Data Visualization in Galileo

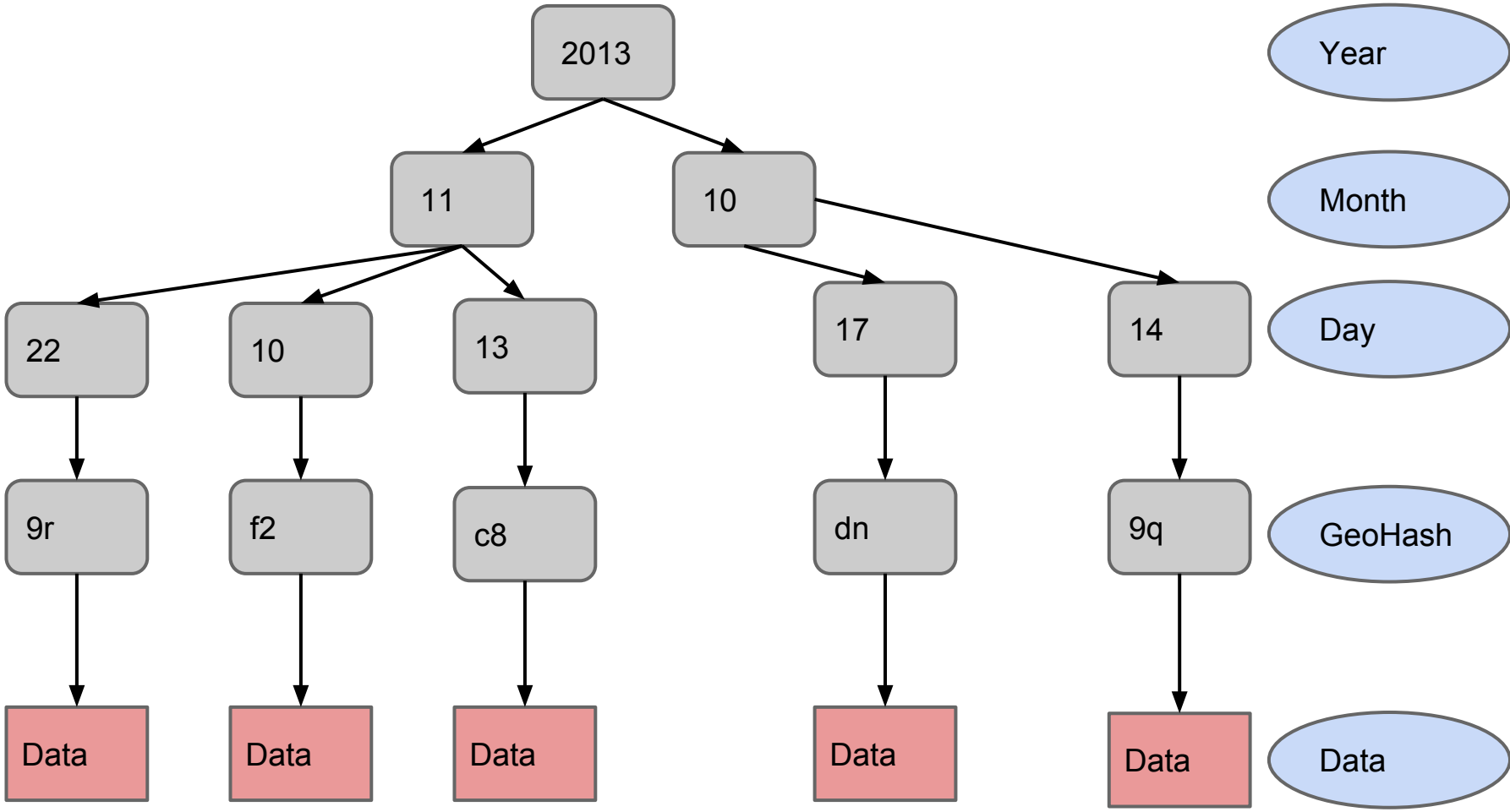
Jared Koontz

16th January 2014

Introduction

- Traditional visualization tools often inadequate with Big Data.
- Perceptual and interactive scalability
- Limited by resolution of visualization, not number of records

Introduction to the Data Model



Aside: Geohashing

- A simple way to encode latitude and longitude and grouping nearby points on the globe
- Single simple string representation. Can be easily saved and transferred around
- Points with same prefix are close by, no advanced algorithms needed.

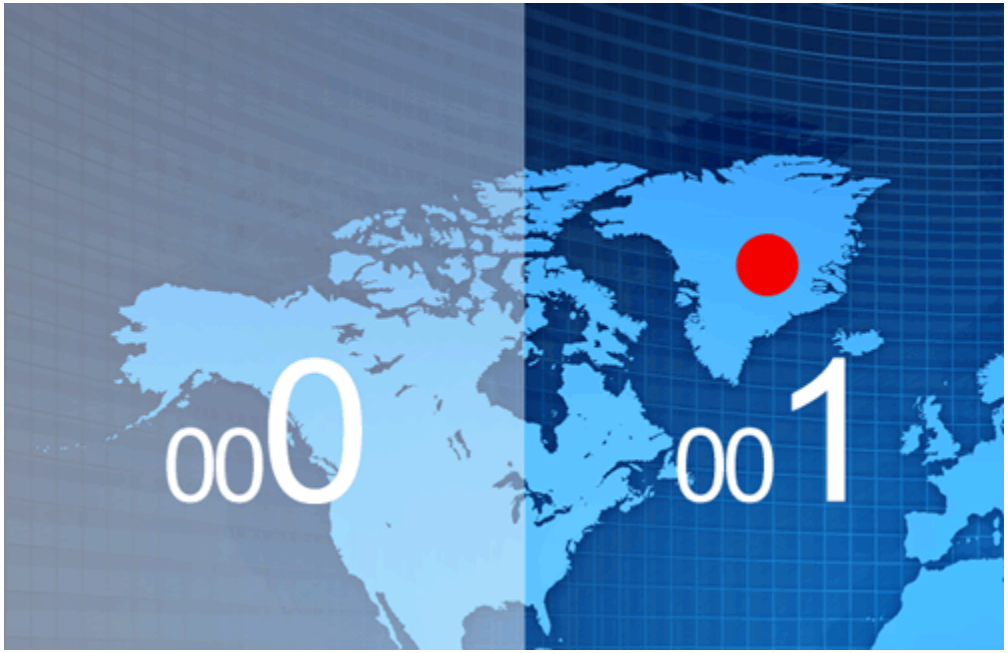
Aside: Geohashing Example

- Suppose we want to get the geohash for this red dot.

Aside: Geohashing Example

Aside: Geohashing Example

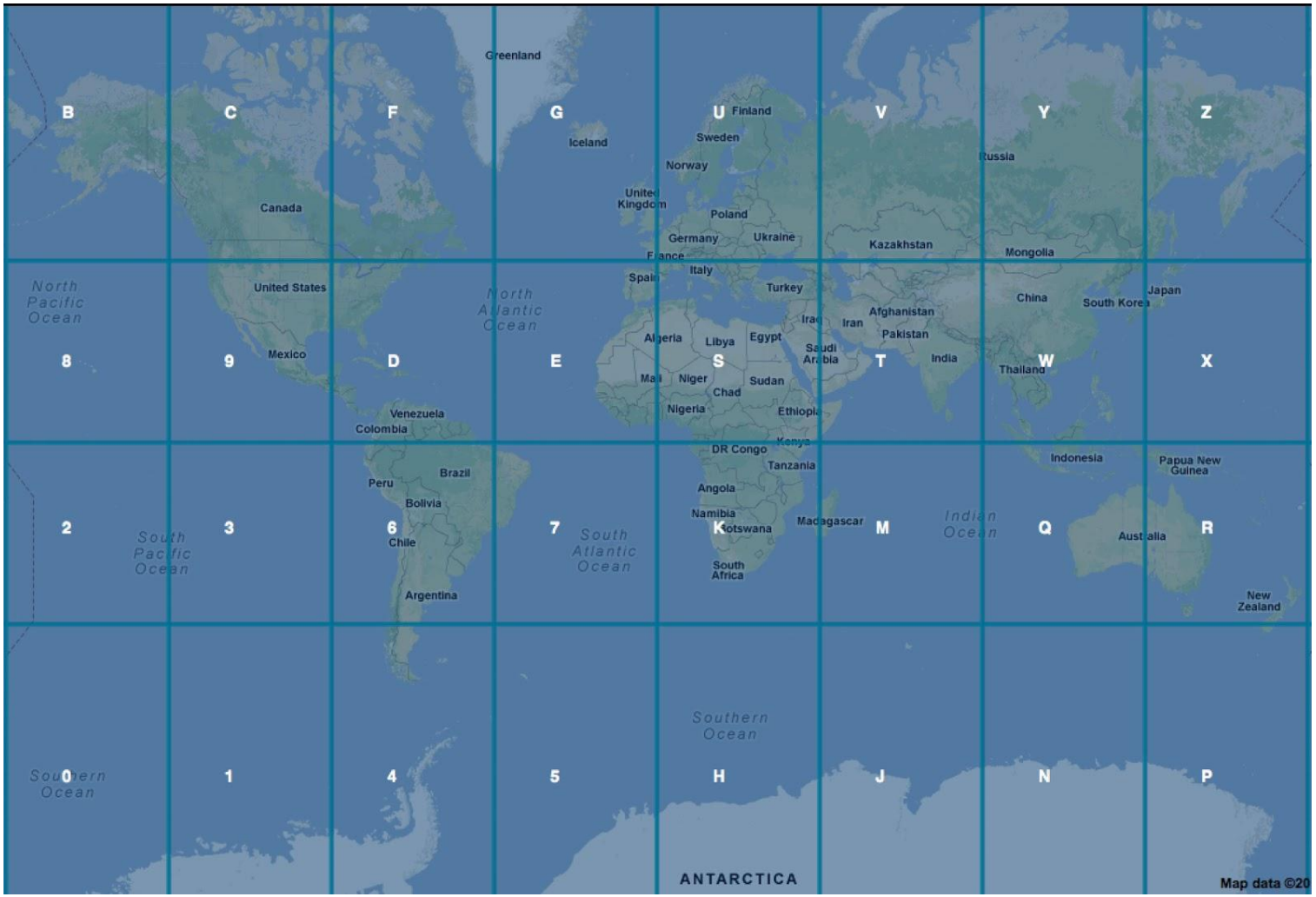
- We are getting closer to our location.



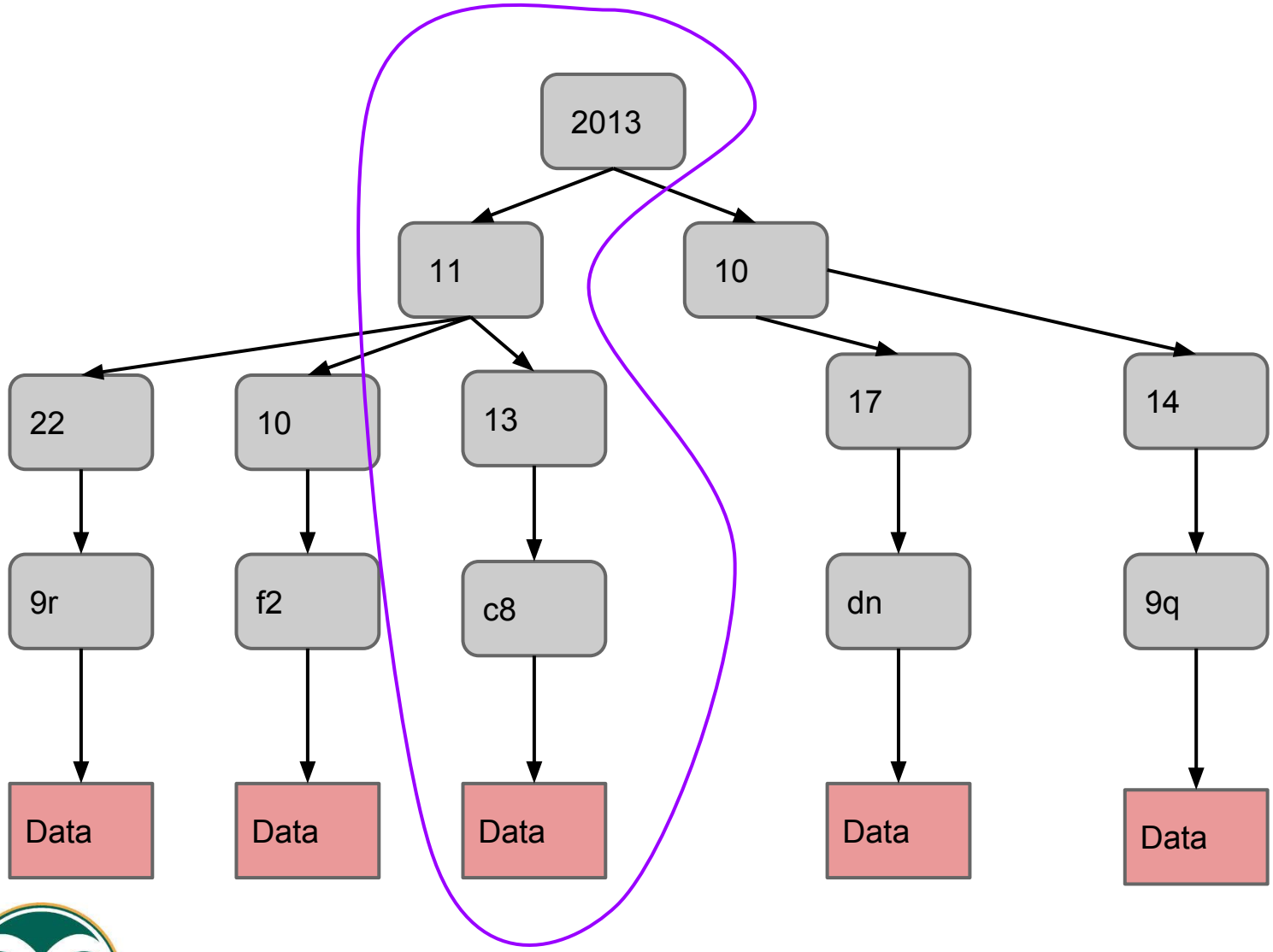
Aside: Geohashing

- We can continue this example out until we get a small enough bounding box for our needs.
  - With a 10 character geohash the next bounding box is only 0.6 meters away.
- If we encode every binary value of length 5 into a character, we get 32 possible 1 character geohashes.

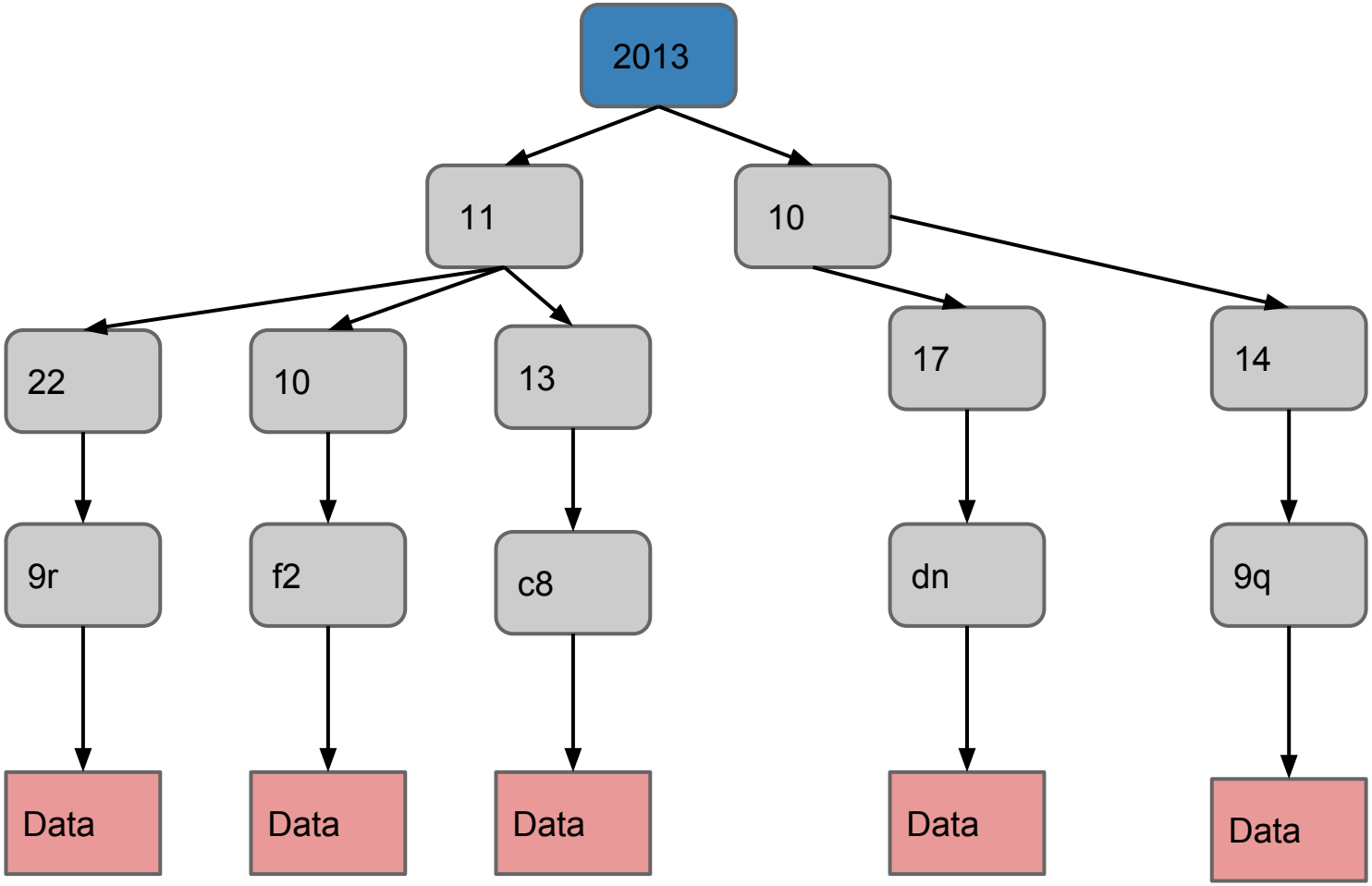
Aside: Geohashing



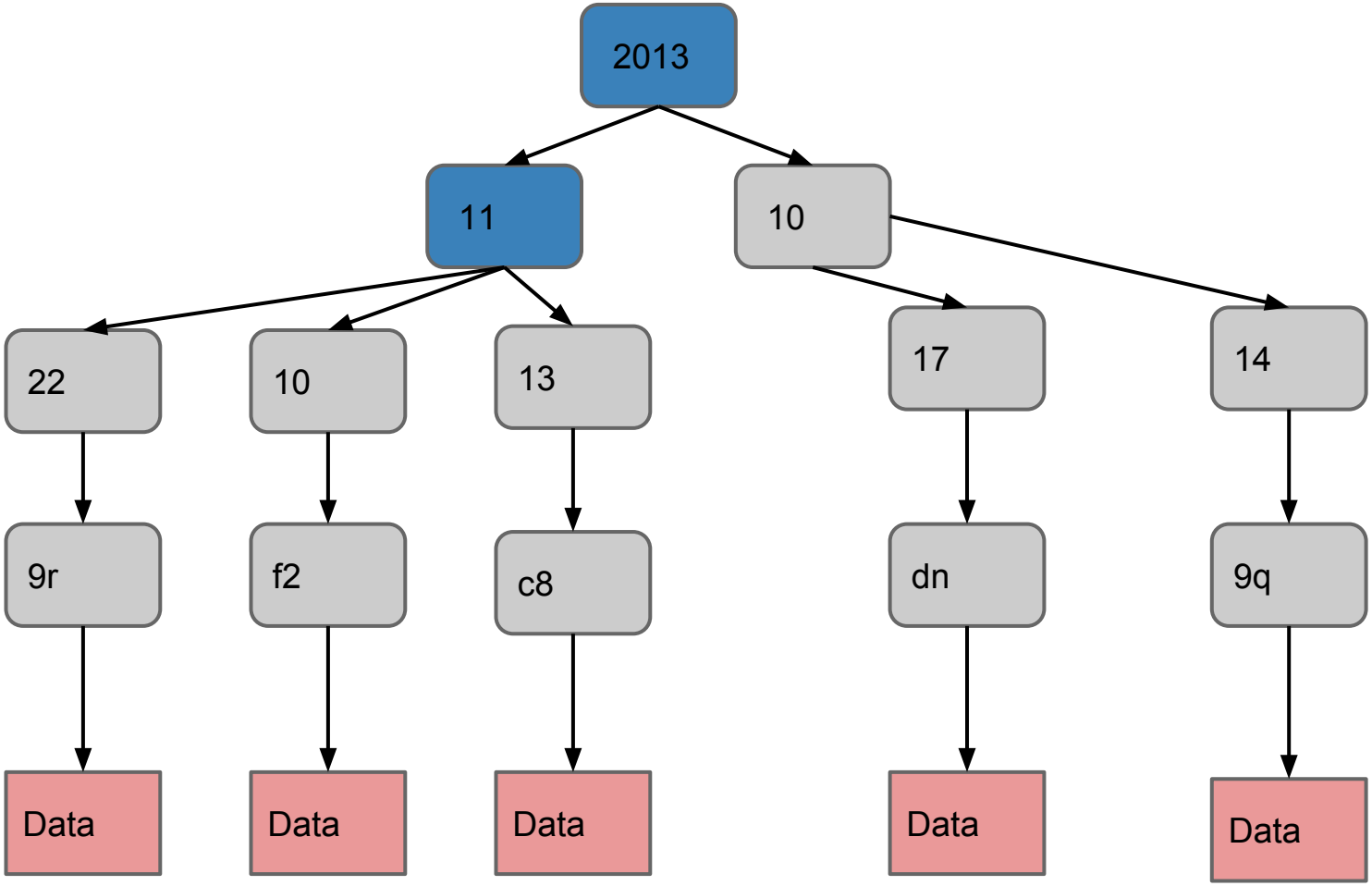
Data Model



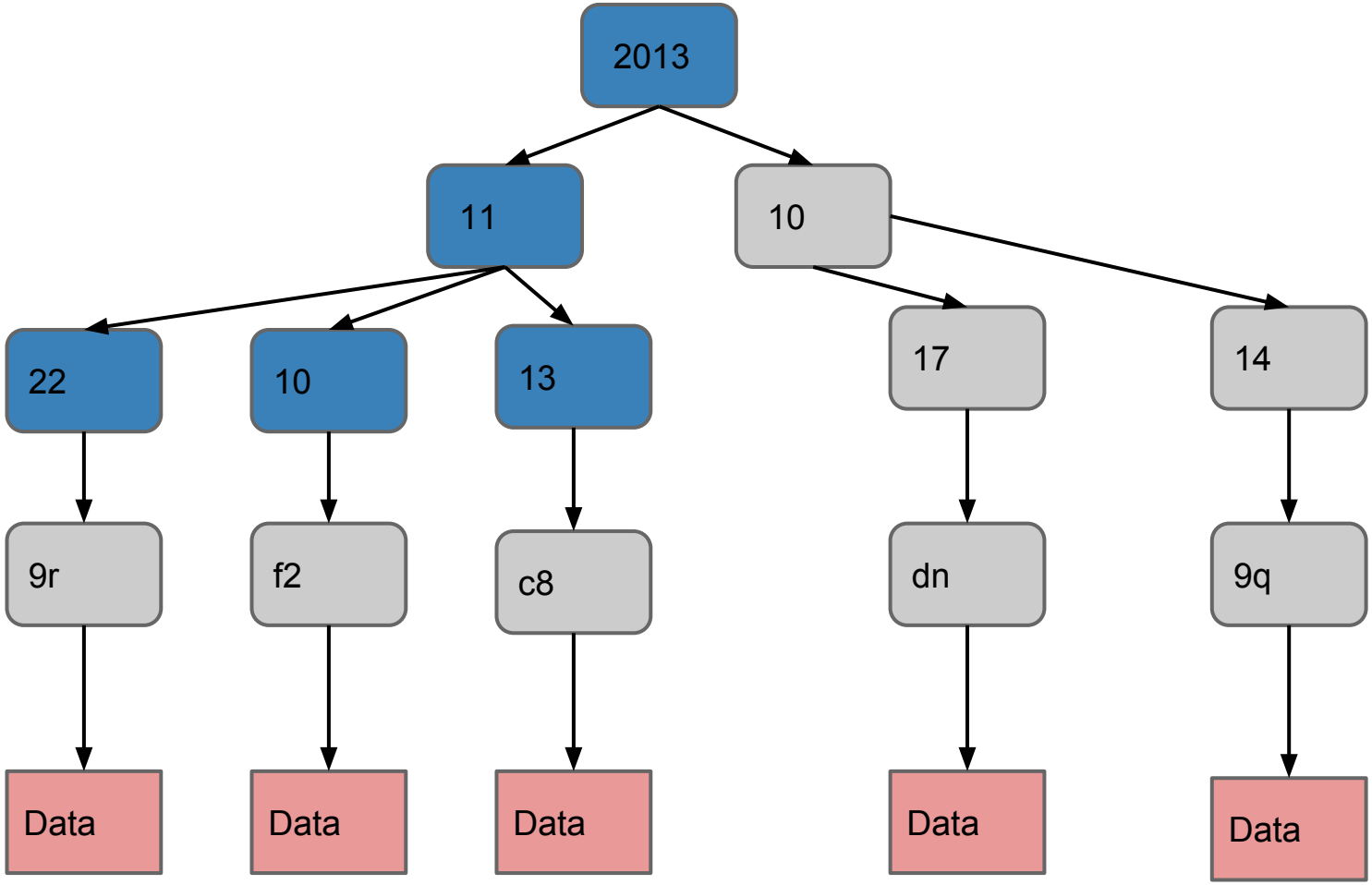
Data Model



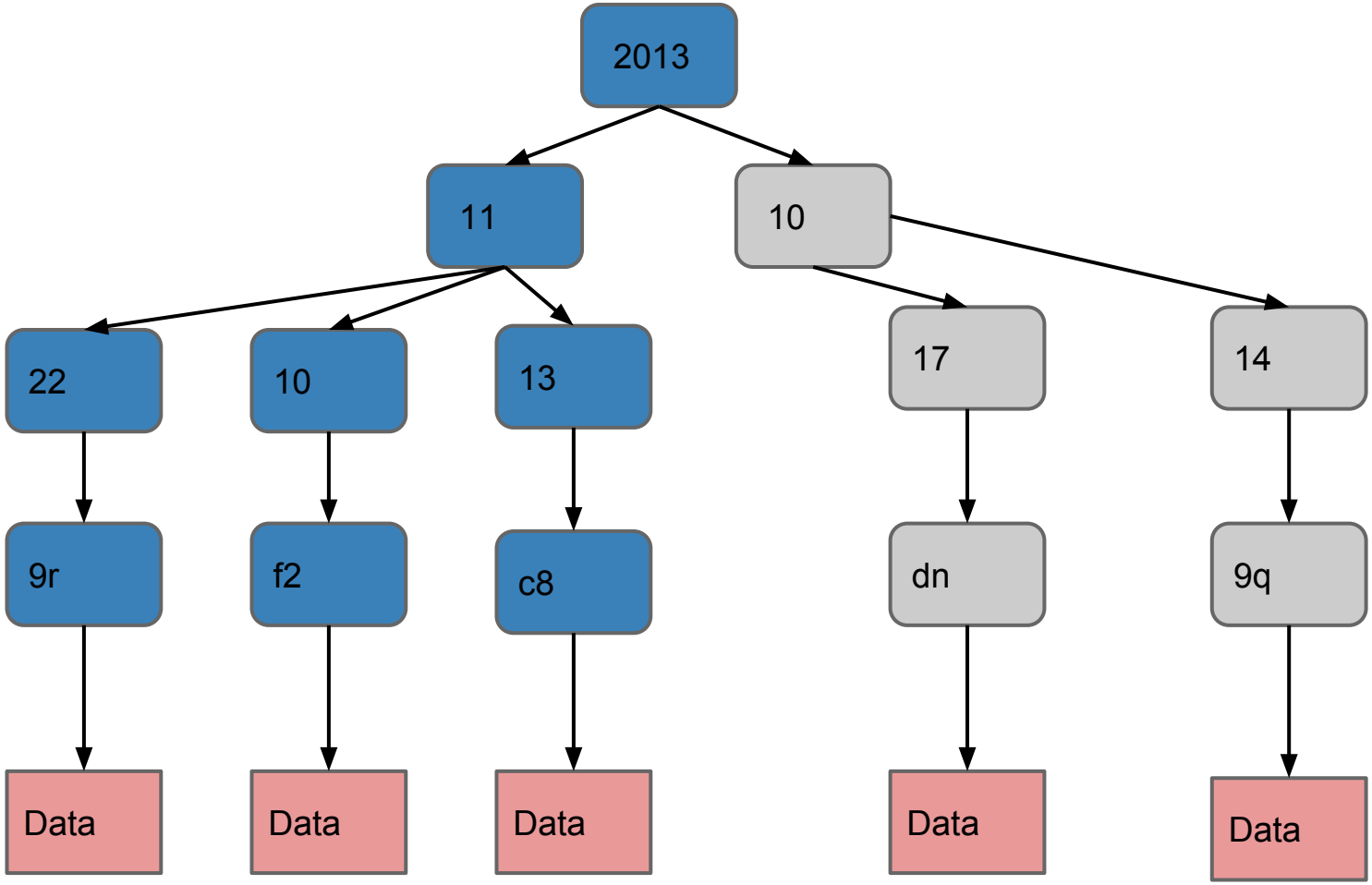
Data Model



Data Model



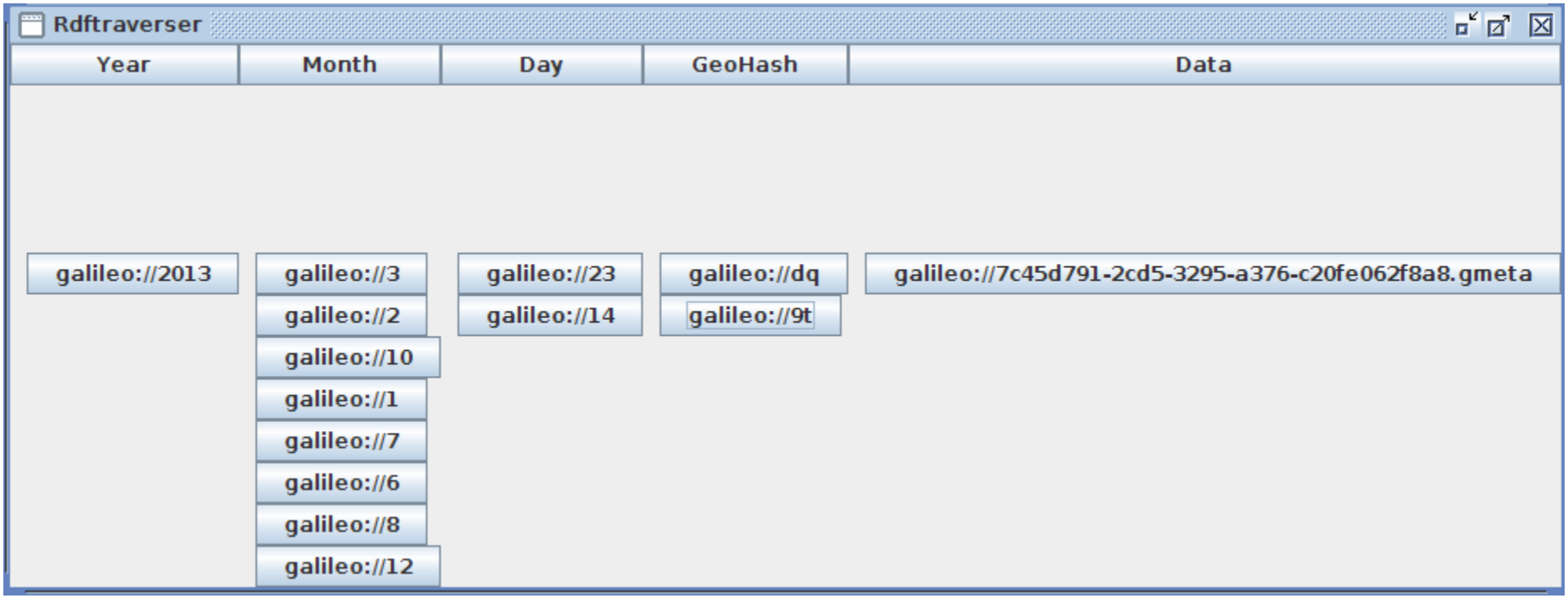
Data Model



Resource Description Framework

- Metadata graph can be transformed into an RDF representation
- RDF is a metadata model, composed of statements about resources in the form of subject-predicate-object expressions.

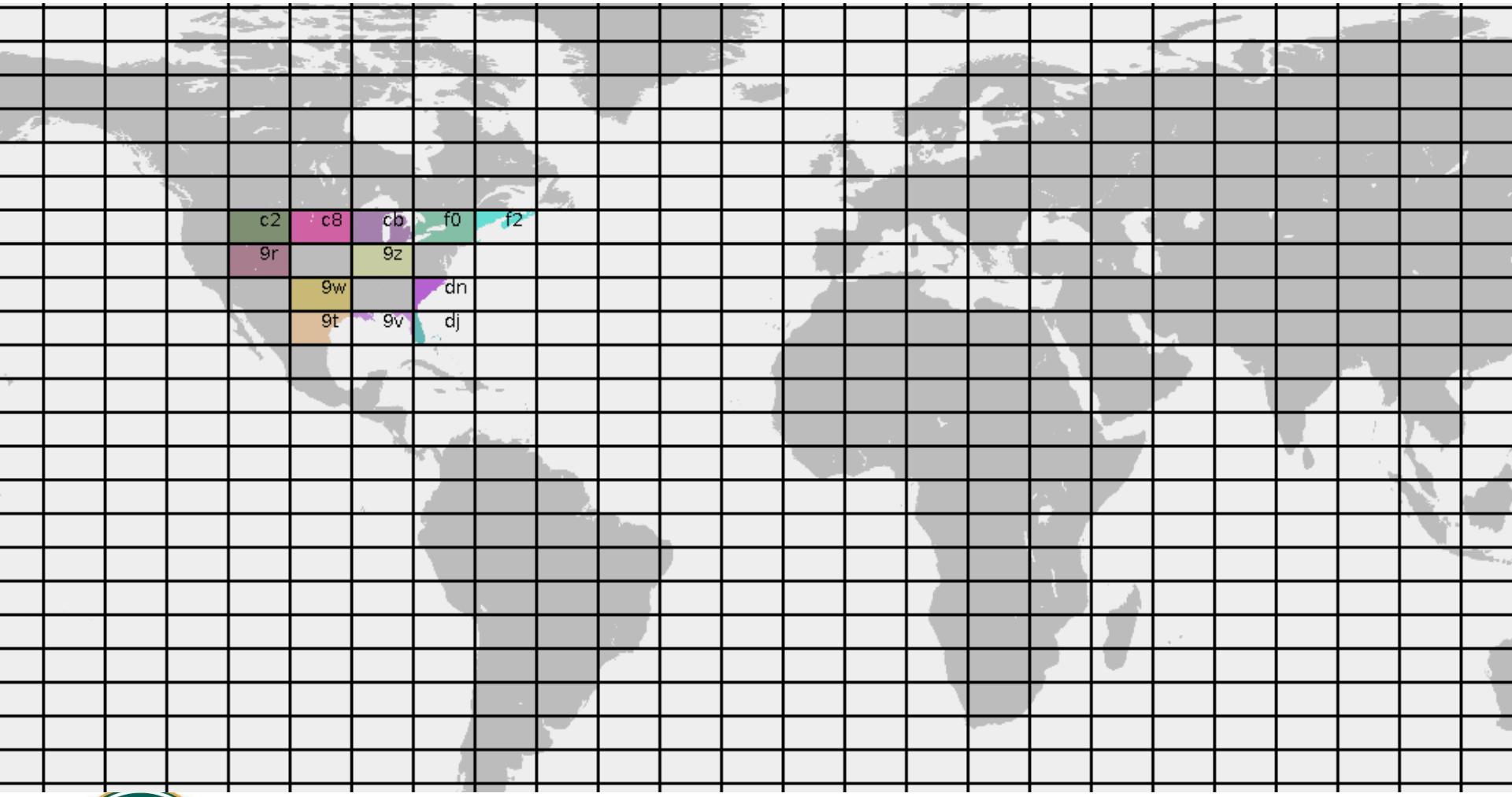
RDFtraverser



RDFtraverser

- Want to visualize data as well as metadata
- Not scalable for large datasets.

GridMap



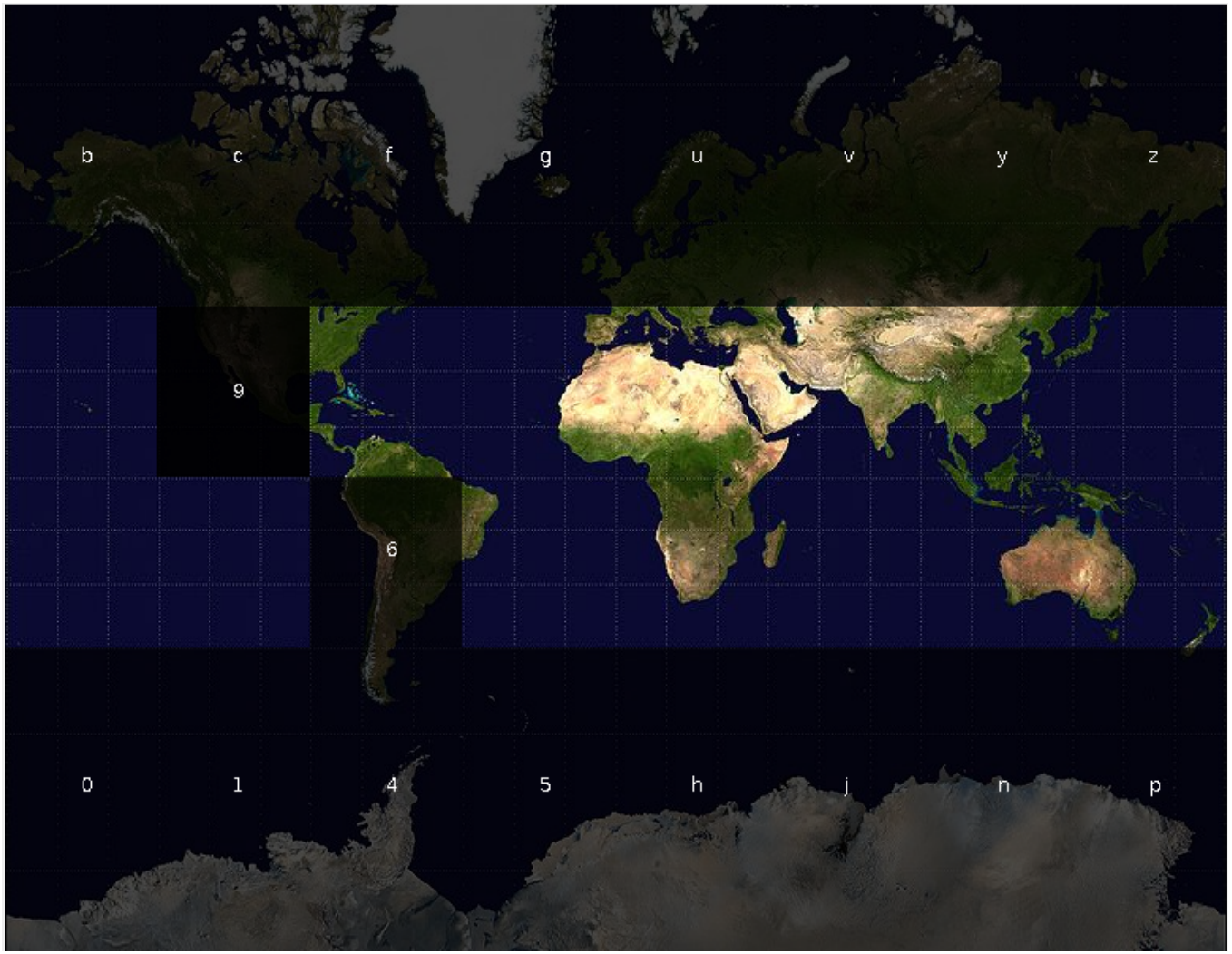
GridMap

- Wasted space storing unchanged image tiles.
- The world is a sphere.

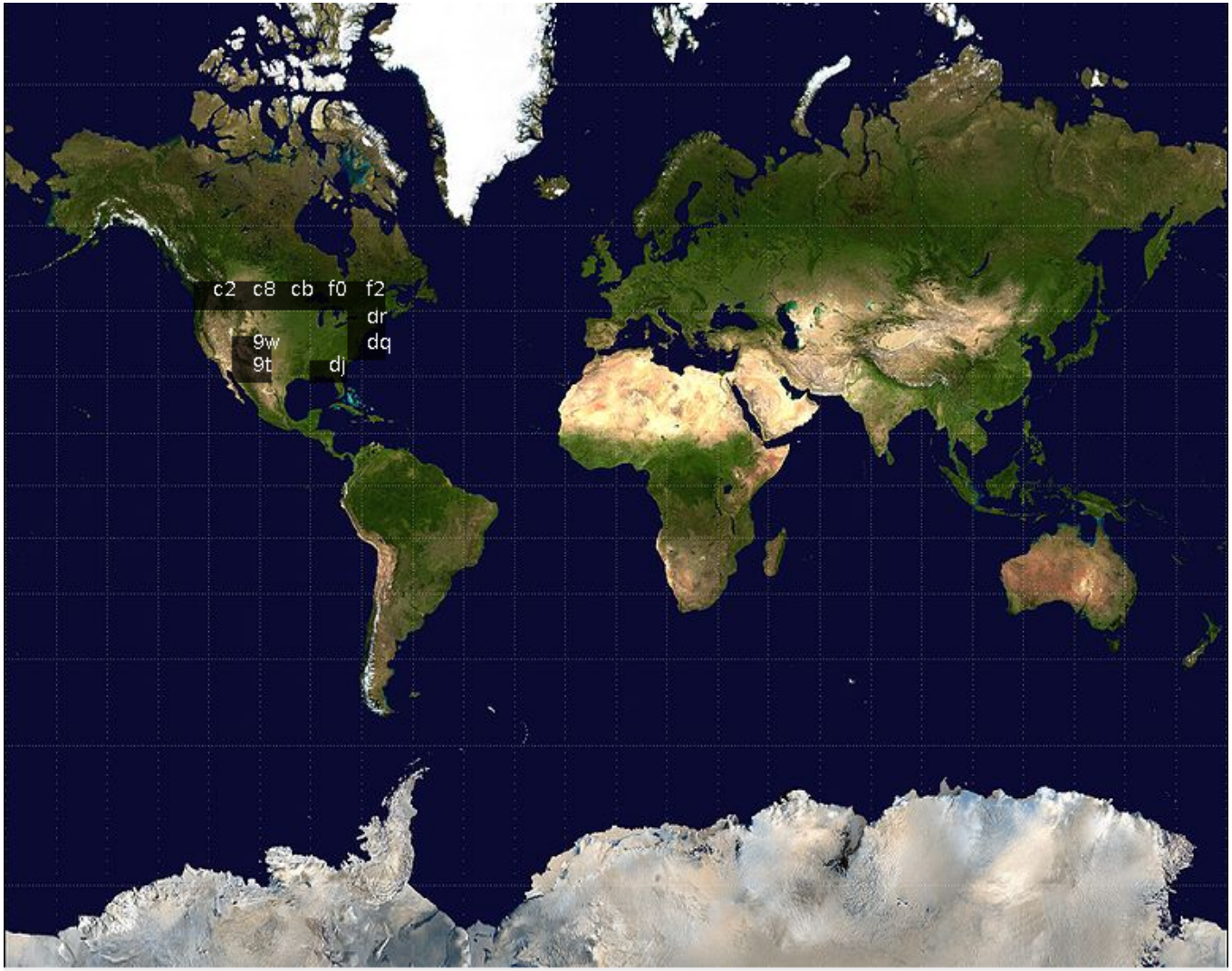
Mercator Projection

- Most common projection
- Geohash -> [Lat,Lon] -> [X,Y]
- Visualize data directly onto the image tiles, and then display the tiles.

Mercator Projection



Mercator Projection



Mercator Projection

- Changing path from root to leaf into a data tile.
- Can create these tiles on the same servers that hold the data. Reduce network I/O
- Send these tiles over the wire, or how to create these tiles.

Addressing Perceptual And Interactive Scalability

- Static map is not perceptually or interactively scalable
- Allow the user to zoom in and zoom out, as well as focus on different areas
- http://www.openstreetmap.org/
  - Makes network requests to get tiles

Interactive Scalability: Creating Queries

- Want to make it easy for the user to query the system

- Select area by drawing a polygon, rather than inputting latitude or longitude coordinates.

Data Flow

1. User creates query through the client side interface

2. Query is sent to Galileo

3. Galileo creates image tiles from query and stores them in the tree like data structure
   a. This will be a different branch than the raw data.
   b. Used to enable brushing and linking

Data Flow

4. Statistics about data will be represented in the image tiles

5. Send tiles back to user to display in interactive map.

Bibliography

- RDF information from http://en.wikipedia.org/wiki/Resource_Description_Framework

- Pictures for geohash from http://www.bigfastblog.com/geohash-intro and http://www.bigdatamodeling.org/2013/01/intuitive-geohash.html

Thank you

Any Questions? Comments?