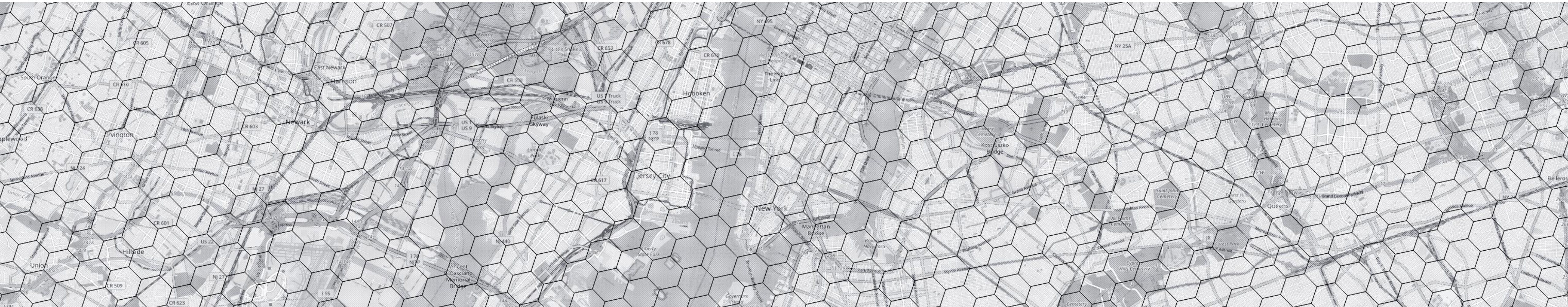
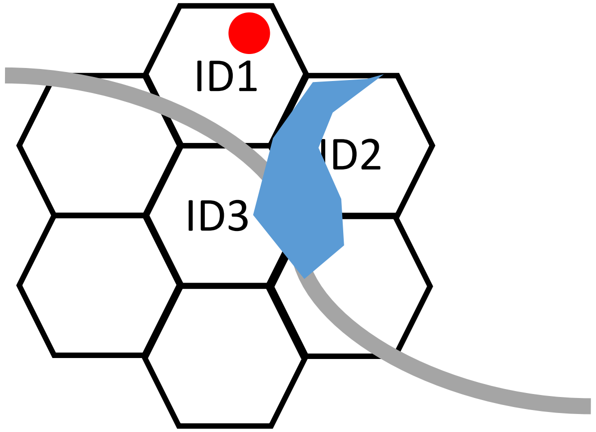
A Planet Scale Spatial-Temporal Knowledge Graph Based On OpenStreetMap And H3 Grid

Martin Böckling, Heiko Paulheim & Sarah Detzler

ESWC 2024, 26.05.2024



Background



Grid Cell ID	hasRiver	hasCampfire	Wildfire
ID1	True	True	False
ID2	False	False	True
ID3	True	False	True

- Use Grid cells as single elements within datasets
- No inclusion of surrounding elements
- Limited data base

UNIVERSITY OF MANNHEIM

State of the art

Spatio-(Temporal) Knowledge Graphs

WorldKG

GeoKG

LauNuts

CrowdGeoKG

OSM2RDF

GEKG

CrispKG

Linked GeoData

AugGKG

KnowWhereGraph

UNIVERSITY OF MANNHEIM

State of the art

Spatio-(Temporal) Knowledge Graphs - WorldKG

WorldKG constructs a Spatial Knowledge Graph using OpenStreetMap transposing the features from OpenStreetMap

- WorldKG contains the data for the complete planet for one time snapshot
- Contains 828 million triples and 113 million entities
- Each geometry associated to a OpenStreetMap feature is translated into a Point WKT format

Dsouza, A., Tempelmeier, N., Yu, R., Gottschalk, S., & Demidova, E. (2021). Worldkg: A world-scale geographic knowledge graph. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 4475-4484).

State of the art
Spatio-(Temporal) Knowledge Graphs - KnowWhereGraph

KnowWhereGraph contains a Spatio-Temporal Knowledge Graph modeling multiple datasets into the Knowledge Graph

- KnowWhereGraph models thematic datasets (Wildfires, Earthquakes, …) and Place Centric datasets (S2 Grid, ZIP codes, …) over time span
- Contains 13+ billion triples
- Models place centric datasets and thematic datasets based on spatial predicates

Janowicz, K., et al. (2022). Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. *AI Magazine*, *43*(1), 30-39.

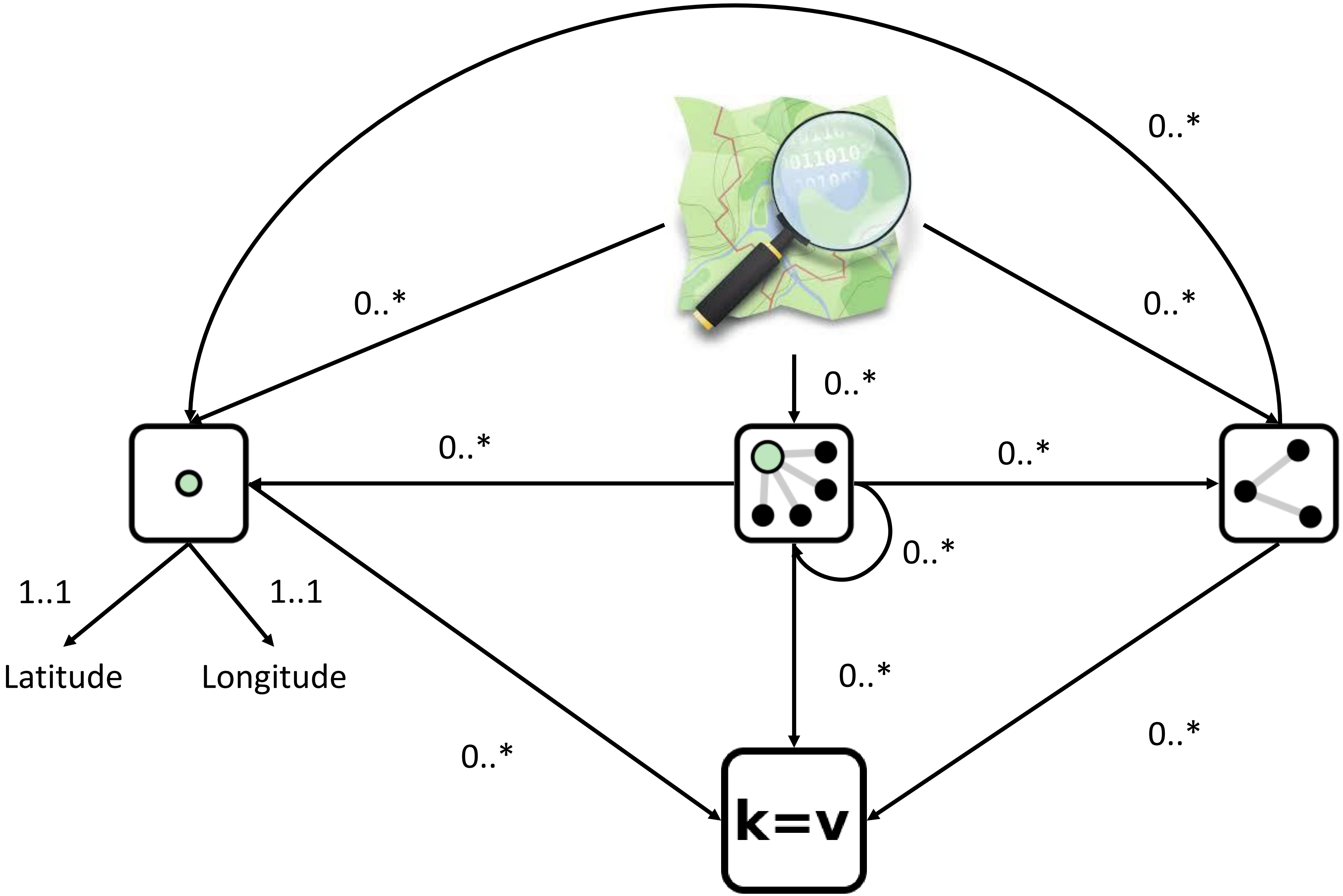
Motivation & Research Goals

Cover spatial data over global extent

Provide temporal dimension to KG

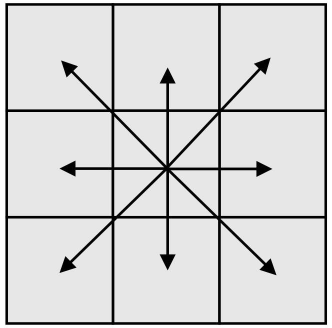
Provide extensibility to KG

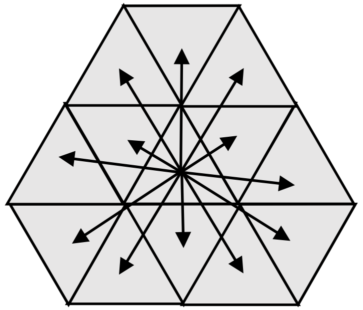
OpenStreetMap data structure

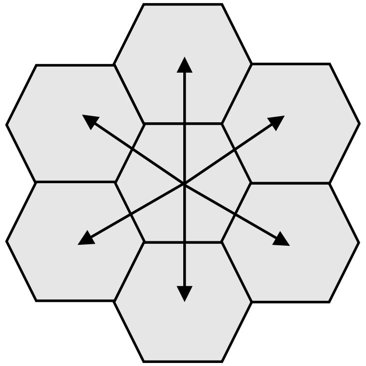


UNIVERSITY
OF MANNHEIM

Extensibility through spatial grid

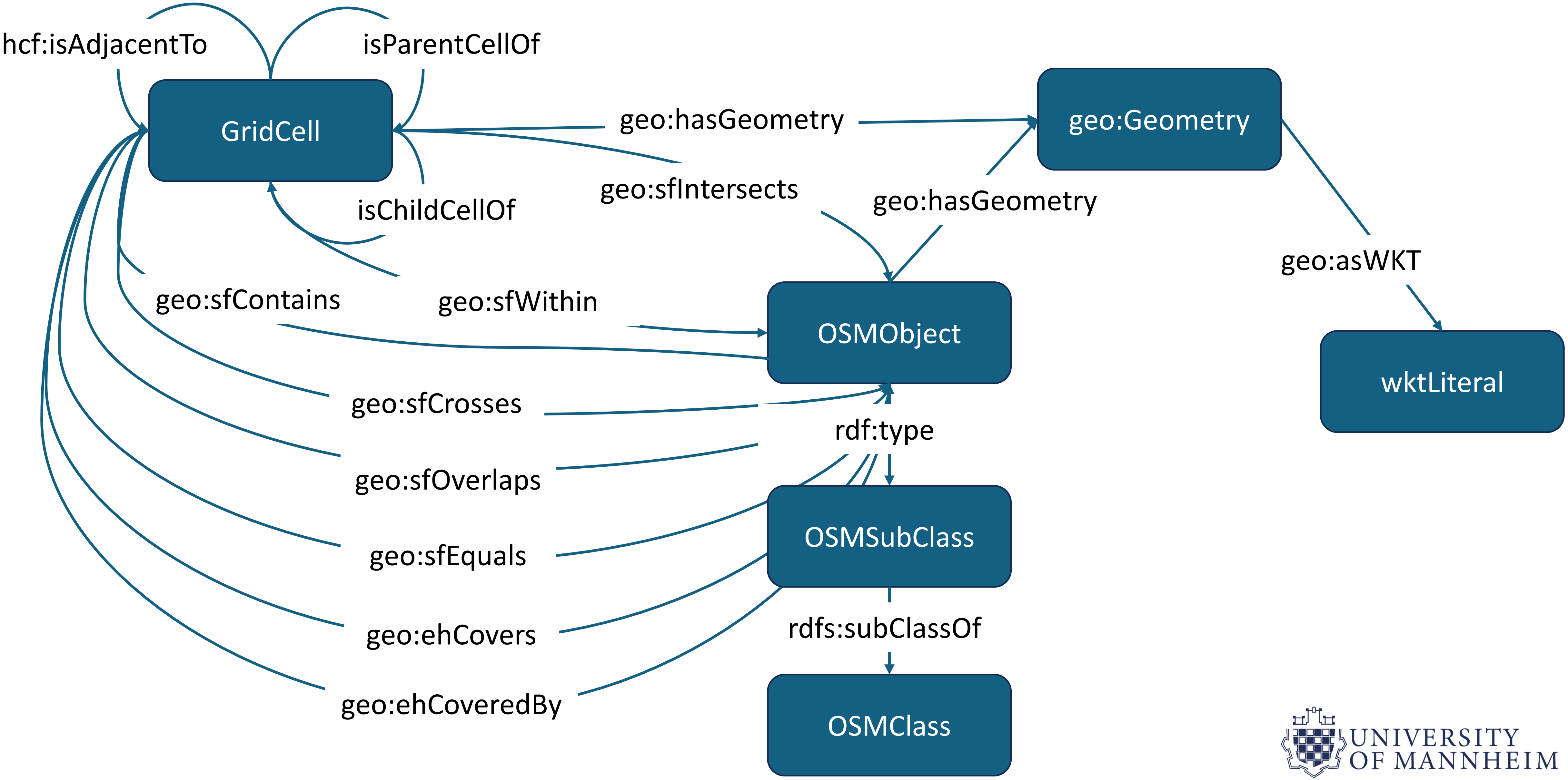






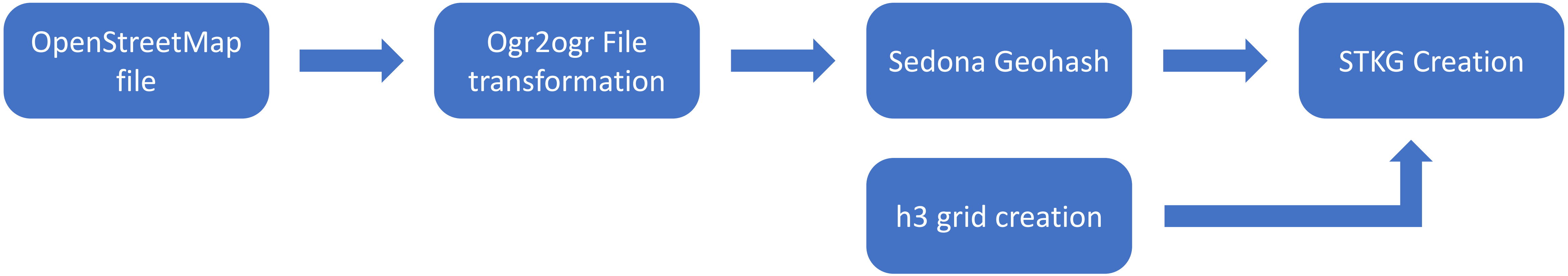
- Selection for regular grids between triangular, square and hexagonal grids
- Usage of hierarchical spatial grid to make sure global representation on different resolutions
- Selection to use hexagonal grid system
    - More accurate representation during tesselation of space
    - Modeling of neighbor information has same distance to all neighboring cells
- Selection to use h3 grid system for a hierarchical discrecte global grid
- Grid aggregation possibility based on population density

Knowledge Graph ontology



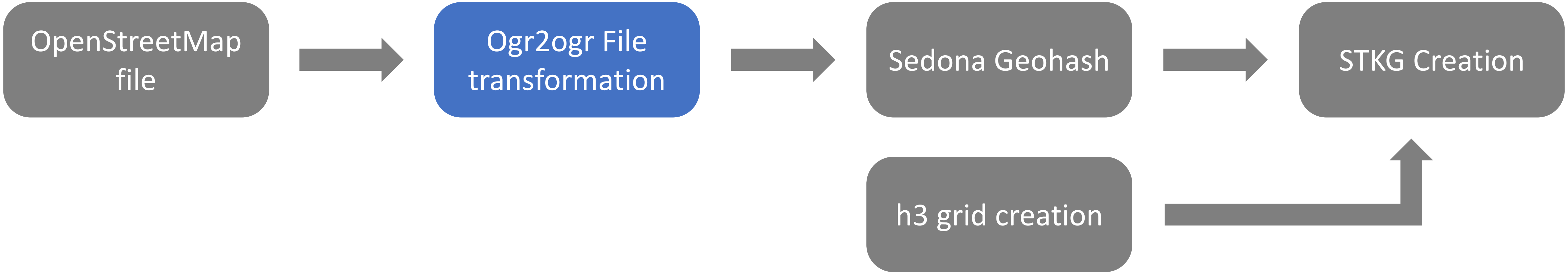
UNIVERSITY OF MANNHEIM

Spatio-Temporal Knowledge Graph creation



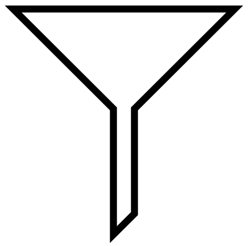
UNIVERSITY OF MANNHEIM

Spatio-Temporal Knowledge Graph creation

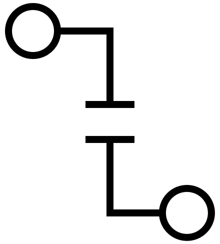


OpenStreetMap to GeoParquet

Transform the pbf file into a GeoParquet file to support processing of Apache Sedona for Spatial Temporal Knowledge Graph creation



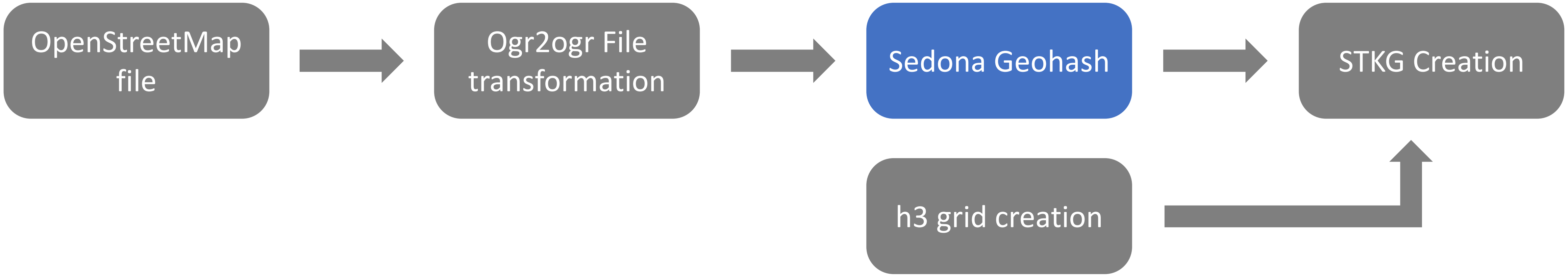
Filtering of unnecessary OpenStreetMap tags



Transform invalid geometries into valid geometries

Skipping of features in case transformation failed

Spatio-Temporal Knowledge Graph creation

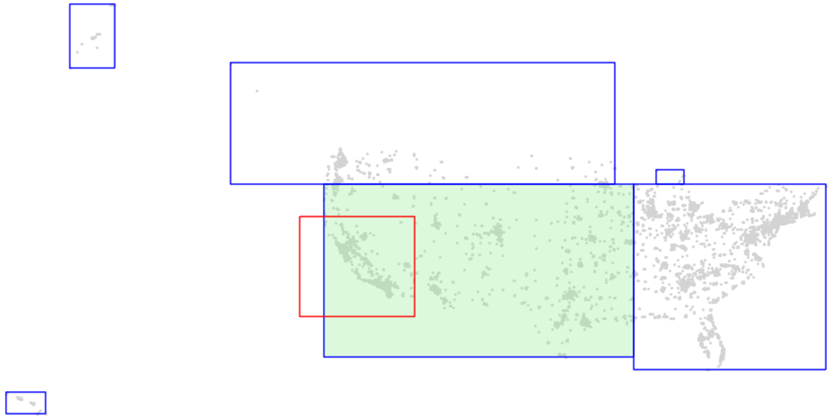


UNIVERSITY OF MANNHEIM

Geohash writer with Apache Sedona

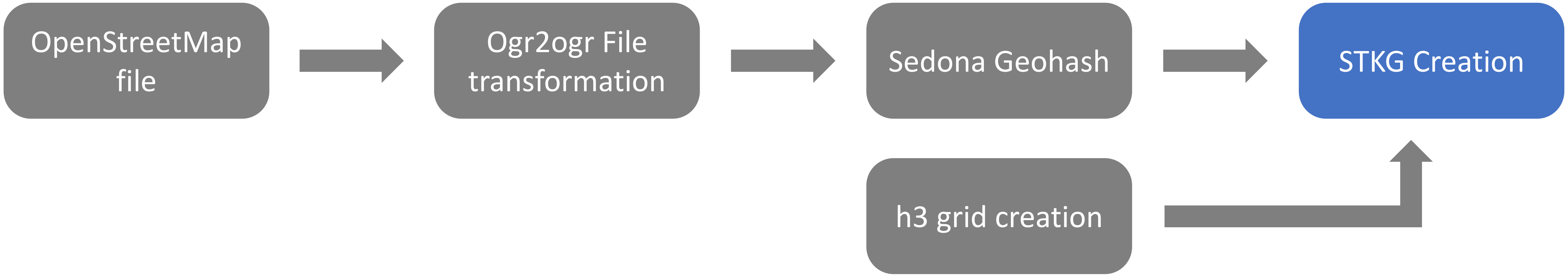
Possibility to push queries down to GeoParquet based on bbox property in GeoParquet metadata

Increased query performance with sorted geohashes



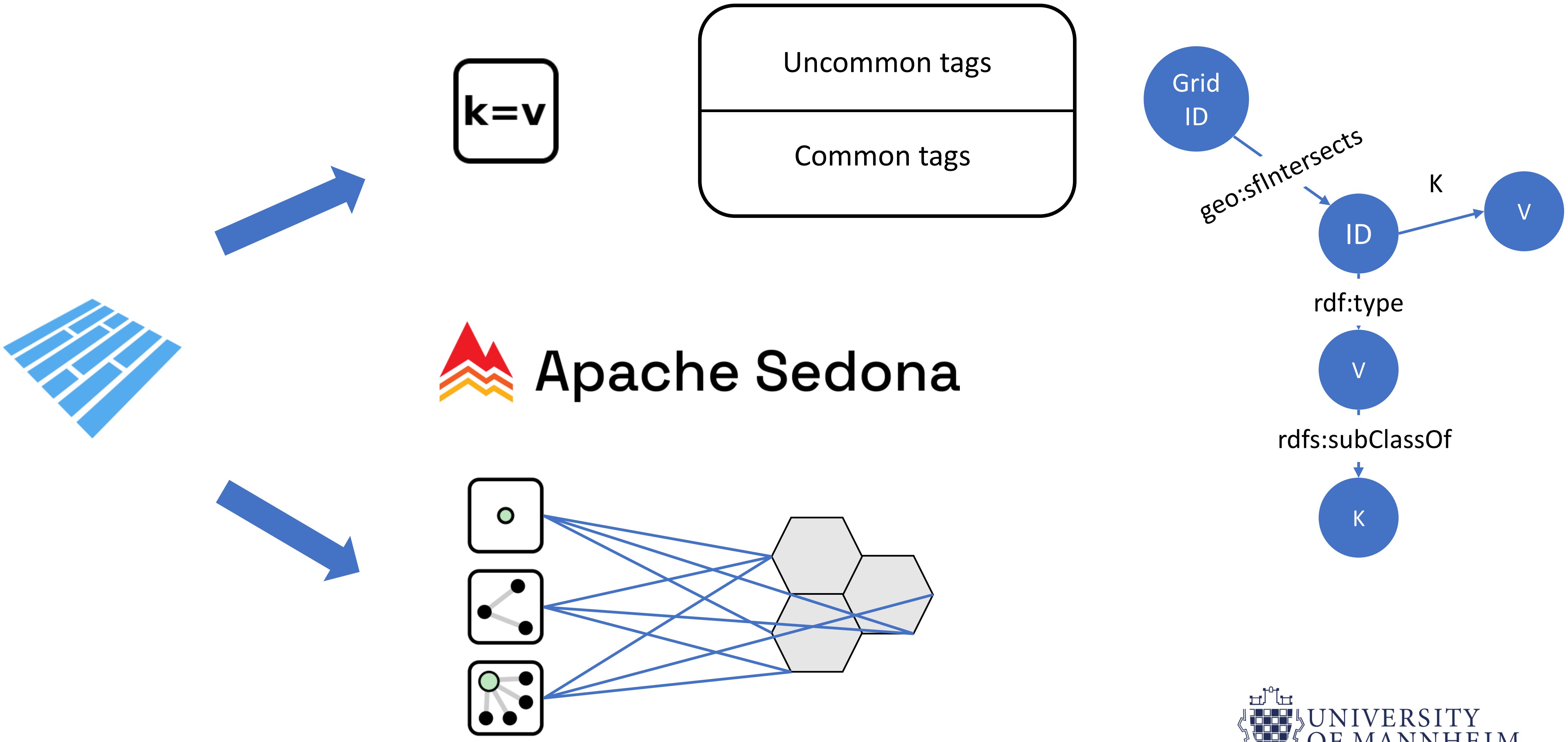
Query window constructed based on join predicate and geometry location

Spatio-Temporal Knowledge Graph creation



UNIVERSITY OF MANNHEIM

Spatio-Temporal Knowledge Graph creation



UNIVERSITY OF MANNHEIM

Instance of Spatio Temporal Knowledge Graph

Way: Universität Mannheim B6 (240974013)

Version #12

addr:street -> place (bzw. entfernt)

Edited 11 months ago by Ropino
Changeset #138367415

Tags

addr:city	Mannheim
addr:place	B6
addr:postcode	68159
amenity	university
name	Universität Mannheim B6
operator	Universität Mannheim
operator:type	university

Nodes

▸ 14 nodes

Download XML

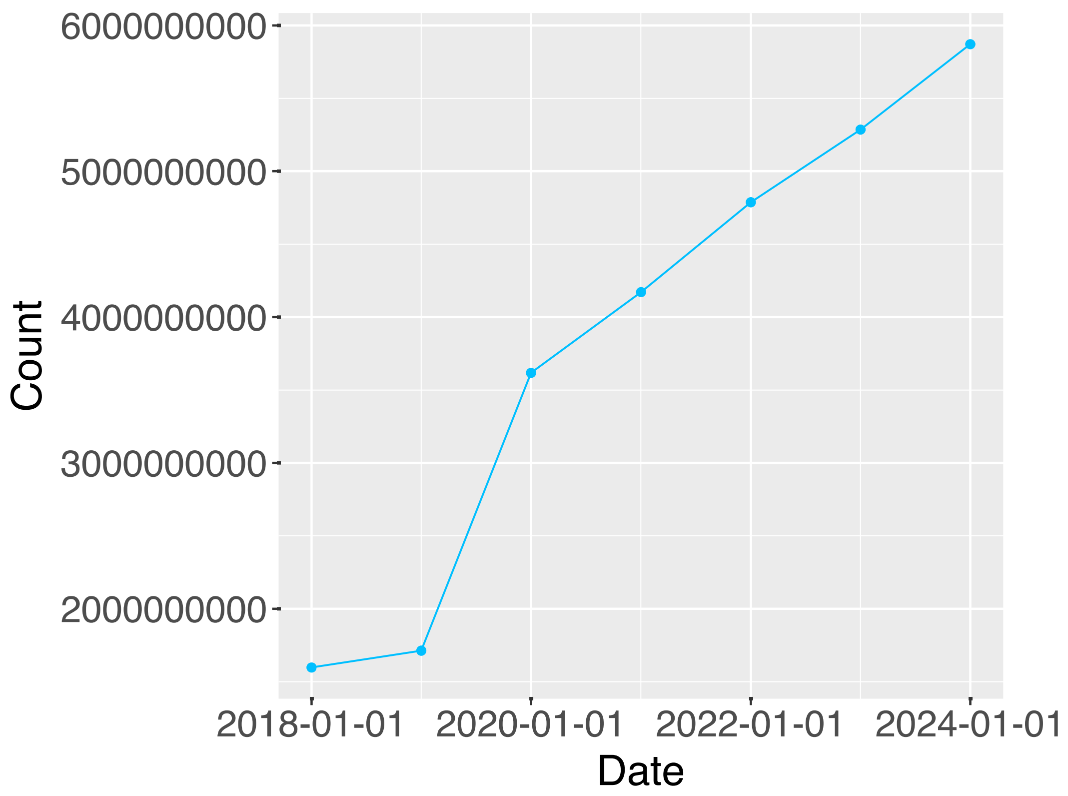
« Version #1 · View History · Version #12 »

Instance of Spatio Temporal Knowledge Graph

Subject	Predicate	Object	Date
240974013	rdf:type	university	2024-01-01
university	rdfs:subClassOf	amenity	2024-01-01
240974013	addr:city	Mannheim	2024-01-01
240974013	addr:postcode	68159	2024-01-01
240974013	name	Universität Mannheim B6	2024-01-01
240974013	operator	Universität Mannheim	2024-01-01
240974013	operator:type	university	2024-01-01
240974013	geo:hasGeometry	geo240974013	2024-01-01
geo240974013	geo:asWKT	POLYGON(…)	2024-01-01
881fae61b9fffff	geo:sfContains	240974013	2024-01-01
881fae61b9fffff	geo:ehCovers	240974013	2024-01-01
881fae61b9fffff	geo:sfIntersects	240974013	2024-01-01
881fae61b9fffff	hcf:isAdjacentTo	881fae6183fffff	2024-01-01

Spatio Temporal Knowledge Graph characteristics

Characteristic	Count
Triple	27,042,753,856
Distinct entities	1,841,912,579
Distinct predicates	98,955



UNIVERSITY OF MANNHEIM

Conclusion and outlook

In total our Knowledge Graph provides over 27 billion entities over a temporal period of 2018 to 2024 (yearly snapshots).

Usage of scalable framework allows large scale processing

No direct possibility to use SPARQL on Knowledge Graph

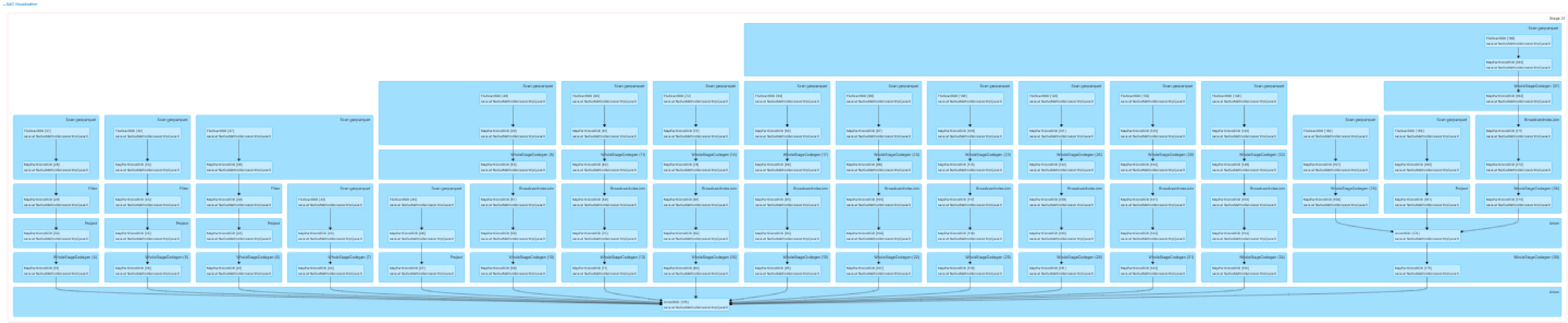
Extension to model OSM geometry relation between each other

Thank you

UNIVERSITY
OF MANNHEIM

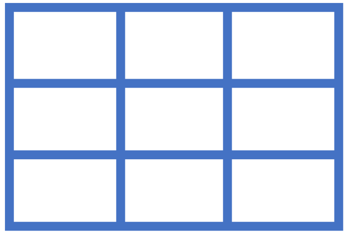
Appendix

Spark DAG

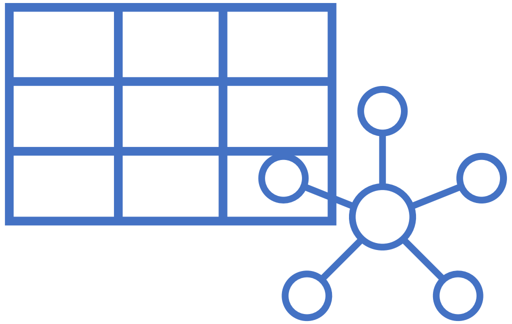


Evaluation Plan

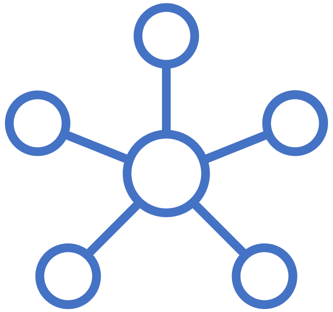
Different spatial datasets are compared by dividing the dataset into three different dataset constellations and performing machine learning tasks



Tabular Cases



Hybrid Cases



Graph Cases

Research questions are evaluated based on ablation studies in the different dataset constellations

UNIVERSITY OF MANNHEIM

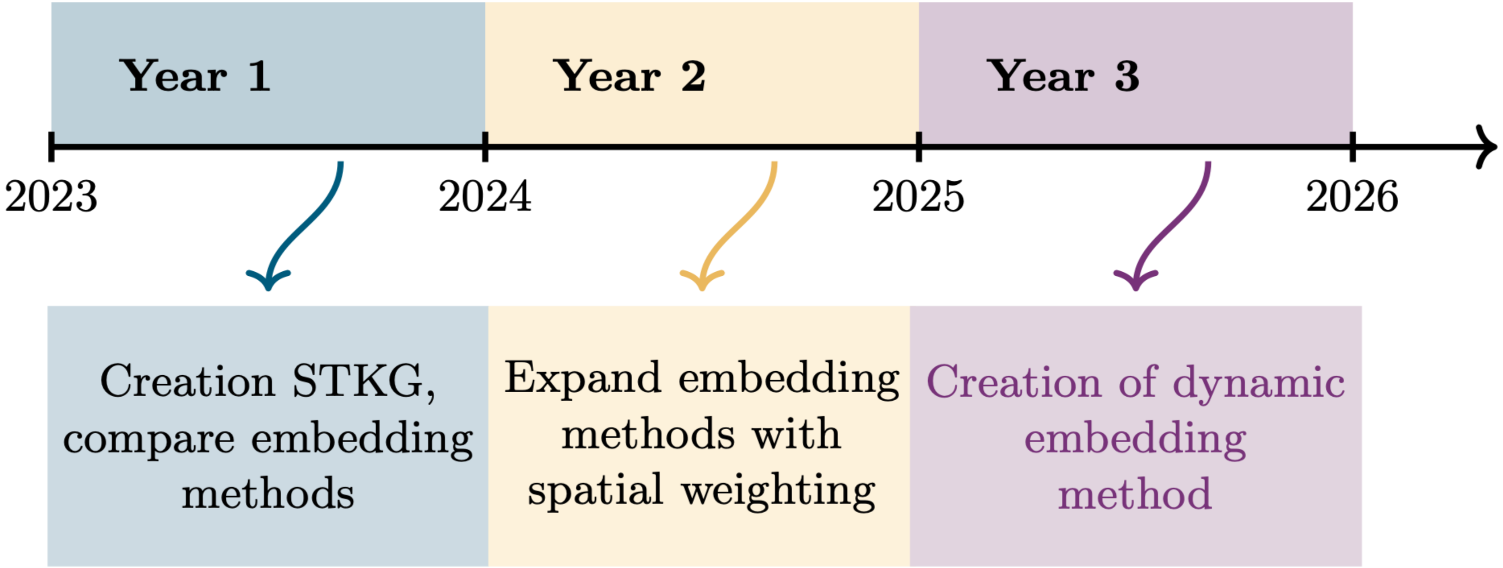
Preliminary results

Dataset containing wildfires over time span of 2010-2021 in California containing weather, landscape and modeled on hexagonal grid cells

Dataset	F1	AUC
Tabular Case	0.3478	0.6816
Hybrid Case	0.3803	0.8748
Network Case	0.0107	0.5341

Based on results derived from Böckling, M., Paulheim, H., & Detzler, S. (2023). Wildfire Prediction Using Spatio-Temporal Knowledge Graphs

Research Methodology and Approach



Data Preparation - Weather

- Pointwise measurements of weather variables
- Need to interpolate data over created spatial grid
- Used interpolation technique:

  Kriging:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i * Z(s_i)$$

Weight λ_i is determined by a semivariogram

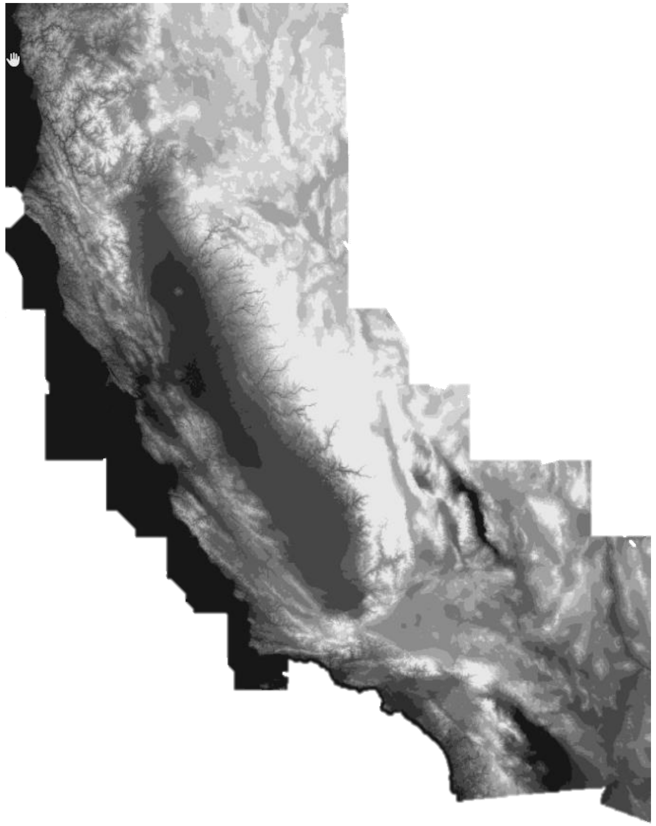
Semivariogram determines spatial autocorrelation and fits function to data

- Each constructed grid cell has now interpolated values for weather variables

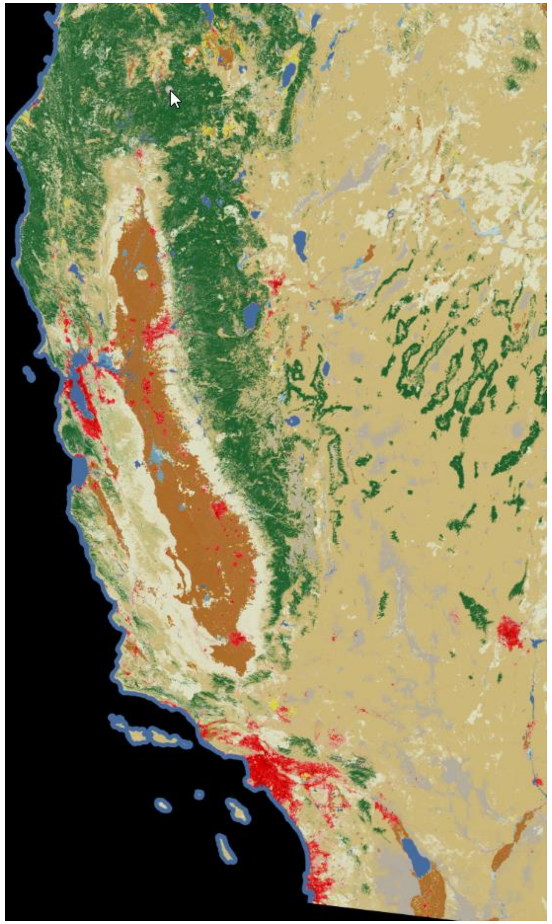
- λ_i: Weight at i
- $Z(s_i)$: Value at point s_i
- $\hat{Z}(s_0)$: Prediction at point s_0

Data Preparation – Elevation & Landcover data

Elevation



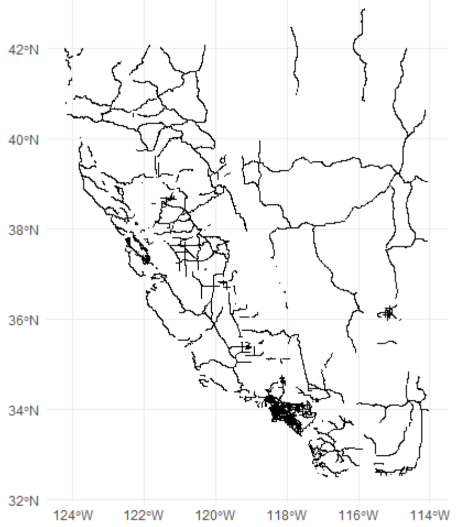
Land cover

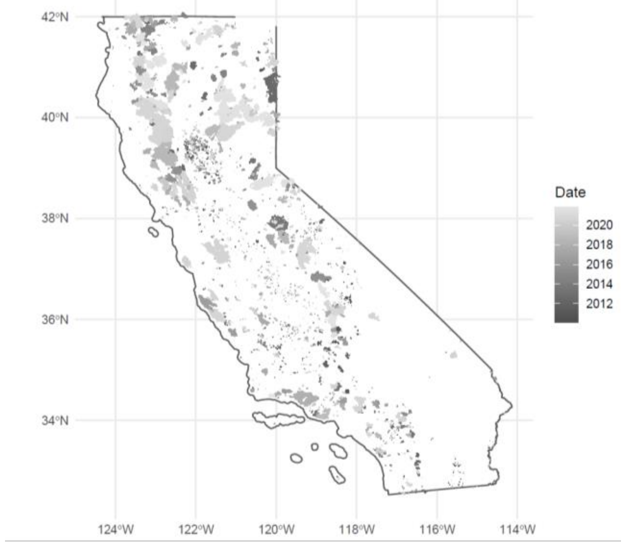


Data Preparation – Elevation & Landcover data

- Both datasets fine granular
  - Elevation 60m*60m tiles
  - Landcover 90m*90m tiles
- Elevation numeric dataset
- Landcover categorical dataset
- Elevation dataset gets aggregated with weighted mean to single grid cell
- Landcover dataset gets aggregated with weighted majority vote to single grid cell

Data Preparation – Openstreetmap & Wildfire data





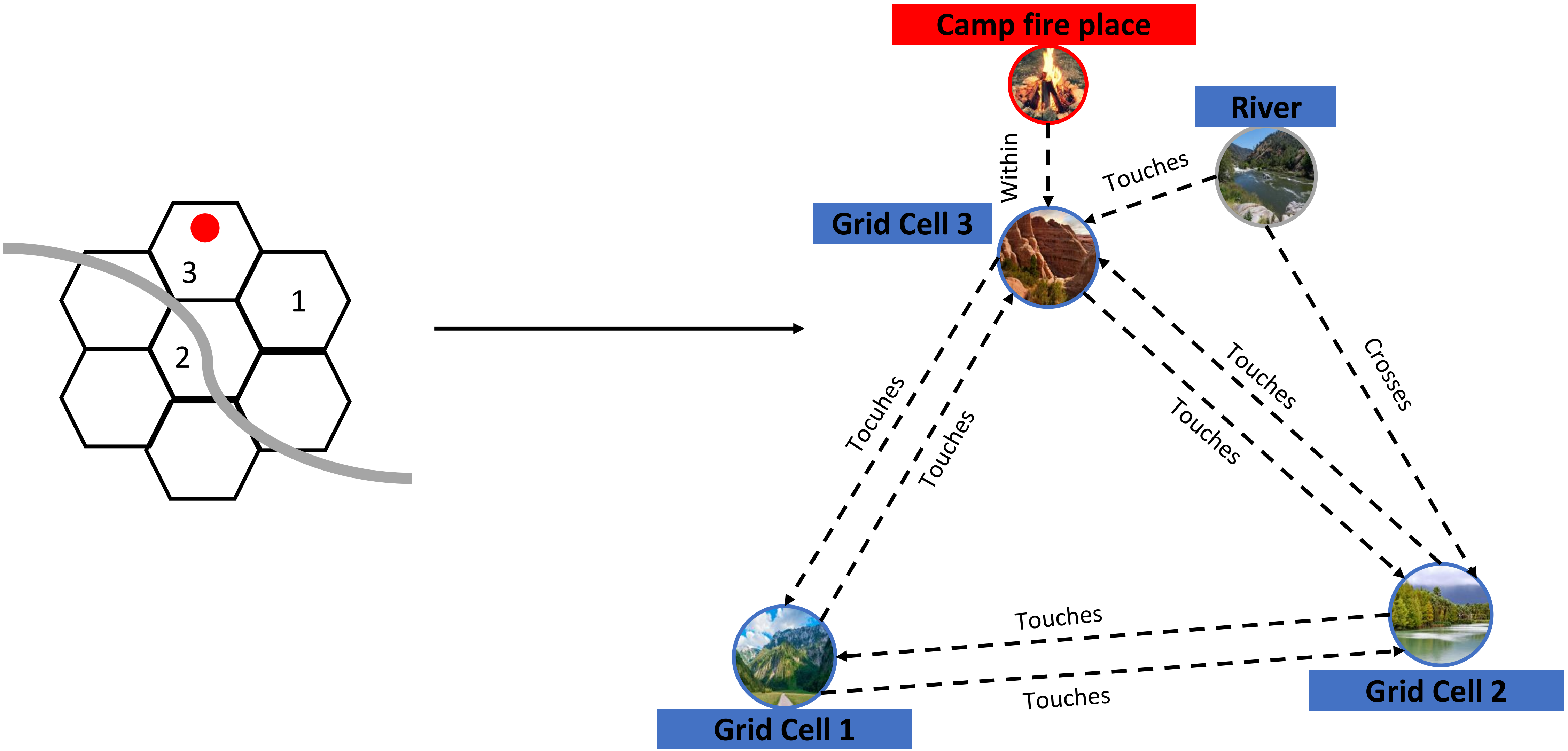
Openstreetmap

- Extract columns related to potential wildfires
- Extract necessary geometry types
- Join Openstreetmap to Grid Cell based on overlap

Wildfire

- Transform year and days to date
- Join wildfire to Grid Cell based on relation overlap

Transform data to spatial knowledge graph



Build up spatial knowledge graph – DE-9IM

- DE-9IM is topological model to build relationships between geometric objects
- DE-9IM model is based upon a 3x3 intersection matrix
  - Measures the dimension dim between two geometries a and b
  - Differentiates between Interior I, Boundary B and Exterior E aspects of geometries

$$\mathrm{DE9IM}(a,b) = \begin{bmatrix} \dim(I(a) \cap I(b)) & \dim(I(a) \cap B(b)) & \dim(I(a) \cap E(b)) \\ \dim(B(a) \cap I(b)) & \dim(B(a) \cap B(b)) & \dim(B(a) \cap E(b)) \\ \dim(E(a) \cap I(b)) & \dim(E(a) \cap B(b)) & \dim(E(a) \cap E(b)) \end{bmatrix}$$

- Value range for dim: {∅ (empty set), 0 (point), 1(lines), 2 (areas)}
- For Knowledge graph creation all geometry objects are related with DE-9IM to created spatial grid
- The pattern of result can be transformed into spatial predicates like the following:
  - Overlap, Touches, Within