název jmenného prostoru se skládá z prefixu jmenného prostoru, dvojtečky a lokální části, např: *xsl:template*
- Některé prefixy jsou již standardně předdefinované (např.: *xml*)
- Umožňuje spojovat různé XML dialekty do jednoho fyzicky uloženého dokumentu

XSL Transformation



- **eXtensible Stylesheet Language**
- Původně to byl nástroj, pomocí kterého bylo možné měnit styl XML dokumentů
- Dnes to je spíše popis obecných transformací z XML dokumentů do (libovolného) formátu
- XSL je XML dokument, který funguje jako „slovník“, XSLT procesor představuje „překladače“ (XSLT procesorem je např. FME)
- Na rozhraní mezi dotazovacími a programovacími jazyky, obsahuje např. proměnné, podmínky, cykly, ...

XML v FME III.

- **XMLValidator** validuje XML syntaxi nebo schéma dokumentu oproti jinému XML souboru
- **XMLNamespaceDeclarer** doplňuje chybějící jmenné prostory v XML dokumentu na základě srovnání prefixů s e vzorovým XML souborem, v němž jsou jmenné prostory správně definovány
- **XSLTProcessor** aplikuje na XML dokumenty XSL transformace, výstupem může být text, RSS, SVG nebo CSV

GML

- **Geography Markup Language**
- XML gramatika definovaná OGC pro vyjádření geografických prvků
- Modelovací a výměnný jazyk geografických systémů na internetu
- Je obecný – ukládá vektory (polygony, linie, body), ale také coverages a může obsahovat i topologii
- Definice GML vznikla v roce 1999
- Současná verze 3.2.1, byla v roce 2007 publikována jako ISO 19136
- GML je základní formát pro poskytování dat podle INSPIRE

GML - struktura

Silnice

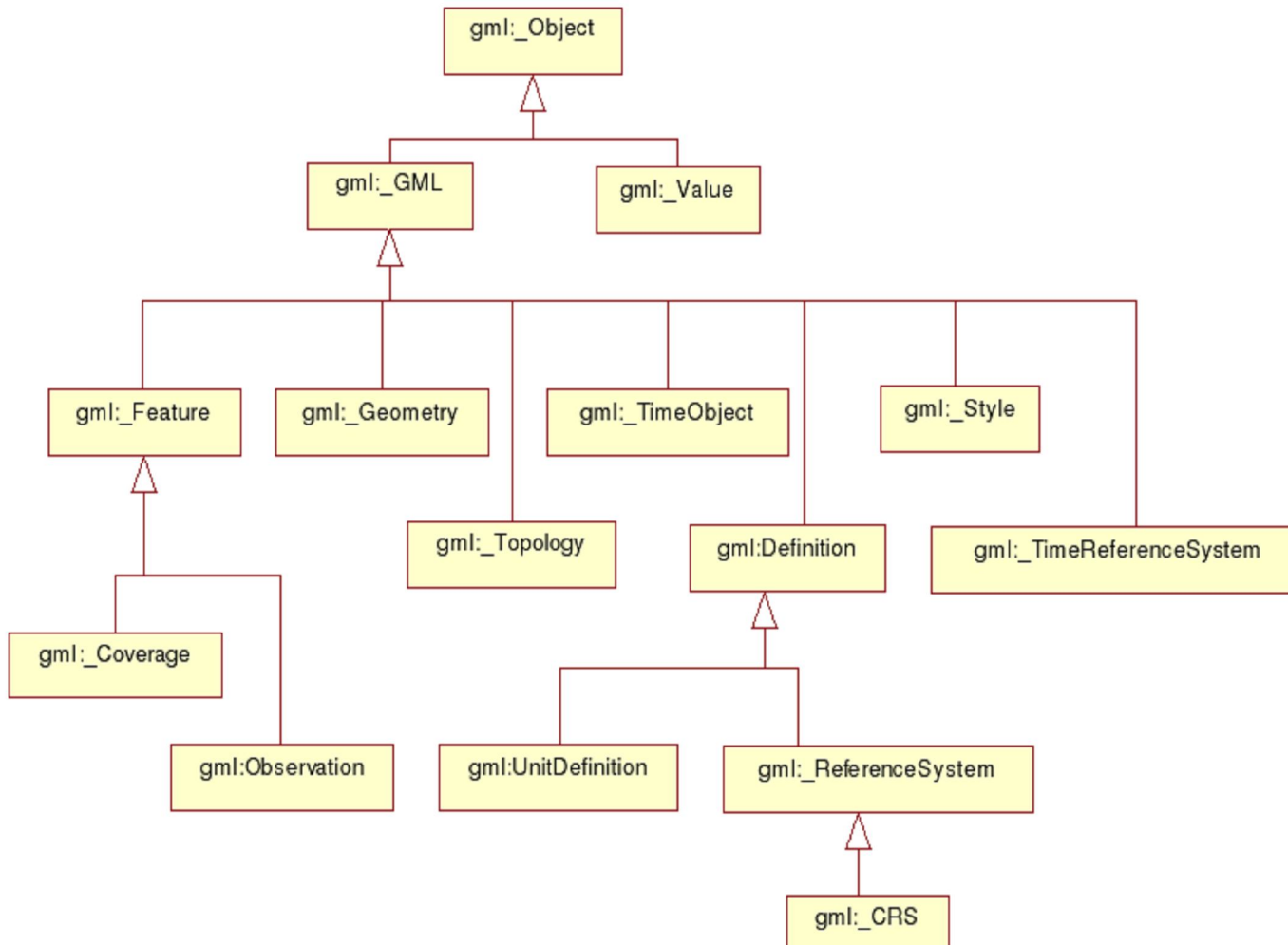
nazev *D1*

trida *dálnice*

stred_linie **gml:Curve**

spravce *ŘSD*

```
<Silnice gml:id="o.1f75dc">  
  <nazev>D1</nazev>  
  <trida>dálnice</trida>  
  <centerLine>  
    <gml:Curve>...</gml:Curve>  
  </centerLine>  
  <spravce>ŘSD</spravce>  
</Silnice>
```



GML 2 x GML 3

- GML verze 2 tvoří 3 základní schémata:
 - *Geometry.xsd* (definuje geometrickou složku geoprvků)
 - *Feature.xsd* (definuje hlavní model prvek – vlastnost; rámec pro vytváření prvků)
 - *Xlink.xsd* (definuje funkce pro odkazování)
- GML verze 3 je rozšířeno o dalších 25 schémat
- Např. OpenLayers nezvládají práci s GML 3
- V FME jsou to odlišné *Writtery*


```
<?xml version="1.0" encoding="UTF-8"?>
<gml2:FeatureCollection xmlns:gml2="http://www.safe.com/gml2" xmlns:gml="http://www.
<gml:boundedBy>
<gml:Box srsName="LL-WGS84"><gml:coordinates>19.7005487829139,42.6893578687921 19.72
</gml:boundedBy>
<gml:featureMember>
<gml2:desnarijeka_line>
<gml2:REDNI_BROJ>329</gml2:REDNI_BROJ>
<gml2:NAZIV_TOKA>Desna rijeka</gml2:NAZIV_TOKA>
<gml:lineStringProperty>
<gml:LineString srsName="LL-WGS84"><gml:coordinates>19.726738129173,42.6893578687921
</gml:lineStringProperty>
</gml2:desnarijeka_line>
</gml:featureMember>
</gml2:FeatureCollection>
```

river3.gml

```
<?xml version="1.0" encoding="UTF-8"?>
<gml:FeatureCollection xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www
<gml:boundedBy>
<gml:Envelope srsName="EPSG:4326" srsDimension="2">
<gml:lowerCorner>42.6893578687921 19.7005487829139</gml:lowerCorner>
<gml:upperCorner>42.7036185059674 19.726738129173</gml:upperCorner>
</gml:Envelope>
</gml:boundedBy>
<gml:featureMember>
<fme:desnarijeka_line gml:id="idc0c218af-f267-483f-blcc-91883563ff5f">
<fme:REDNI_BROJ>329</fme:REDNI_BROJ>
<fme:NAZIV_TOKA>Desna rijeka</fme:NAZIV_TOKA>
<gml:curveProperty>
<gml:LineString srsName="EPSG:4326" srsDimension="2">
<gml:posList>42.6893578687921 19.726738129173 42.6895140673165 19.7265132034818 42.68
</gml:LineString>
</gml:curveProperty>
</fme:desnarijeka_line>
</gml:featureMember>
</gml:FeatureCollection>
```

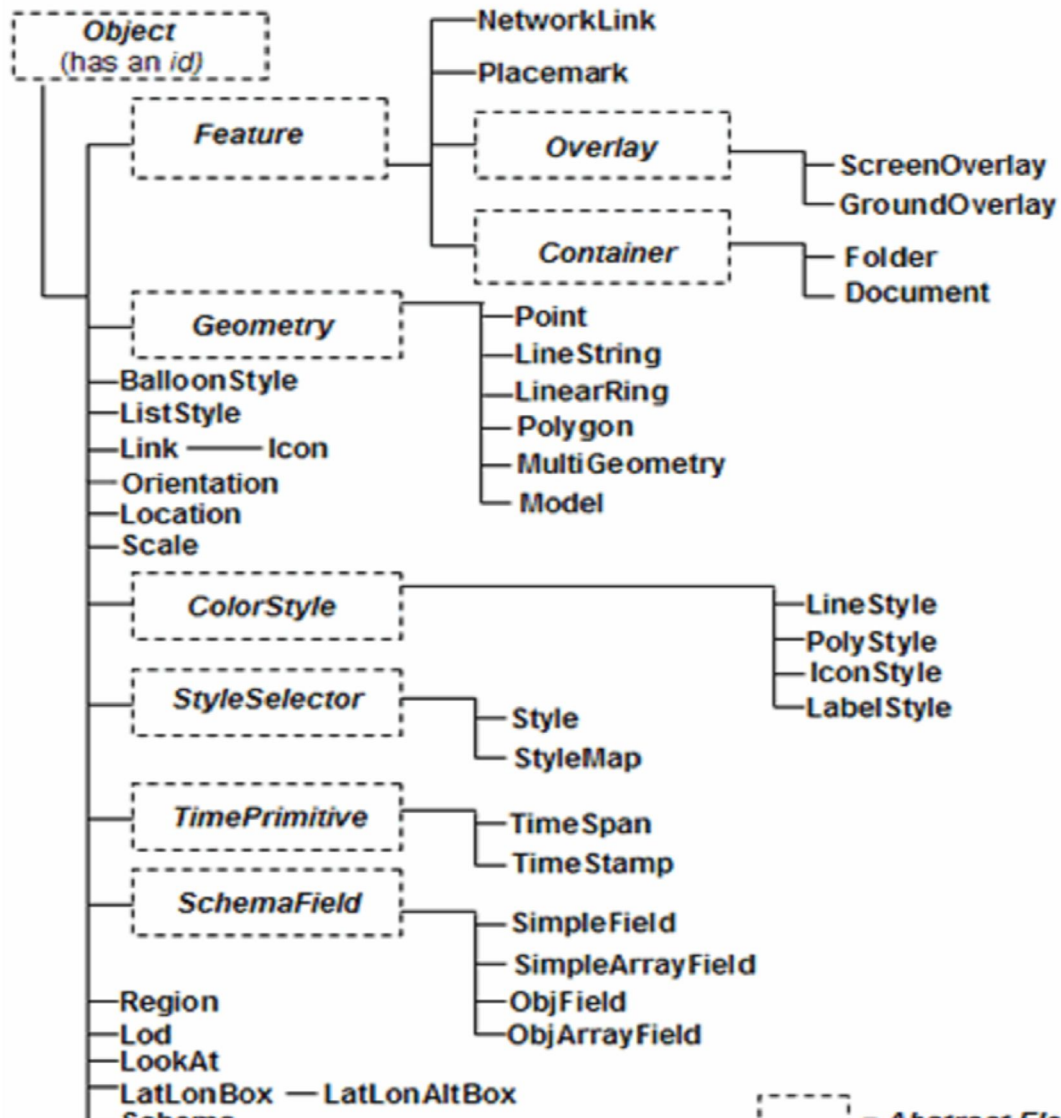
GML v FME

- **GMLFeatureExtractor** vytváří GML2 dokument ze vstupních prvků, když je ukládá jako specifické atributy prvků
- **GMLFeatureReplacer** vytváří GML prvky z atributů uložených ve vstupním dokumentu

KML

- **Keyhole Markup Language** nebo **Keyhole Mapping Language**
- KML byl původně vyvinut firmou *Keyhole, Inc.* jako API pro její virtuální glóbus Earth Viewer.
- V roce 2004 firmu koupila společnost *Google, Inc.* a produkt přejmenovala na Google Earth
- V roce 2008 se stal KML ve verzi 2.2 standardem *Open Geospatial Consortium* (OGC)
- Slouží k modelování a ukládání prostorových dat (body, čáry, rastrové data, mnohoúhelníky a 3D modely)

KML – datový model



KML v FME I.

- **KMLPropertySetter** nastavuje KML prvkům obecné vlastnosti
 - **Navigation Tree:** Name, Summary, Visible; **Balloon:** Content Type, Content, Include Attribute Table; **Geometry:** Geometry Type, Altitude Mode, Raster Altitude, Raster Opacity, Extrude, Follow Terrain
- **KMLRegionSetter** určuje vlastnosti související s lokalizací
 - **Bounding Box:** Calculate, Minimum X, Y/Maximum X, Y; **Display Criteria:** Minimum Display Size, Maximum Display Size, Minimum Fade Extent, Maximum Fade Extent
- **KMLStyler** nastavuje vizualizaci (vzhled) KML prvků
 - **Allow Unique Styles Per Feature;** **Color:** Color, Fill Color, Opacity, Fill Opacity; **Icon:** Name, Scale; **Line Style:** Line Width; **Label Style:** Scale

KML v FME II.

Transformer

Transformer Name:

Time

Type:

Timestamp:

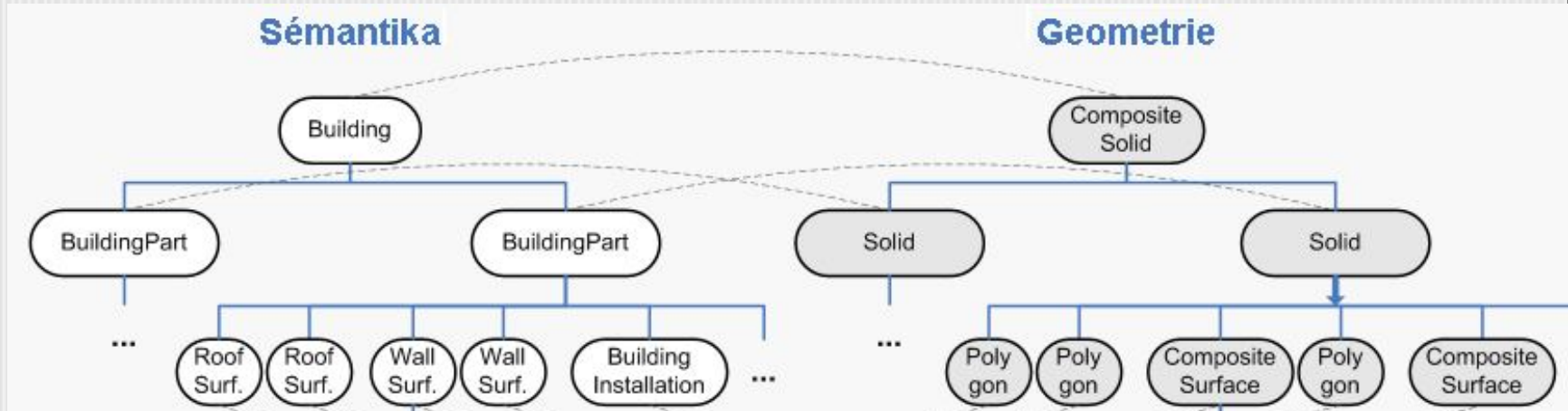
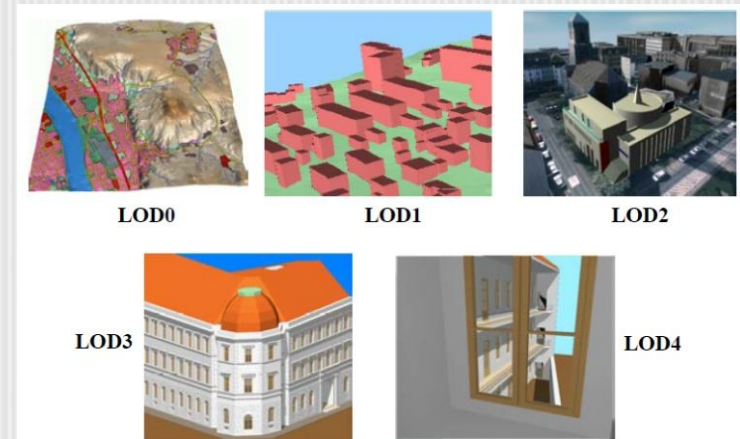
Period Start:

Period End (optional):

- **KMLTimeSetter** vytváří časové určení KML prvkům
 - Type, Timestamp, Period Start, Period End
- **KMLTourBuilder** vytváří ze vstupních prvků průlet
 - **Tour:** Tour Name, Tour Duration; **Transition:** Type; **Balloon:** Display; **Delay:** Type, Duration; **View:** Perspective, Range, Heading, Tilt
- **KMLViewSetter** nastavuje parametry pohledu na KML prvky
 - **Location:** Longitude, Latitude, Altitude, Altitude Mode; **View:** View Type, Heading, Tilt, Roll, Range

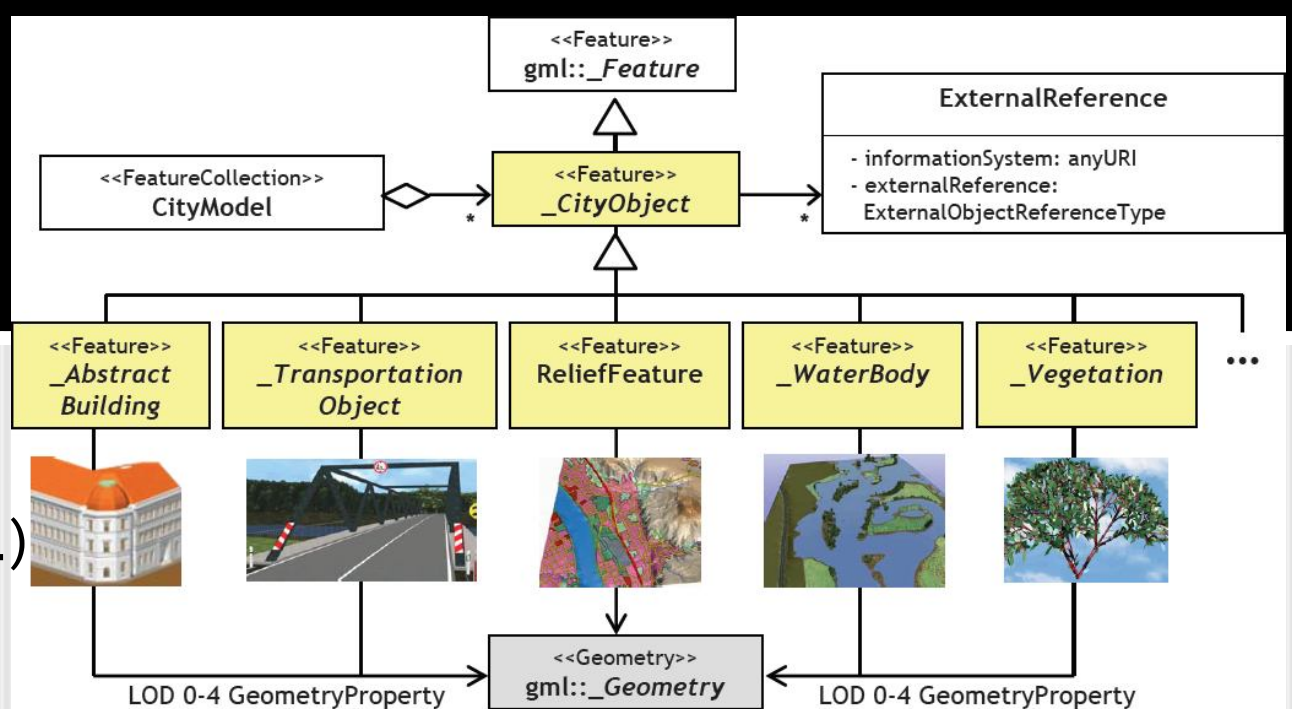
CityGML I.

- Vývoj zahájen v roce 2002, v říjnu 2008 se stalo (ve verzi 1.0) standardem organizace OGC
- Základem je GML 3.1.1 a standardy ISO
- Víceúrovňová reprezentace ($LOD_0 - LOD_4$)
- Geometrie (GML 3 + specifické koncepty)
- Topologie (*XLink*), textury a materiály
- Klíčový princip: sémantické modelování



CityGML II.

- Tematický model
 - Terén (GRID, TIN, ...)
 - Budovy
 - Dopravní objekty
 - Hydrografie
 - Vegetace
 - LandUse
 - CityFurniture (městský mobiliář, např. světla, semaforey, ...)
- Rozšiřitelnost
 - Generické elementy a atributy
 - ADE = rozšíření, které mají vlastní XSD schéma, např. NOISE ADE (součást OGC specifikace CityGML 1.0 a podporováno v FME)



CityGML v FME I.

- Důležitou částí při tvorbě CityGML modelů je rovněž nastavení správného typu geometrie pro jednotlivé třídy
- K tomu slouží **GeometryTraitSetter**, jenž definuje geometrii prvků (*INPUT*) na základě zvoleného atributu (*Source Attributes*). Původní typ geometrie může být přepisován (*Overwrite Existing Traits – Yes*) nebo nemusí
- Ne všechny třídy mají geometrii

The screenshot displays two transformers in an FME workspace: 'AttributeCreator_3' and 'GeometryTraitSetter_6'. The 'AttributeCreator_3' transformer has an 'INPUT' and an 'OUTPUT' port. The 'GeometryTraitSetter_6' transformer also has an 'INPUT' and an 'OUTPUT' port. Arrows indicate a data flow from the 'OUTPUT' of 'AttributeCreator_3' to the 'INPUT' of 'GeometryTraitSetter_6'. Below the workspace, two parameter windows are shown. The 'AttributeCreator Parameters' window shows the transformer name 'AttributeCreator_3' and a table of new attributes:

Attribute	Value
citygml_lod_name	lod1Geometry
citygml_level_of_detail	1

The 'GeometryTraitSetter Parameters' window shows the transformer name 'GeometryTraitSetter_6', the source attribute 'citygml_lod_name', and the 'Overwrite Existing Traits' option set to 'Yes'.

CityGML v FME II.

- Je nutné správně vytvořit sémantickou složku modelu
 - Validní pojmenování vytvářených tříd
 - Správné provázání těchto tříd
 - *FeatureRole* se zadává jako atribut, např. *cityObjectMember*, *noiseRoadSegmentProperty*
 - Pomocí atributů *gml_id* a *gml_parent_id* se určuje třída (rodič) a její potomka (mají shodnou hodnotu těchto atributů)
- Vytvoření korektního 3D modelu pro účely hlukového mapování vyžaduje také správné nastavení *Writeru*

