Swiss Re

General Public Release

GIS and Risk Rating provide Smart Analytics

Dr. Octavian Iercan - SwissRe

SWISS RE
150
YEARS

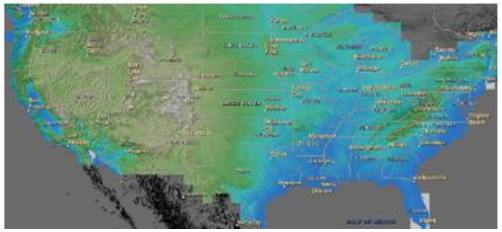
Importance of geographic information in the insurance industry

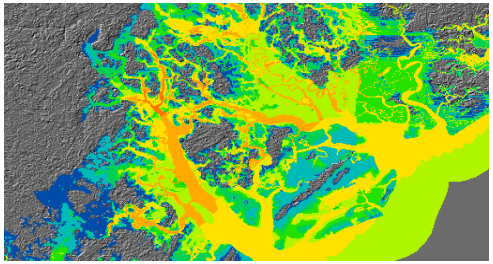
- Location is key to determine the exposure of a risk
- Available geographic information gets more and more detailed (country – province / cresta – zip – address – lat/lon)
- Heterogeneous data quality on a global scale
- Even an exact address might not be the best solution
- Use of GIS tools for risk assessment, underwriting and portfolio management is key
- Global hazard information is hard to get
  → Swiss Re produces such data and shares its knowledge with clients

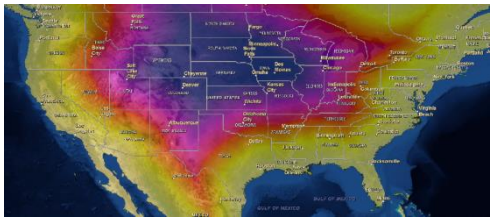
GIS in Swiss Re

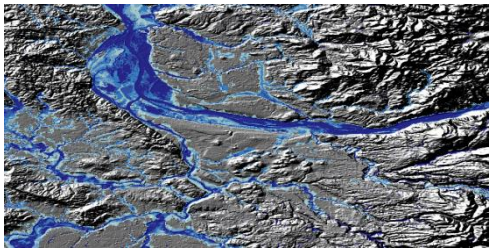
GIS is an essential component to develop in-house models

- Modernize the applications for Cat Modelling using GIS
- Constantly upgrade to high resolution risk data (GRID)





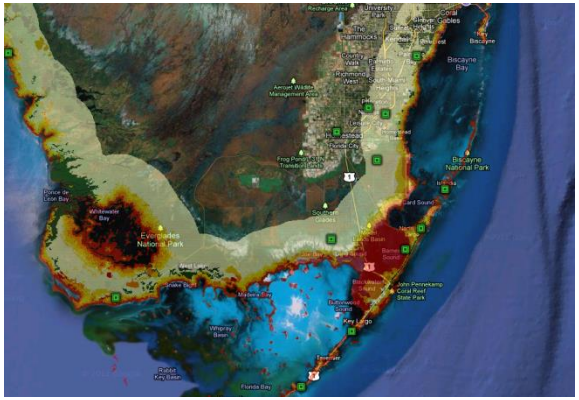


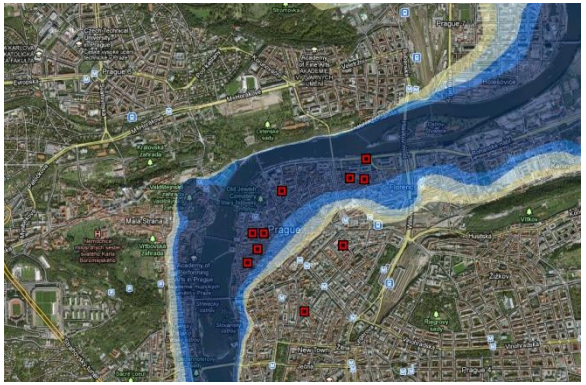


Some examples:
1) Elevation
2) Surge modelling
3) Distance to coast impact
4) Flood zone calculation based on DTM

GIS is instrumental to build and use models

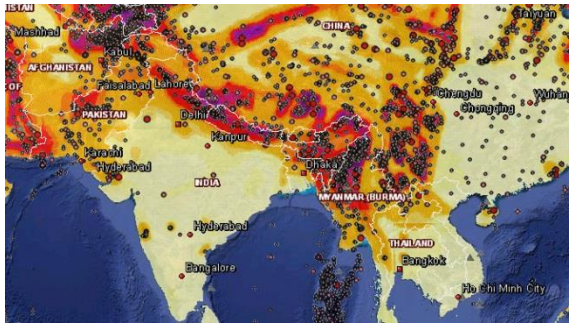
- Where are my insured objects?
  - *GIS: Encoding of addresses*
- What is the risk of natural catastrophes?
  - *GIS : Look up of risks (e.g. flood zones)*
- In which country, province, municipality are my insured objects located?
  - *GIS : Providing geo-information for the insured risks*

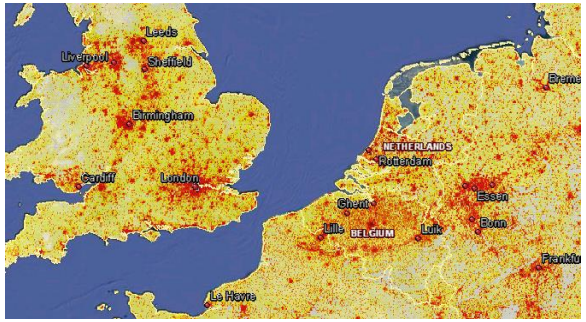




Further GIS related ingredients for risk assessment

- Detailed satellite imagery
- Nat Cat risks maps
- Terrain data
- Population data
- Industry risk information

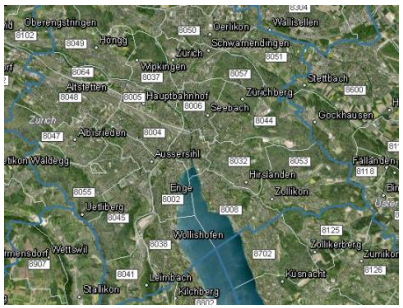




GEO Applications in Swiss Re

Swiss Re's central GEOdatabase

- One company, one GEOdatabase!
- GEOdatabase is Master Source for GEOdata
- Worldwide data provided
- Spatial & attribute data
- All major Swiss Re applications use GEOdatabase



Web application: the Hazard Atlas

Swiss Re's online natural hazard information and webmapping visualization tool

It comes in two versions:

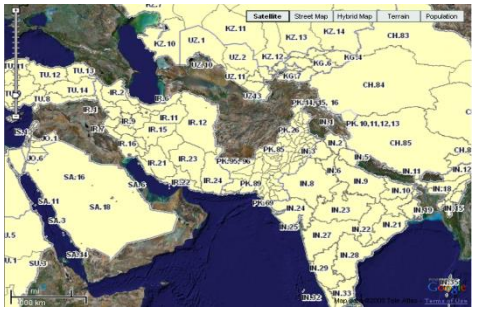
- GEOportal for Swiss Re employees
  - connected to pricing models and tools
- CatNet® for Swiss Re clients

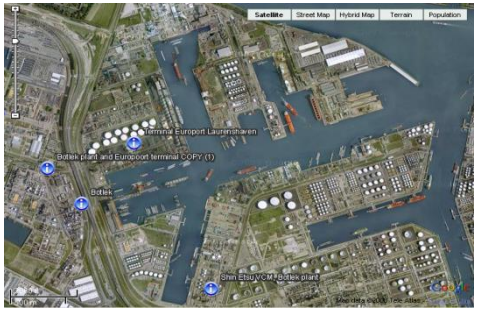
It combines:

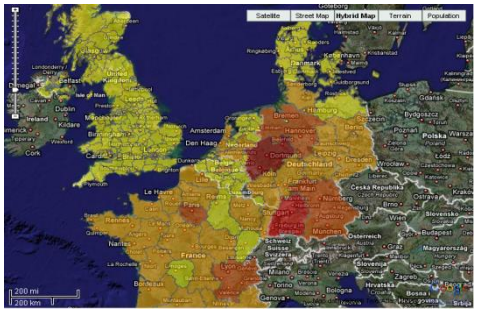
- Nat Cat data
- Administrative / postal data
- Insured risks
- Population data
- Satellite imagery
- ...

with analysis functionality and Swiss Re modelling

- Based on Javascript API for Google Maps
- Maps server by ESRI ArcGis server





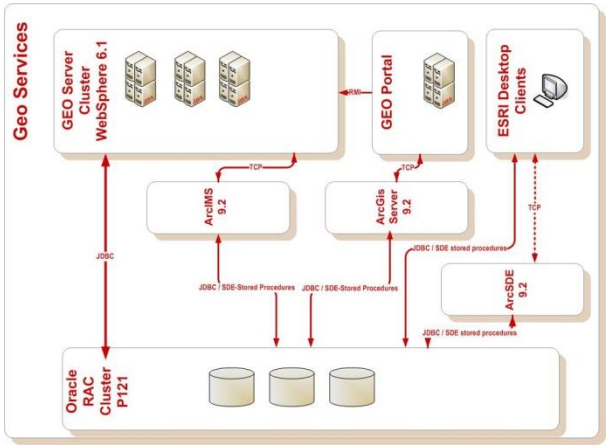


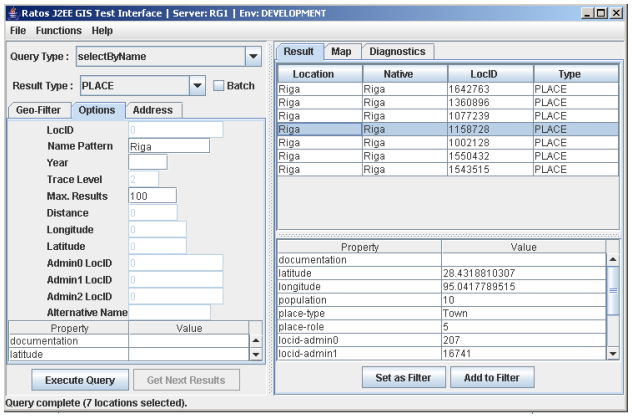
CatNet® WMS: clients integrate our services in their Web Applications

Service Name	Content	Service Type
Coastal Flood	Visualization	Map Service
Hailstorms	Visualization	Map Service
River Flood	Visualization	Map Service
Seismic Hazard	Visualization	Map Service
Tornadoes	Visualization	Map Service
Tsunami	Visualization	Map Service
Wildfires	Visualization	Map Service
Wind Speed	Visualization	Map Service
Risk Look Up	Intersection of one input point with the CatNet data	Geoprocessing Service
Risk Look Up Batch	Intersection of multiple input point with the CatNet data	Geoprocessing Service

SwissRe own GEOServer

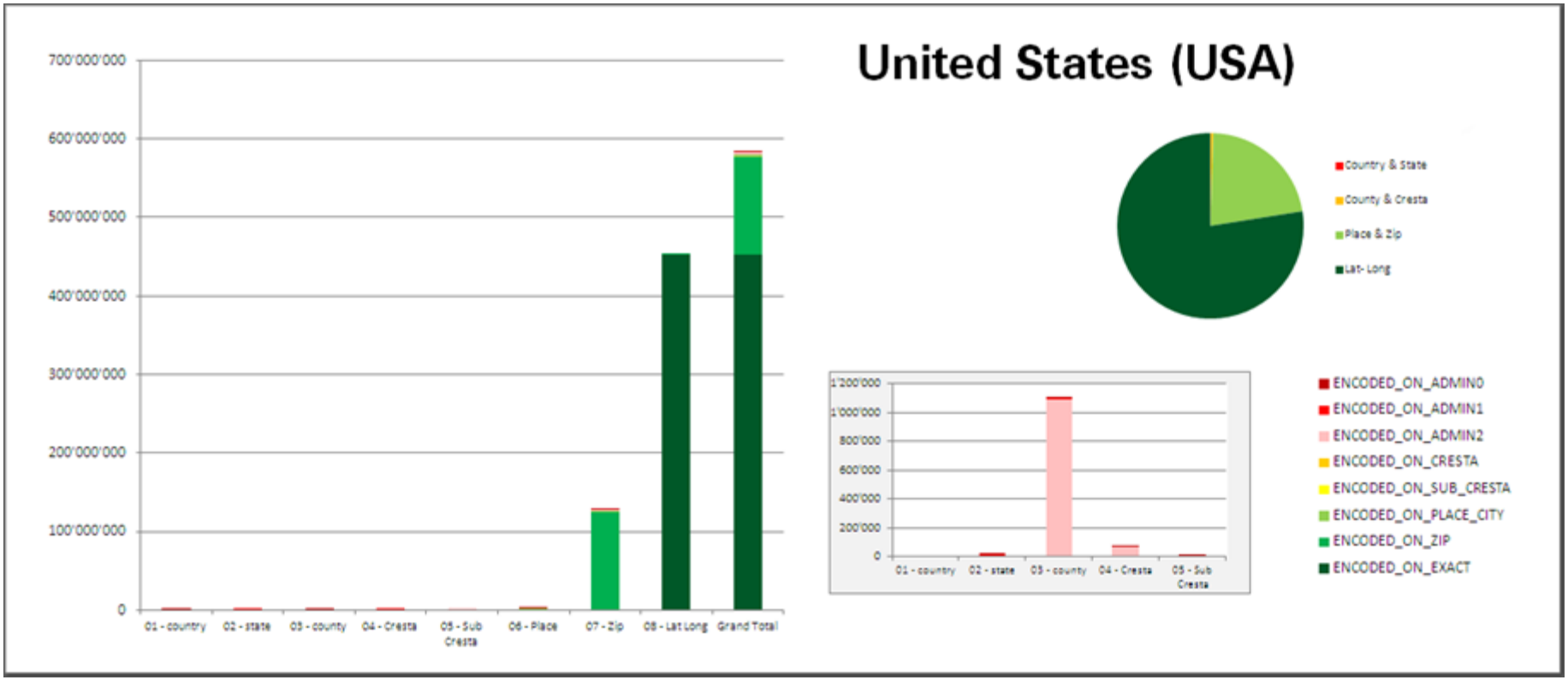
- GEOServer provides data from the Geodatabase (Oracle) fast and convenient to client applications
- Performance of GEOServer is improved by indexing the content using Quadtree methodology
- The GEOServer is a Java application
- GEOServer is running on 16 different servers
- Client applications use our Web services
- Street level encoding is done using Google Maps web services



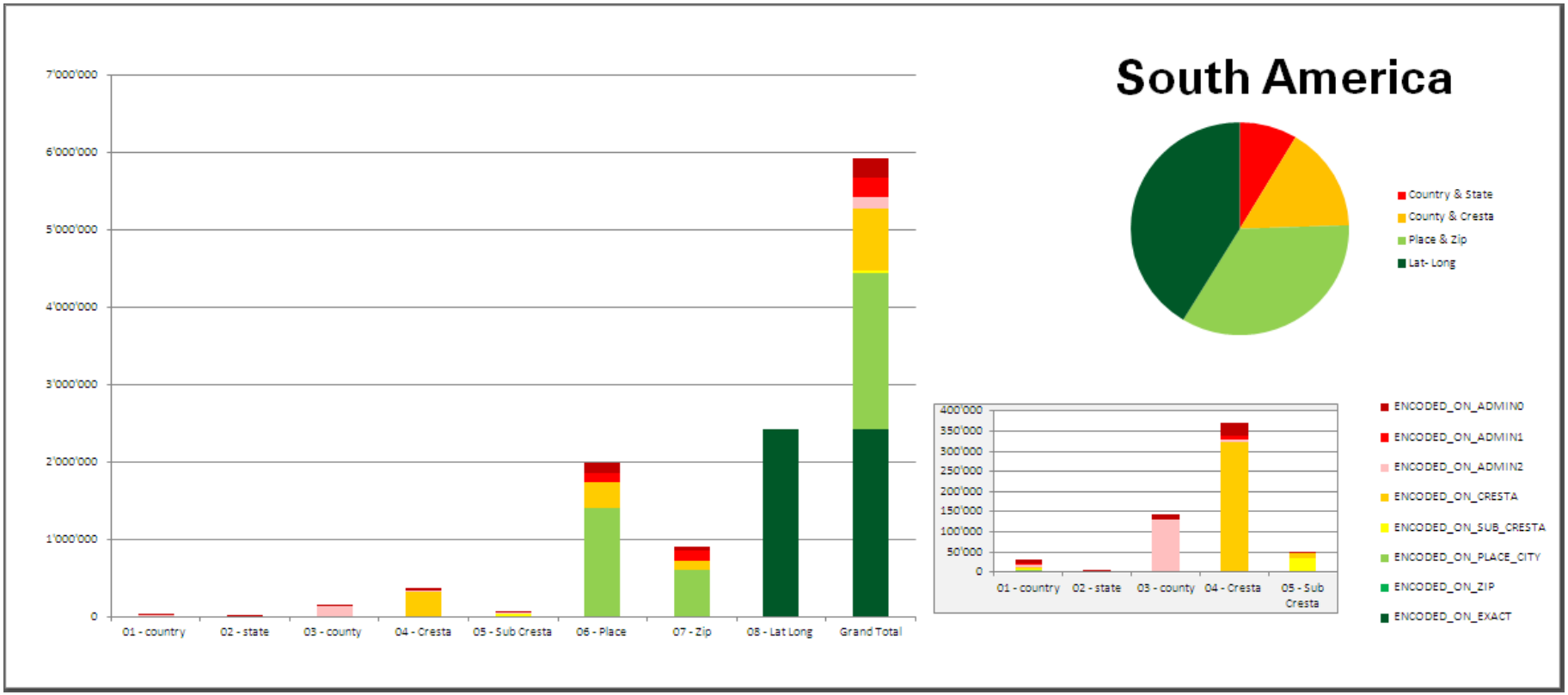


Geocoding and Visualization for accumulation control

Geocoding quality excellent in the US



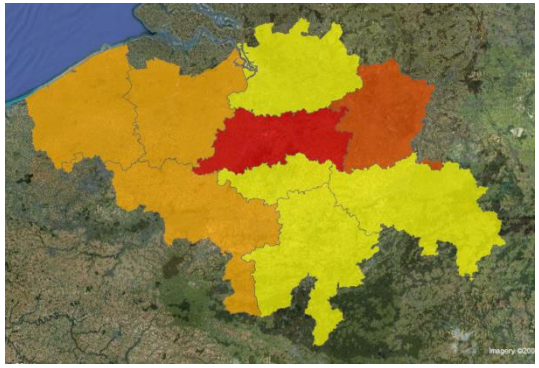
Geocoding quality differs in South America

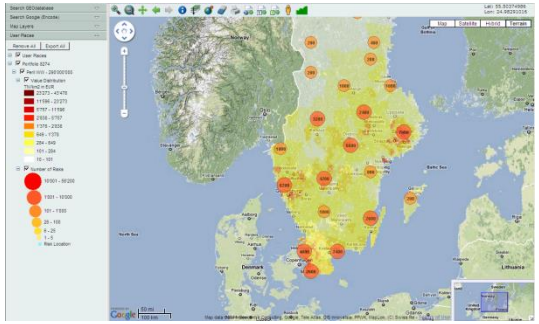


Solution → We upwards encode to the next available quality

GIS vs NonGIS visualization of model output

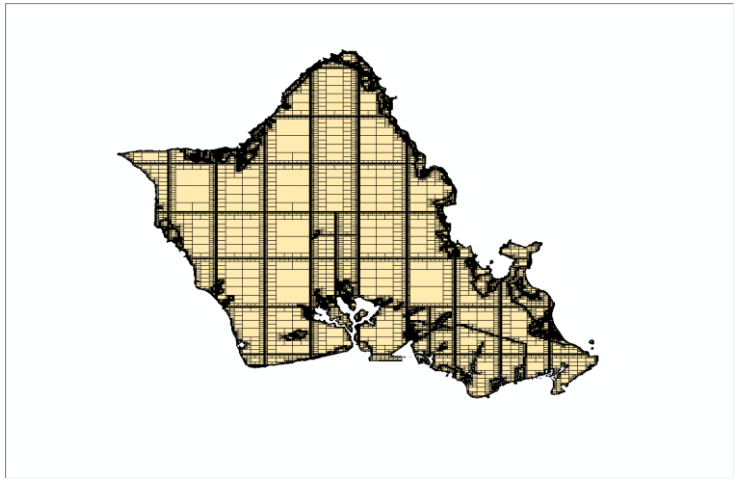
- Summary of insured values by geographic region
- Control of insured values by geographic region or possible event
- Detection of accumulation hotspots

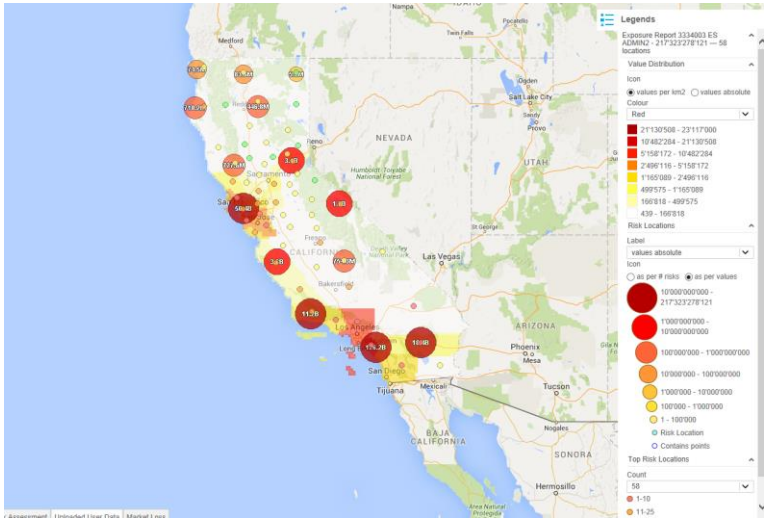




Risk value accumulation with ValueMap

- Visualizing large amounts of data in a very short time is the goal of our Value Map
- In a Nutshell it is a way to display a certain exposure value we receive in an EDM (Exposure Data Model) from a client
- The values are distributed using our Quadtree indexing and like this we can both visualize exposure on features such as counties and square km





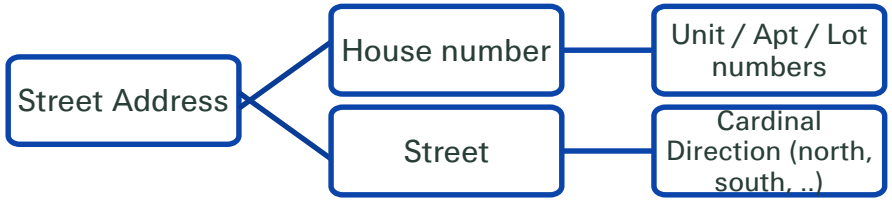
Data Intelligence on EDM Data imports

Exposure Data Intelligence

- A record linking solution was built, that …
  - isolates duplicate risks and gives every distinct risk a separate identifier.
  - can be used to construct a new database of **unique synthetic risks**
- The resulting synthetic risks can be further enriched with external data and used
  - The multi-step record linking algorithm is deployed using the Apache Spark framework for distributed computing.
  - Example is based 68 million risks in Florida use case

- Feature Extraction used the following features

Admin1	Admin2	Admin3	Place
Street Address	Zip code	Latitude	Longitude



Exposure Data Intelligence

Key Findings

- Risk Clusters in Miami FL
  - The bubbles represent the locations of duplicate risks within all of the 2014 treaty portfolios.
  - The size of the bubble scales with the number of identical risks at a specific location

- Running our algorithm on all 68 million risks in Florida (2014 portfolios), we find:
  - **6 million** original **risks** not clustered and thus **truly unique**
  - **4.5 million risk** synthetic **clusters** containing at least two identical risks
  - i.e. 4.5 + 6 = 10.5 million unique risks. **85% of risks are duplicates**

CL ID	ID	pd_v_id	int_itm_id	admin0	admin1	admin2	place	zip	lat	long	Add
133	964960	963587	5311713220	339	12048	316580	0	33015	(25.94	-80.304)	18264 MDITRRNEAN BLVD
133	1874785	963099	5303480320	339	12048	316580	0	33015	(25.94	-80.304)	18264 MDITRRNEAN BLVD
133	40182758	1076562	6481194874	339	0	0	0	33015	(25.94	-80.304)	18264 Mediterranean Blvd
133	3167033	963097	5303469215	339	12048	316580	0	33015	(25.94	-80.304)	18264 MDITRRNEAN BLVD
133	6415386	1005297	5727306120	339	12048	316580	0	33015	(25.94	-80.304)	18264 MDITRRNEAN BLVD
133	26907614	1005295	5727288904	339	12048	316580	0	33015	(25.94	-80.304)	18264 MDITRRNEAN BLVD

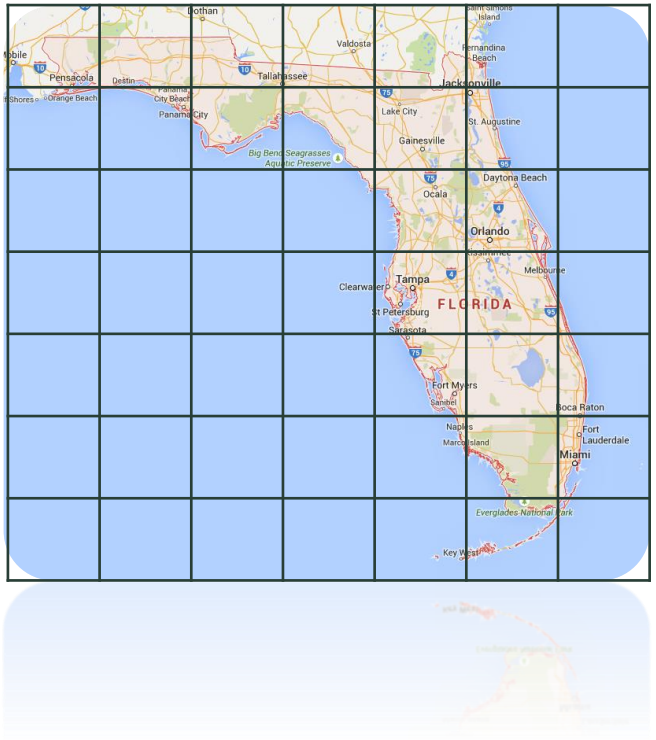
The record linking algorithm in detail

Candidate Finding & Pair-Wise Comparison

- We use a grid partitioning of geographic coordinates as pre-filter to **establish candidates**
  - 1000 x 1000 cells with 10% overlap
  - use an efficient nearest-neighbor algorithm to compute for each risk its *n* nearest other risks

- All candidates are **pair-wise compared** using all features.
  - Levenshtein edit-distance used for address comparison
  - Combined matching probability is computes (using Bayes Theorem and the assumption of independence of features)

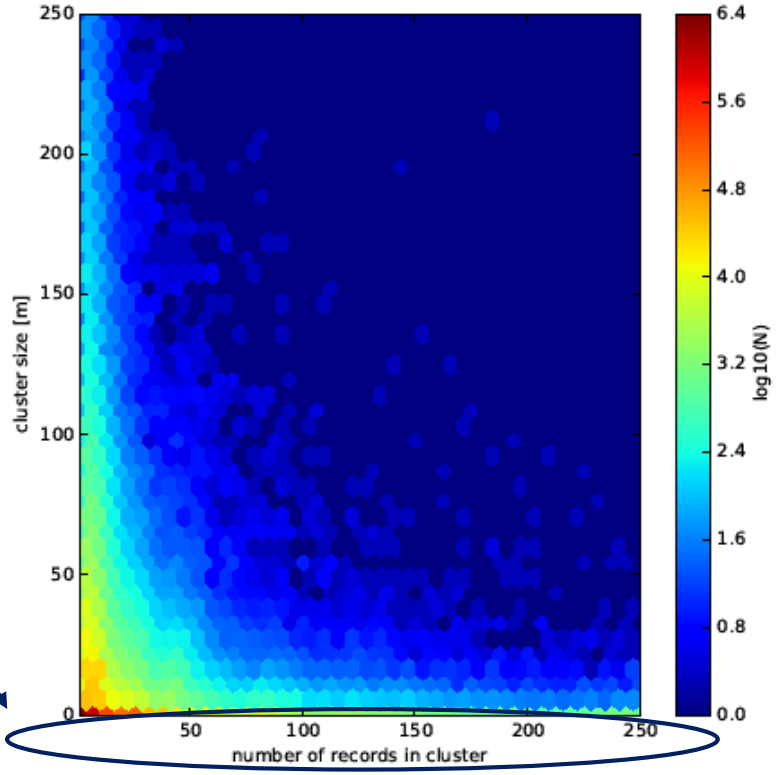
- **Final clusters** are obtained imposing a transitivity relation on pair-wise clusters



Detailed Results

Cluster-size vs. # risks

- Vast majority of clusters have effective size of 0 m
  - i.e. all records within these clusters have exactly the same lat&long coordinates.



- 0.6% of clusters contain more than 200 risks
  - examples:
    - 1 cluster contains 692 risks with address “5000 E VENICE AVE” and coordinates (lat = 27.1, long = -82.3)
    - various large clusters contain risks with address “X” and matching other attributes (admin1, admin2, admin3, place, zip, lat, long)

Thank you!

©2016 Swiss Re. All rights reserved. You are not permitted to create any modifications or derivative works of this presentation or to use it for commercial or other public purposes without the prior written permission of Swiss Re.

The information and opinions contained in the presentation are provided as at the date of the presentation and are subject to change without notice. Although the information used was taken from reliable sources, Swiss Re does not accept any responsibility for the accuracy or comprehensiveness of the details given. All liability for the accuracy and completeness thereof or for any damage or loss resulting from the use of the information contained in this presentation is expressly excluded. Under no circumstances shall Swiss Re or its Group companies be liable for any financial or consequential loss relating to this presentation.