Geospatial Data and Processing in Apache Projects

George Percivall
CTO, Chief Engineer – Open Geospatial Consortium
percivall@apache.org

ApacheCon NA 2019
Using location, we connect people, communities, technology and decision making to create a sustainable future for us, our kids and future generations

- By specializing in making location more Findable, Accessible, Interoperable and Reusable
- Via a proven collaborative and agile process combining standards, innovation and partnerships

Comprehensive global community-driven forward-looking expertise in location
Open Source and Open Standards

Open consensus standards as building blocks in Open Source
– Increase code quality
– Reduce development effort
– Stable and proven APIs and encodings for interoperability
– Interchangeability of software components
Spatial Things

“Anything with spatial extent; size, shape, position, e.g. people, places, objects”

Feature Data

Coverage Data

Metadata

Maps

OGC®
Best Practices for Spatial Data

- Choose **coordinate reference systems** to suit your user’s applications
- Provide **geometries** in a usable way: right level of accuracy, precision, size
- Use **spatial data encodings** that match your target audience
- Make data available through **APIs**
- **Link resources** together to create the Web of data
- **Spatial analytics** - discover meaningful patterns in spatial data

OGC/W3C Spatial Data on the Web Best Practices [https://www.w3.org/TR/sdw-bp/]
In the beginning…

All was simple, everyone agreed: the world was flat.
‘continent’ in different map projections

- Orthographic projection
- Globular projection
- Mercator projection
- Stereographic projection
What is Geodesy?

Surprise: Latitude is not unique nor is Longitude

Due to different Geodetic Datums:

\[ \phi_1 \not= \phi_2 \]
What errors can you expect?

- Wrong geodetic datum:
  - several hundreds of metres

- Incorrect ellipsoid:
  - horizontally: several tens of metres
  - height: not effected, or tens to several hundred metres

- Wrong map projection:
  - entirely the wrong projection:
    - hundreds, even thousands of kilometres (at least easy to spot!)
  - partly wrong (i.e. one or more parameters are wrong):
    - several metres to many hundreds of kilometres

- No geodetic metadata \(\rightarrow\) coordinates cannot be interpreted
  - Datum; ellipsoid; prime meridian; map projection
CRS WKT: Dynamic CRS, ellipsoidal 3D coordinate system

GEOGCRS["WGS 84 (G1762)",
    DYNAMIC[FRAMEEPOCH[2005.0]],
    TRF["World Geodetic System 1984 (G1762)",
        ELLIPSOID["WGS84",6378137,298.257223563,LENGTHUNIT["metre",1.0]]],

    CS[ellipsoidal,3],
    AXIS["(lat)",north,ANGLEUNIT["degree",0.0174532925199433]],
    AXIS["(lon)",east,ANGLEUNIT["degree",0.0174532925199433]],
    AXIS["ellipsoidal height (h)",up,LENGTHUNIT["metre",1.0]]

    ID["EPSG",4269],
    REMARK["1986 realisation"]}
Reference system by EPSG code

```java
CRS.forCode("EPSG::26747"); // NAD27 / California zone VII
```

**WARNING:** Code "EPSG:26747" is deprecated and superseded by 26799. Reason is: Error in dependent projection record.

**Coordinate operations**

```java
import org.opengis.referencing.operation.CoordinateOperation;
// Class declaration omitted for brevity
CoordinateReferenceSystem sourceCRS = // may be parsed from WKT, EPSG code, etc.
CoordinateReferenceSystem targetCRS = // idem
CoordinateOperation op = CRS.findOperation(sourceCRS, targetCRS, region);

// Verify domain of validity and accuracy
System.out.println("Valid in " + CRS.getGeographicBoundingBox(op));
System.out.println("Accuracy " + CRS.getLinearAccuracy(op) + " m");
```
OGC simple features geometries restricted to 0, 1 and 2-dimensional geometric objects that exist in 2-dimensional coordinate space (R2)

http://www.opengeospatial.org/standards/sfa
Topological Relations between Spatial Objects

Equals

A \ cong \ B

Touches

A \ touches \ B

Overlaps

A \ overlaps \ B

Contains

A \ contains \ B

Within

A \ within \ B

Disjoint

A \ disjoint \ B

Intersects

A \ intersects \ B

Crosses

A \ crosses \ B

Defined in OGC Simple Features Access Standard
http://www.opengeospatial.org/standards/sfa

Also defined in GeoSPARQL
http://www.opengeospatial.org/standards/geosparql
OGC® Spatial Temporal Geometry

1 prism = 1 leaf + 1 sweep (&attribute)

End leaf of tracks

OGC Moving Features Standard
Operations between one trajectory object and one or more geometry objects

An example is “intersection” between a geometry object and a trajectory of a moving feature like a car, a person, a vessel, an aircraft, and a hurricane.
Moving Feature Access: Two Trajectories

Operations between two trajectory objects

An example is to calculate a distance of the nearest approach of a trajectory to another trajectory.
Encodings
GeoPackage

• GeoPackage is a universal file format for geodata.
  – open, standards-based, application and platform independent, and self-describing.
  – Works on any desktop or mobile OS using SQLite
  – Connected / limited / disconnected environment use

• GeoPackage - the modern alternative to formats like SDTS and vendor specific

• Experience it here:
  http://www.ogcnetwork.net/geopackage
GeoPackage: Raster Maps, Images, Feature Data in One File

Imagery
Low – High Resolution

Raster Maps:
Small – Large Scale

Tile Pyramids
24 zoom levels

Feature Data

Single File Sqlite Database
containing all data for direct-use on mobile platforms & handheld devices
Two Different Use Patterns

- **Thematic communities describe spatial datasets**: Cadastre, Topography, Geology, Hydrography, Meteorology, Aviation, City Models, etc.

- **Embed location in other XML grammars**: GeoRSS, GeoSPARQL (OGC), Geopriv (IETF), POI (W3C), Sensor Web (OGC), etc.
CityGML: Simple to Complex objects with structured geometry

- Geometric entities know **WHAT** they are
- Semantic entities know **WHERE** they are and what their spatial extents are
OGC GeoTIFF 1.1

- OGC standard for georeferencing information to be embedded in a TIFF file.
  - map projection, coordinate systems, ellipsoids, datums, etc. for spatial reference of data

- GeoTIFF compliant with TIFF
  - Software incapable of reading geo metadata will still open a GeoTIFF file

- Widely supported
  - including GDAL and libgeotiff
Geospatial Data and Processing in Apache Projects

APIs and Web Services
1000s of Services, 100Ks Datasets Implement OGC Web Services

Web Map Service (WMS)
Web Map Tile Service (WMTS)
Web Feature Service (WFS)
Web Coverage Service (WCS)
KML, GML, GeoPackage
GeoTIFF, NetCDF, HDF

CityGML
Emergency / Disaster Management

DigitalGlobe
Aviation Flight Information / Safety

Eurocontrol

OneGeology.Org


OpenIOOS.Org

Meteorology, Hydrology, Ocean Monitoring
OGC API Standards

Modular API building blocks to spatially enable Web APIs in a consistent way

- Modernized service architecture vs. OWS
- Leverages OpenAPI specification
- Current Web architecture and Spatial Data on the Web Best Practices,
- Focus on developer experience and usability
- Modular building blocks for fine-grained access to spatial data that can be used in data APIs,
- Open development; Public GitHub repo, Early implementations, In-depth validation, Slow release

OGC API - Features

OGC API - Coverages
OGC API - Map Tiles
OGC API - Processing
OGC API – Features – Part 1: Core

Resources of the Web API

Table 1. Overview of resources, applicable HTTP methods and links to the document sections

<table>
<thead>
<tr>
<th>Resource</th>
<th>Path</th>
<th>HTTP method</th>
<th>Document reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landing page</td>
<td>/</td>
<td>GET</td>
<td>7.2 API landing page</td>
</tr>
<tr>
<td>Conformance declaration</td>
<td>/conformance</td>
<td>GET</td>
<td>7.4 Declaration of conformance classes</td>
</tr>
<tr>
<td>Feature collections</td>
<td>/collections</td>
<td>GET</td>
<td>7.12 Feature collections</td>
</tr>
<tr>
<td>Feature collection</td>
<td>/collections/{collectionId}</td>
<td>GET</td>
<td>7.13 Feature collection</td>
</tr>
<tr>
<td>Features</td>
<td>/collections/{collectionId}/items</td>
<td>GET</td>
<td>7.14 Features</td>
</tr>
<tr>
<td>Feature</td>
<td>/collections/{collectionId}/items/{featureId}</td>
<td>GET</td>
<td>7.15 Feature</td>
</tr>
</tbody>
</table>

http://docs.opengeospatial.org/DRAFTS/17-069r1.html#tldr
Indexes 2D space as regular grids with scales in a projected CRS

- Path = /collections
  - Returns metadata describing \collections at this API

- Path = /collections/{collectionId}
  - Returns metadata describing {collectionId}

- Path = /tileMatrixSet
  - Returns all available tile matrix sets

- Path = /tileMatrixSet/{tileMatrixSetId}
  - Returns a tiling scheme by id
OGC API - Coverages

• Geospatial Coverage: “spatial function” or “field”, Spatial domain to Values Range

• OpenAPI = URL-oriented request language, aligned across OGC Coverage standards
  
  - download coverage c001

    http://acme.com/oapi/collections/{collectionid}/coverages/c001

  - lat/long cutout, time slice \( t=2009-11-06T23:20:52 \)

    http://acme.com/oapi/collections/{collectionid}/coverages/c001?
    SUBSET=Lat(40,50)&SUBSET=Long(10,20)

  - coverage c001, in GeoTIFF": either http accept header, or:

    http://acme.com/oapi/collections/{collectionid}/coverages/c001?
    F=“image/tiff“
Geospatial Data and Processing in Apache Projects

Discrete Global Grid Systems
Discrete Global Grids

Slide source:
Matthew B. J. Purss,
Robert Gibb, Faramarz
Samavati, Perry
Peterson, Jin Ben,
Roger Lott
Discrete Global Grid Systems

“...a spatial reference system that uses a hierarchical tessellation of cells to partition and address the globe. DGGS are characterized by the properties of their cell structure, geo-encoding, quantization strategy and associated mathematical functions.”

- OGC DGGS Standard
Standardising Discrete Global Grid Systems

Different Cell Shapes

- Square = Familiar
- Triangular = Fast
- Hexagonal = Fineness of Fit

Unique Cell Indices

- Hierarchy-based, Space-filling Curve, Axes-based or Encoded Address

nD Spatial Analyses

↓

1D Array Processes
Hexagons!

Geospatial and Temporal Forecasting at Uber
September 09, 2019 Apachecon
Chong Sun, Brian Tang
Geospatial Data and Processing in Apache Projects

Geovisualization
3D Geospatial Visualization

The approach: Support multiple 3D data formats

<table>
<thead>
<tr>
<th>X3D</th>
<th>glTF</th>
<th>3D Tiles</th>
<th>I3S</th>
<th>KML</th>
</tr>
</thead>
</table>

OGC 3D Portrayal Service

Content Delivery
Query

Berlin with 3D and Textures for Visualization

New York City portrayal of attributes

OGC®
Indexed 3D Scene Layer (I3S)

- Storage and transmission of large, heterogeneous 3D geospatial data sets
- 3D geospatial content, various coordinate systems along with a rich set of layer types
- Expandable to accommodate new data types and access patterns

- Developed by Esri now also an OGC Community Standard

- 3D Objects
- Points
- Integrated Meshes
- Point Clouds
OGC Community Standard
Massive heterogeneous 3D geospatial datasets with semantics

Spatial data structures + glTF + Styling + Metadata

Photogrammetry
Point clouds
3D buildings
Terrain
Ecosystem

Exporters / Tilers

Visualization engines

Validator in-progress

Built on

Slide source: Cesium at OGC TC
Geospatial Data and Processing in Apache Projects

Geospatial Trends
Location Powers: Data Science

13th & 14th November 2019
Google, Mt View, CA

• Sessions
  1. Foundations
  2. Analytics
  3. Ripe Trends
  4. Outcomes
  5. Actions to Take

• Explosive availability of data on every aspect of human activity; with revolutionary advances in computing technologies is transforming geospatial data science.

Data Science – source NIST

http://locationpowers.net
GeoSpark

http://datasystemslab.github.io/GeoSpark/

• Geospatial DBMS on top of Apache Spark
• Spatial RDD, Spatial SQL, Spatial DataFrame
• Distributed geospatial visualization
• Interactive SQL/map visualization (with Apache Zeppelin)

"GeoSpark comes close to a complete spatial analytics system. It also exhibits the best performance in most cases."

"How Good Are Modern Spatial Analytics Systems?" Varun Pandey, Andreas Kipf, Thomas Neumann, Alfons Kemper, PVLDB 2018

Google “GeoSpark ASU”

In production!

8K - 10K monthly downloads
Geospatial Track - ApacheCon NA 2019

0900 Geospatial Data and Processing - Reusable Building Blocks
  • George Percivall, OGC

1000 Geospatial Data Management in Apache Spark
  • Jia Yu, ASU & Mohamed Sarwat, ASU

1130 Apache Science Data Analytics Platform Apache (SDAP)
  • Frank Greguska, JPL

1400 Using GeoMesa on top of Accumulo, HBase, Cassandra, and big data file formats for massive geospatial data - a LocationTech Project
  • James Hughes, CCRI

1500 Geospatial Indexing and Search at Scale with Apache Lucene
  • Nick Knize, Elastic

1630 Geospatial and Temporal Forecasting in Uber Marketplace
  • Chong Sun and Brian Tang, Uber

1730 Realtime Geospatial Analytics with GPUs, RAPIDS, and Apache Arrow
  • Josh Patterson, NVIDIA

Join geospatial@apache.org