OpenBridgeGraph: Integrating Open Government Data for Bridge Management

Jia-Rui Lin[a,b]

[a]Department of Civil Engineering, Tsinghua University, China
[b]Tsinghua University-Glodon Joint Research Center for Building Information Modeling, Tsinghua University, China
E-mail: lin611@tsinghua.edu.cn

Abstract –

Due to limited funds, road authorities around the world are facing challenges related to bridge management and the escalating maintenance requirements of large infrastructure assets. Nowadays, many government organizations have published a variety of data to enable transparency, foster applications, and to satisfy legal obligations. Open governments data like bridge data, weather data would help to better assess the condition of bridges for maintenance purpose and allocation of funds. However, these data sets are fragmented in different systems or formats, and their value in bridge management are not fully explored. This paper proposes a graph-based bridge information modeling framework to integrate open government data for bridge management. The framework represents bridge inventory data as a labeled property graph model and extends the model with weather data. Implementation of the framework employs python scripts for data processing, and neo4j database for data management. The framework is demonstrated using data from national bridge inventory (NBI) and national oceanic and atmosphere administration (NOAA). The results show that the proposed framework can potentially facilitate the integration and retrieval of public government data, and effectively support and provide services to bridge management. Scripts and used data are also shared on GitHub to foster future explorations.

Keywords –

Bridge management; Open government data; Bridge information modeling; Graph database; Data retrieval; Knowledge.

1 Introduction

As essential infrastructures for transportation, bridges are widely seen in both cities and rural areas. Since most of the bridges have been built for decades, the deterioration of structural assets and more specifically, deficiencies related to ageing bridges have become a common problem throughout the world[1]. Given that theirs conditions are getting worse year by year, there is a huge demand for resources and funds to maintain bridges every year. For instance, UK has more 160,000 bridges and it will cost about £180 m to maintain or repair bridges in England only according to previous research[2].

Due to limited funds, it is hard to fulfill the escalating maintenance requirements, and the road authorities have to make decisions wisely to avoid potential structural failures. Thus, structural condition assessment and rating are always utilized to choose bridges that with the worst conditions[1, 3]. This calls for a national bridge database[2] as well as bridge management systems, to estimate costs of bridge maintenance accurately. Thus, National Bridge Inventory is established as a unified database to analyze bridges and judge their conditions, for safety and management purposes[2] in the United States since 1968. And similar databases are later created in other countries.

Usually, typical bridge maintenance scenarios like risk evaluation or condition rating require not only structural characteristics, inspections but also other factors such as environmental parameters[4]. However, environmental parameters such as temperature, precipitation are not included in the national bridge databases like NBI. And environmental data are still missing in most of the bridge management systems[2] such as Pontis[5] in the United States, DANBRO in Denmark. Lack of environmental data will lead to inaccurate assessment of bridge conditions and impact the decision-making process.

Recently, many government organizations have published a variety of data to enable transparency, foster applications. For example, the public sector information directive in Europe, the open data initiative in 2009 in the United States, are proposed and open government data portals such data.gov.uk, data.gov, and data.gov.sg are provided for citizens and stakeholders[6]. The Federal

Highway Administration (FHWA) of the U.S. Department of Transportation has opened the NBI database for public access[7]. While there are also quite a few datasets related to environmental data available from national oceanic and atmosphere administration (NOAA)[8]. Integration of all these datasets will bring new opportunities for bridge management and maintenance.

However, to the best knowledge of the author, few attentions have been paid to integration of open government data for bridge maintenance. Therefore, this research explores how to integrate open government data and how to use them for bridge management purpose. First of all, research methodology and framework are proposed in section 2. Then, a labeled property graph model is introduced to model bridge inventory data and environmental data in section 3, procedures and scripts for integration of open government data to create the graph model are also explained in this section. While section 4 demonstrates how to use the established graph model in various data retrieval scenarios related to bridge management. Finally, benefits of the proposed method and potential future works are concluded and discussed in section 5.

2 Framework

As mentioned above, various government data have been opened for public access, and more open data will be accessible in the future. However, these data are usually represented in different formats. To integrate heterogeneous data from different sources, graph based data model are preferred than relational data model according previous investigations[9].

Thus, this research utilized a similar approach based on labeled property graph. As shown in Figure 1, the proposed framework consists of five steps, namely, data collection, graph modeling, graph creation, model extension and application, which are explained in detail as follows.

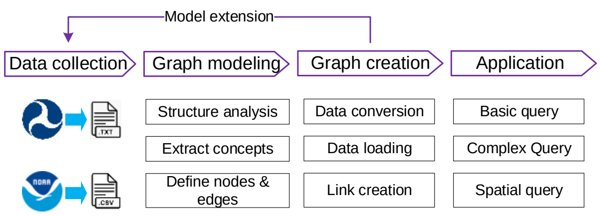

Figure 1. Workflow of the proposed framework

1. **Data collection:** first of all, bridge information from opened bridge databases like the NBI database[7] are downloaded as text files similar to csv format, and environmental data is also obtained from the website such as the NOAA[8] as csv files.
2. **Graph modeling:** considering that the data types, property names and values are encoded and represented with different rules, for instance, NBI database is recorded and encoded following this guide[10], structure of the collected data are analyzed. Then, key concepts as well as their properties are extracted by analyzing corresponding guides on data formatting. At last, key concepts are modeled as nodes of a graph, with their names or classes attached as labels, and relationships between concepts are represented as edges.
3. **Graph creation:** to create the defined graph model in the previous step, a data conversion process is firstly implemented based on python scripts. Then, all data are loaded into a graph database called neo4j, and links between datasets from different sources with neo4j cypher scripts if needed.
4. **Model extension**: since more open data could be accessed, they could be integrated in the graph model through model extension. Generally, engineers or developers could follow the same way as the above-mentioned 3 steps, and the same concepts used in different models and links between different concepts should be identified to merge different data sources and connect different concepts together.
5. **Application**: finally, the created graph is demonstrated with different data retrieval scenarios. In this research, basic query showing the capacity of graph database, complex query for finding bridges linking two states, and spatial query to get environmental data from nearest weather stations of a bridge are provided.

3 Graph-based Modeling of Bridge and Environmental Information

3.1 Concept Graph Model

Following the above-mentioned framework, a concept graph model in Figure 2 is first established for bridge management purpose.

In the middle of Figure 2, concept Bridge is used to represent basic information of a bridge. Structure type, year built, length, and a few other properties defined in NBI database are included in Bridge. Instead of taken all data related to a bridge into a single row as NBI database did, Route, Feature, Traffic, Navigation, Inspection, SpecialInspection, and Improvement are introduced as new concepts. Route stands for a road a bridge carries, and Features are rivers, creeks, etc., that a bridge intersects. Traffic and Navigation capture properties related to traffic loads (i.e., cars, trucks, buses) and navigation control on waterway. Within these two concepts, average daily traffic, vertical and horizontal clearance of navigation are usually considered properties. Meanwhile, Inspection and SpecialInspection are also introduced to model data related to bridge inspections. Date of inspection, inspection frequency as well as category of special inspections like underwater inspection, fractural inspection is considered in the data model. Finally, improvements made to a bridge are also modeled as Improvement in the proposed model. Cost, date and other properties of improvements are considered.

Moreover, the concepts State, County, and Agency are also introduced to represent state, county a bridge locates in and the agency which the bridge belongs to. According to the NBI database, states are also responsible for the maintenance of bridges, and thus a relationship ResponsibleFor is defined.

To better illustrate the proposed graph model, an exemplary graph of bridge inventory which shows most of the previously mentioned concepts and relationships are provided in Figure 3.

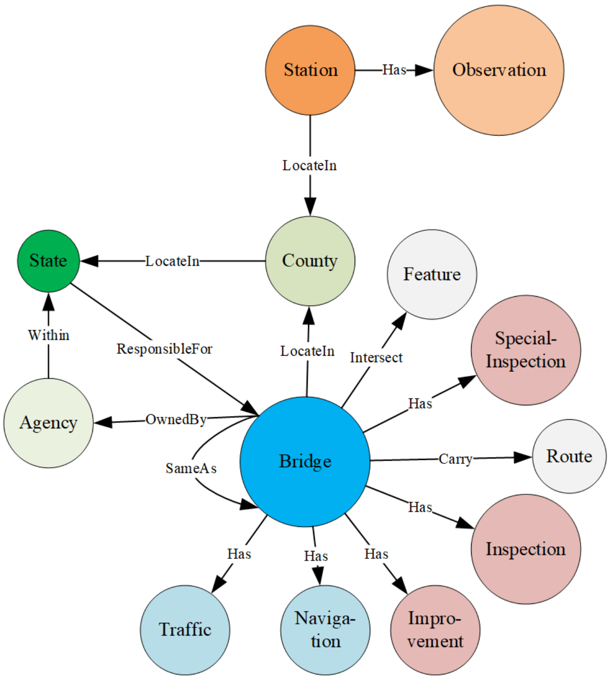

Figure 2. Concept graph model for bridge management

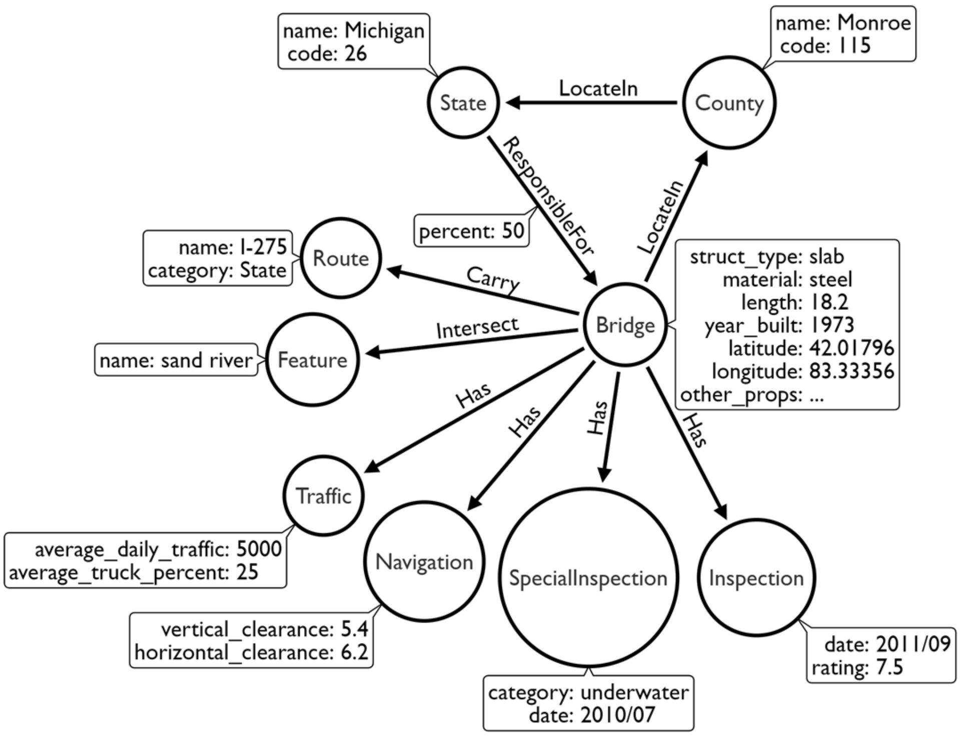

Figure 3. Exemplary Graph of Bridge Inventory

At the top of Figure 2, two concepts, Station, and Observation, are introduced to model environmental from NOAA. Station captures the ID, name, latitude, longitude, and elevation of a weather station, while Observation represents a few environmental features such as temperature, precipitation, wind, observed at a curtained time. And a Has relationship is used to model the relationship between Station and Observation. In this way, an exemplary graph model as shown in Figure 4 could obtained based on raw data from NOAA website.

Note that the Has relationship defined in the graph model is not only used to link bridges and inspections, features, navigations, but also used to connect weather stations and their observations. This is where graph model shows its power. Another benefit of labeled graph model is that relationships can also have properties, as ResponsibleFor relationship does in Figure 3, which is quite straightforward for engineers.

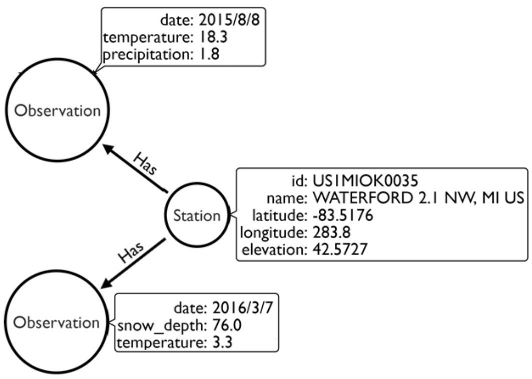

Figure 4. Exemplary Graph of Climate Data

3.2 Data Integration and Graph Creation

To convert multiple source data into the proposed graph model, a data integration and graph creation method is adopted in this research.

Firstly, to transform a data row containing all data related to bridges into different concepts, data mapping rules are defined based on python scripts. Figure 5 shows how to map NBI data tables to Bridge concept and ResponsibleFor relationship of the graph model. According to the upper part of Figure 5, label of the Bridge concept followed by NBI properties used to generated its ID are firstly provided. Then, how to map each column of the NBI data table to properties of the Bridge concept is then defined. Similarly, label and property mapping configurations of relationship ResponsibleFor are defined at the bottom part of Figure 5. Meanwhile, labels as well as NBI properties used to created IDs of source and target concepts are also provided, which links between different concepts can be further created.

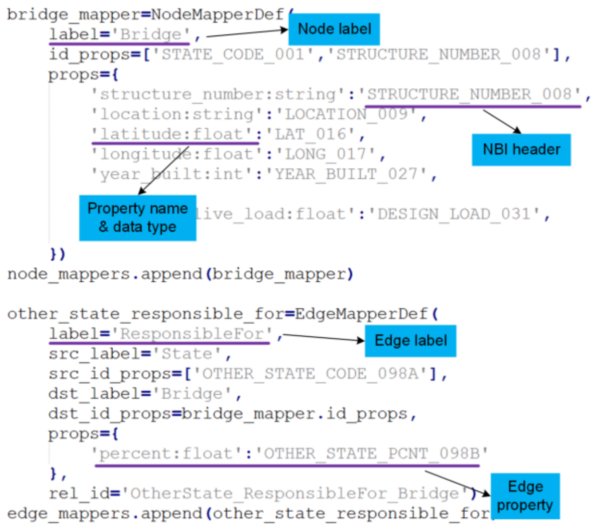

Figure 5. Mapping scripts to create graph model

In addition, since data values of the NBI database are encoded following specific rules, python scripts to decode and convert the data are also needed. For example, latitude and longitude are encoded as a string with 6 digits in the NBI database, and degree, minute, second are represented by the first two digits, the middle two digits, and the last two digits respectively. In this way, python scrips to convert the string to a float number representing the latitude and longitude in radian is provided in Figure 6. Another example shown in Figure 6 is the python script used to convert the year information represented with a two digits string in the NBI database.

```
def get_latlong(raw_val):
    degree=float(raw_val[:-6])
    minute=float(raw_val[-6:-4])
    second=float(raw_val[-4:])/100.0
    return degree+minute/60.0+second/3600.0

def get_date(raw_val):
    month=int(raw_val[:-2])
    year_str=raw_val[-2:]
    if int(year_str[0])<3:
        year=int('20'+year_str)
    else:
        year=int('19'+year_str)
    return datetime.datetime(year,month,1)
```

Figure 6. Scripts to convert NBI properties

Therefore, with defined mapping rules and data conversion scripts, bridge information of the NBI database are converted to different csv files, each of which persists data of a concept or a relationship of the proposed graph model.

Then, the generated csv files could be loaded in neo4j database and the graph is established. As shown in Figure 7, bridge information could be imported with the csv loading script of neo4j, and nodes representing bridges are automatically created.

```
$ LOAD CSV WITH HEADERS FROM
  "file:///Bridge.csv" AS row merge
  (n:Bridge{id:row.id}) on match set
  n+=row on create set n=row;
```

```
$ LOAD CSV WITH HEADERS FROM
  "file:///State_To_Bridge.csv" AS row
  match (s:State{id:row.src_id}),
  (d:Bridge{id:row.dst_id}) merge (s)-
  [r:ResponsibleFor]→(d) on match set
  r+={percent:row.percent} on create
  set r={percent:row.percent};
```

Figure 7. Data loading scripts

Furthermore, once nodes representing bridges and states are created, the ResponsibleFor relationship can also be created by loading corresponding csv files with script provided in Figure 7.

Following the same way, data tables obtained from the NOAA website could also be converted and imported into neo4j database, thereby creating data nodes and relationships related to environmental data.

If connections exist between data nodes generated with different datasets, extra rules could be defined based on neo4j script. For example, if ID of concept ConceptA from a dataset equals a property PropB of concept ConceptB generated by another dataset, then the following script in Figure 8 could be used to create the relationship RelAtoB.

```
$ match (a:ConceptA),(b:conceptB) where
  a.ID=b.PropB create (a)-[:RelAtoB]→(b)
```

Figure 8. Data script to create relationships

4 Demonstration

Following the previously mentioned workflow, this research build a graph model based on data retrieved from the NBI and NOAA website. Specifically, NBI data of Michigan state from 2011 to 2016 and Wisconsin state in 2016 as well as environmental data in Michigan and Ohio state from 2015/1/1 to 2016/12/31 are utilized. In addition, codes and names of all states and counties are also included when creating the graph. At last, there are 190, 525 nodes with 13 labels and 283, 616 relationships with 8 labels created in total. Details of nodes and relationships are listed in Table 1.

According to Table 1, it is concluded that there are 25, 410 bridge inventories carrying 1, 509 routes and intersecting with 6, 955 features in total. While 30, 737 inspections, 1, 506 special inspections as well as 3, 879 improvements were made from 2011 to 2015. In other words, about 1.27 inspections were made for each bridge on average and only 15.26% of them were improved during this period, reflecting that there is a huge demand of funds for bridge maintenance.

Table 1. Number of Nodes and Relationships in Generated Graph Model

Category	Label	Amount
Node	State	52
Node	County	3, 228
Node	Agency	26
Node	Bridge	25, 410
Node	Route	1, 509
Node	Feature	6, 955
Node	Traffic	25, 747
Node	Navigation	15, 952
Node	Improvement	3, 879
Node	Inspection	30, 737
Node	SpecialInspection	1, 506
Node	Station	179
Node	Observation	75, 345
Relationship	LocateIn	28, 628
Relationship	OwnedBy	25, 411
Relationship	Within	26
Relationship	ResponsibleFor	25, 513
Relationship	Carry	25, 413
Relationship	Intersect	25, 438
Relationship	Has	153, 166
Relationship	SameAs	21

To further illustrate how the established graph could be used for various data retrieval scenarios. Due to limited content of this paper, three scenarios are chosen to show basic query, complex query and spatial query capacity of the neo4j graph database.

4.1 Scenario 1: Query Data Related to Bridges

Top of Figure 9 shows a basic query retrieving bridges and all nodes they direct to. In the query, putting noting in the square brackets means matching all relationships, and the arrow implies that only relationships directing from bridges to other nodes are considered. To limit the results within 25, graph shown at the middle of Figure 9 is obtained.

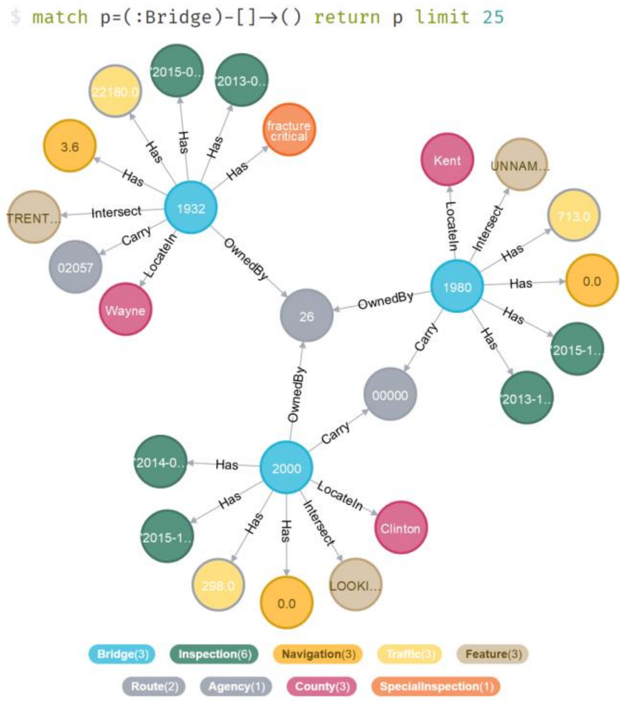

Figure 9. Query bridges and its connected data

It is found that there are 3 bridges built in the year of 1932, 1980, and 2000 respectively, and two of them carry the same route. Meanwhile, a special inspection considering fracture critical details is also conducted when inspecting the bridge built in 1932. In this scenario, it is quite easier to retrieve all information related to a bridge with a simple query. It is also possible to match certain nodes connected with a specific relationship by specifying the relationship's label in the square brackets.

4.2 Scenario 2: Query Bridges Linking Two States

Furthermore, it is also possible to retrieve bridges linking two states. Query shown at the top of Figure 10 generates a small graph at the bottom of Figure 10.

As illustrated in Figure 10, two bridges connecting Michigan state and Wisconsin state are identified with the query. Since the NBI database assigns different unique numbers to bridges located in different states, there are two bridge nodes representing the same bridge in Figure 10. It is also illustrated in Figure 10 that the bridges are maintained together by the two states.

In this way, it is possible to query bridges carrying the same route, intersecting with the same feature, or owned by the same agency.

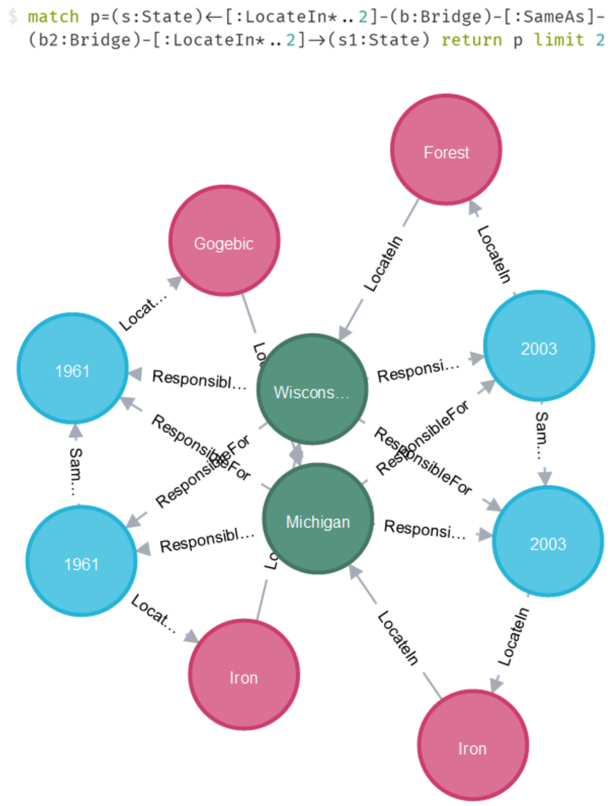

Figure 10. Bridges linking two states

4.3 Scenario 3: Query Nearby Environmental Data of Bridges

At last, given that longitude and latitude of bridges and weather stations are available in the created graph database, it is also possible to retrieve nearby environmental data of a bridge. For example, Figure 11 shows a query retrieving temperature and precipitation observed by nearby weather stations of a bridge for underwater inspection purpose. And part of the results of the query are listed in Table 2

```
$ match p=(b:Bridge)-[:Has]→
(ins:SpecialInspection{category:'underwater inspection'}),
q=(s:Station)-[:Has]→(o:Observation) where ins.date=o.date
and o.precipitation>0 and exists(o.temperature) return
b.structure_number,ins.date,o.temperature,o.precipitation,di
stance(b.loc_point,s.loc_point)/1000 as distance order by
distance limit 5
```

Figure 11. Query nearby environmental data of bridges for special inspection

It is found that underwater inspection of bridge No. 000000000011328 is conducted on 2015/4/1 and the temperature and precipitation observed by the nearest weather station (about 5.92 km away) at the same day are 1.7℃ and 0.3mm respectively. While similar results could be also obtained for the other two bridges.

In this way, researchers, engineers, and road authorities could extract environmental data from weather stations, thereby developing better model for condition evaluation and making wise decisions for bridge maintenance.

Table 2. Nearby environmental data of bridges for special inspection

Bridge No.	Date	Temperature/℃	Precipitation/mm	Distance/km
000000000011328	2015/4/1	1.7	0.3	5.92
B27007400000000	2015/11/1	6.1	3.8	24.05
000000000011884	2015/4/1	-2.2	0.3	27.90

5 Discussion

In this work, a framework integrating multi-source open data for bridge management and maintenance is proposed and demonstrated in a few scenarios.

The framework is validated with data collected from the NBI database and the NOAA website, since they are the public available and could be accesses easily. Despite this, it is possible and easy to adapt the proposed framework for integrating other bridge databases and data sources. For example, bridge database of the UK could be accessed from data.gov.uk, then the same workflow proposed in Figure 1 can be directly used to create a graph model for bridge management and maintenance. Similarly, open weather data could also be integrated with the same way. Moreover, except for bridge data and environmental data, as other open data is published, it is also possible to further extend the established graph model and continuously integrate more data sources.

With the proposed framework and graph model, it is possible for the managers to query related bridge information, environmental data as well as their connections for bridge management purpose. In this way, decision-makers could take open government data as a supplement of their private data to further improve bridge maintenance process. For example, a manager could take a target bridge operated by the agency he/she belongs to, query bridges that have similar states and environmental conditions with the target bridge from the graph database, and by comparing the retrieved bridges with the target bridge, he/she could get insights on how to improve the inspection and maintenance decisions.

However, when conducting this research, it is also found that the quality and richness of open government data still need further improvements. Detailed inspection information, other factors such as geotechnical stability, subsurface conditions are still not available from open data sources. Thus, it is recommended that the governments as well as other data providers should open more data if possible, and an approach that integrating both open data and private data for bridge maintenance are suggested.

6 Conclusions

As the government keeps opening data for public access, there is a trend in integrating multisource open data to foster applications and innovations in different areas. However, though bridges are essential for delivering goods and people around the urban and rural ear and there is a huge demand of funds for bridge maintenance, value of open government data for bridge management has not been explored yet.

This research investigates how to integrate the bridge information and the environmental data opened by the government based on graph modeling for bridge maintenance. A labeled graph model and python scripts for data conversion and integration are proposed. Demonstration in different scenarios show the flexibility and feasibility of the proposed framework. With the proposed method, it is possible to query various data related to bridges and retrieve environmental data nearby bridges, which are valuable for flood risk assessment, vulnerability, damage modeling [5] and decision-making, etc.

In the future, stakeholders are encouraged to open more data, such as data of highway network, traffic data collected by mobiles or sensors, which could be integrated and innovative application in bridge management would emerge quickly. In addition, it is also recommended to integrate both open data and private data as they could complement each other, thus bridge managers could make better decisions for bridge maintenance purpose.

What's more, to foster more explorations and applications, scripts used for data conversion and graph creation as well as a few exemplary graph queries are shared through GitHub:

https://github.com/smartaec/OpenBridgeGraph. Other researchers could access and modify the scripts for their further works.

Acknowledgement

This research is supported by the Tsinghua University Initiative Scientific Research Program (No. 2019Z02UOT), the Natural Science Foundation of China (No. 51908323), and the Beijing Natural Science Foundation (No. 8194067). The author also thanks Prof. Kincho Law (Stanford University) for his valuable comments.

References

[1] Rashidi M., Samali B. and Sharafi P. A new model for bridge management: Part A: condition assessment and priority ranking of bridges. *Australian Journal of Civil Engineering*,14(1):35-45, 2016.

[2] Pregnolato M. Bridge safety is not for granted-a novel approach to bridge management. *Engineering Structures*,196:109193, 2019.

[3] Aktan A., Farhey D., Brown D., Dalal V., Helmicki A., Hunt V. and Shelley S. Condition assessment for bridge management. *Journal of Infrastructure Systems*, 2(3):108-117, 1996.

[4] Wang H., Hsieh S., Lin C. and Wang C. Forensic diagnosis on flood-induced bridge failure. I: Determination of the possible causes of failure. *Journal of Performance of Constructed Facilities*, 28(1):76-84, 2014.

[5] Thompson P., Small E., Johnson M. and Marshall A. The Pontis bridge management system. *Structural Engineering International*, 8(4):303-308, 1998.

[6] Attard J., Orlandi F., Scerri S. and Auer S. A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4):399-418, 2015.

[7] FHWA. National Bridge Inventory. On-line: https://www.fhwa.dot.gov/bridge/nbi.cfm, Accessed: 14/03/2020.

[8] NOAA. Climate Data Online. On-line: https://www.ncdc.noaa.gov/cdo-web/, Accessed: 14/03/2020.

[9] Yoon B., Kim S. and Kim S. Use of graph database for the integration of heterogeneous biological data. *Genomics & informatics*, 15(1):19-27, 2017.

[10] FHWA. *Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges*. Washington, D.C., 1995.