

How a Project to Digitize Legacy Paper Records Can Be Used as the Basis for a Transit System Knowledge Graph

Joseph Busch

Project background

- Situation at a major regional transit authority
 - 3,000 office-based employees spread across more than 10 aging buildings.
 - ✓ Deferred maintenance costs
 - ✓ Safety issues
 - ✓ Fragmented workforce



Goals, challenges, and solution

- Goals

- Consolidate workforce in 3 regional buildings
- Break down silos and encourage collaboration among departments and teams

- Challenges

- New ways of working and space types (hoteling seating, collaboration areas, etc.)
- Limited space for personal materials (paper files, reference materials, etc.)

- Solution

- Change management, engagement, and communications
- Eliminate legacy paper records and reference materials
 - Develop metadata strategy (aka taxonomy), do scanning, and build repository

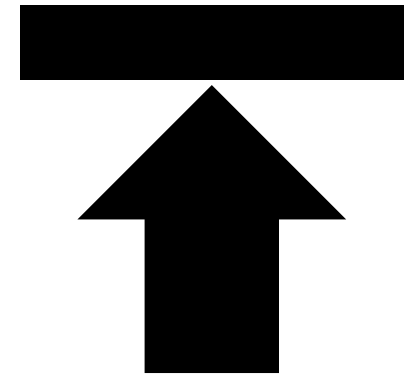
Accidental digital transformation

- The pandemic caused the Authority to make an abrupt shift to remote work from home
- So the focus of this engagement shifted to helping a bureaucracy which had no culture of remote work be able to work remotely

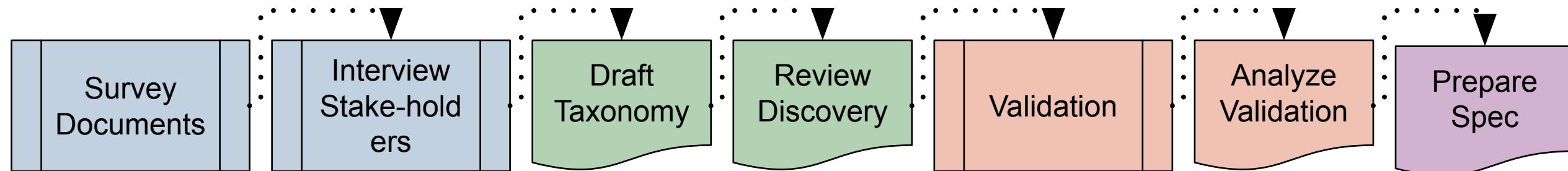


... literally overnight.

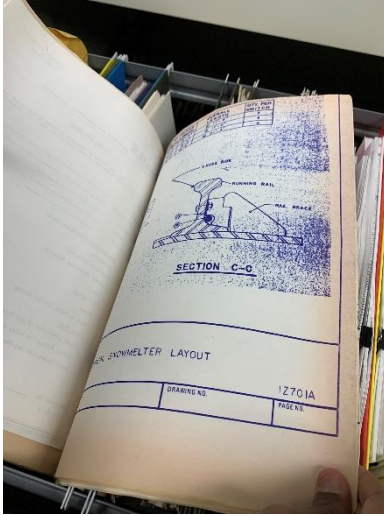
Enterprise taxonomy process in this project



- Bottom up organizational survey
 - Survey each department
 - Interview dept stakeholders
 - Define dept use cases, metadata, and vocabularies (the taxonomy)
 - Validate taxonomy with department
 - Generate specification for department
 - Apply best practices and emerging commonalities to each Department taxonomy sequentially



Document survey



Drawings



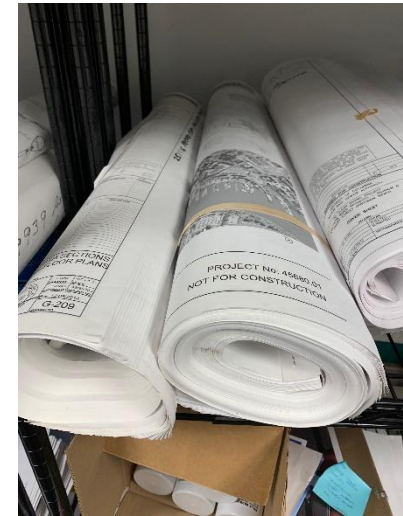
Aperture Cards



Contracts



Personnel Files



Plans



Claims

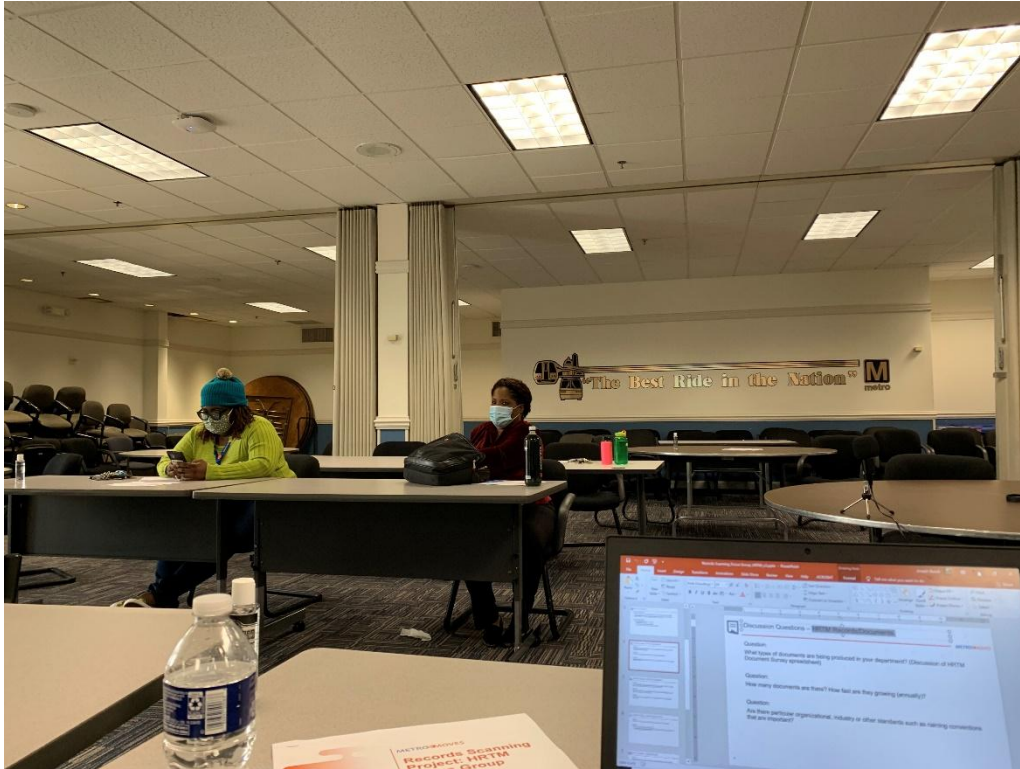


Binders



Oversized Materials

Stakeholder interviews



Draft taxonomy template

	A	B	C	D	E	F	G
1	Metadata Attribute	Attribute Type	Mandatory	Attribute Definition	Populataion Method		
2	Common Document Type						
3	Document Type	Document Type	X	Document types based on Office Record Survey.	Transcribed from box.		
4	Name						
5	Last Name	Name		File owner's last name.	Transcribed from document		
6	First Name	Name		File owner's first name.	Transcribed from document		
7	Title						
8	Title	Title		Document title.	Transcribed from document		
9	Date						
10	Date	Date		Full document date in the format YYYY-MM-DD.	Transcribed from document		
11	Year	Date		Year document covers.	Transcribed from document		
12	Organizational Unit						
13	Org Unit	Organization	X	Abbreviation for organizational unit.	Default to Office Name for all documents.		
14							
15							
16							
17							
18							
19							
20							



Validation

This is a study preview. No data will be saved.

Task 4 of 9

Who are the parties related to Master Agreement 01 (MA01)?

Home

Common Document Types

Master Agreement

← I'd find it here

Digitization specification



Project team



Project demographics

Overall

- 63 Departments participated
- 2 years (Oct 2020-Oct 2022)
- Average of 2.6 Departments per month
- Average of 9.6 attributes per Department
- Document Type and Business Unit were mandatory attributes for every Department

Most common attributes

Attr Type	Attribute Name	%
Date	Date (MM/DD/YYYY)	59%
Name	Last Name	52%
Name	First Name	51%
Identifier	Contract#	25%
Identifier	Employee/Contractor ID	40%
Subject	Title	40%
Date	Year	37%
Location	Line/Station Code	25%
Name	Vendor/Manufacturer	25%
Date	Month/Year	21%

Things, Not Strings

- Introducing the Knowledge Graph: things, not strings

May 16, 2012

Amit Singhal,

“In other words, Google would return what is known as structured information about people, places, events, movies, and other concepts, not just the traditional list of web links containing words matching the search terms.” -- Friedrich Lindenberg

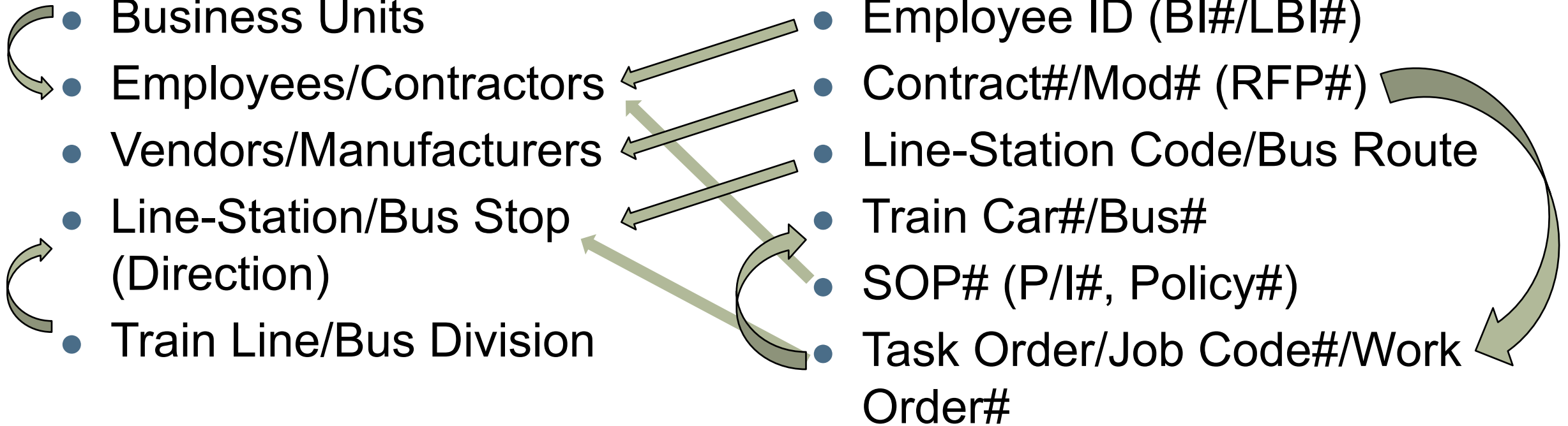
Named entity and identifier relationships

Names

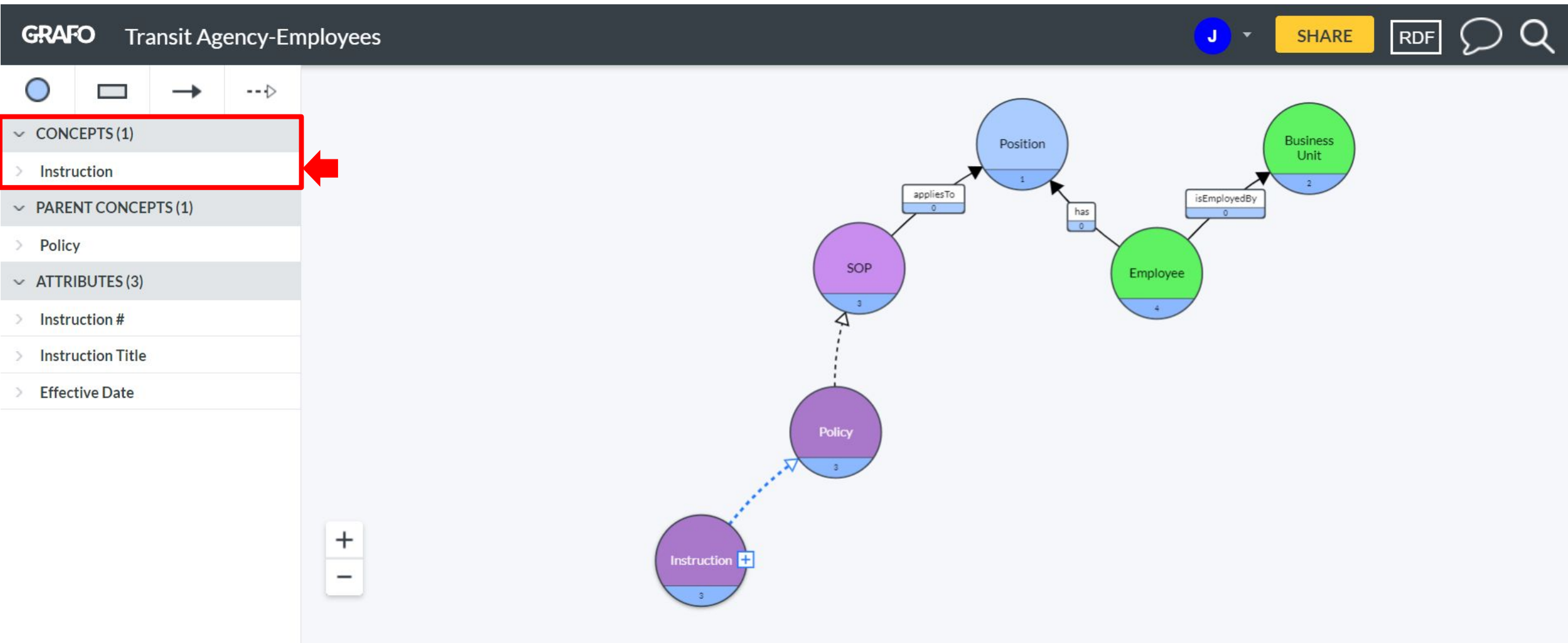
- Business Units
- Employees/Contractors
- Vendors/Manufacturers
- Line-Station/Bus Stop (Direction)
- Train Line/Bus Division

Identifiers

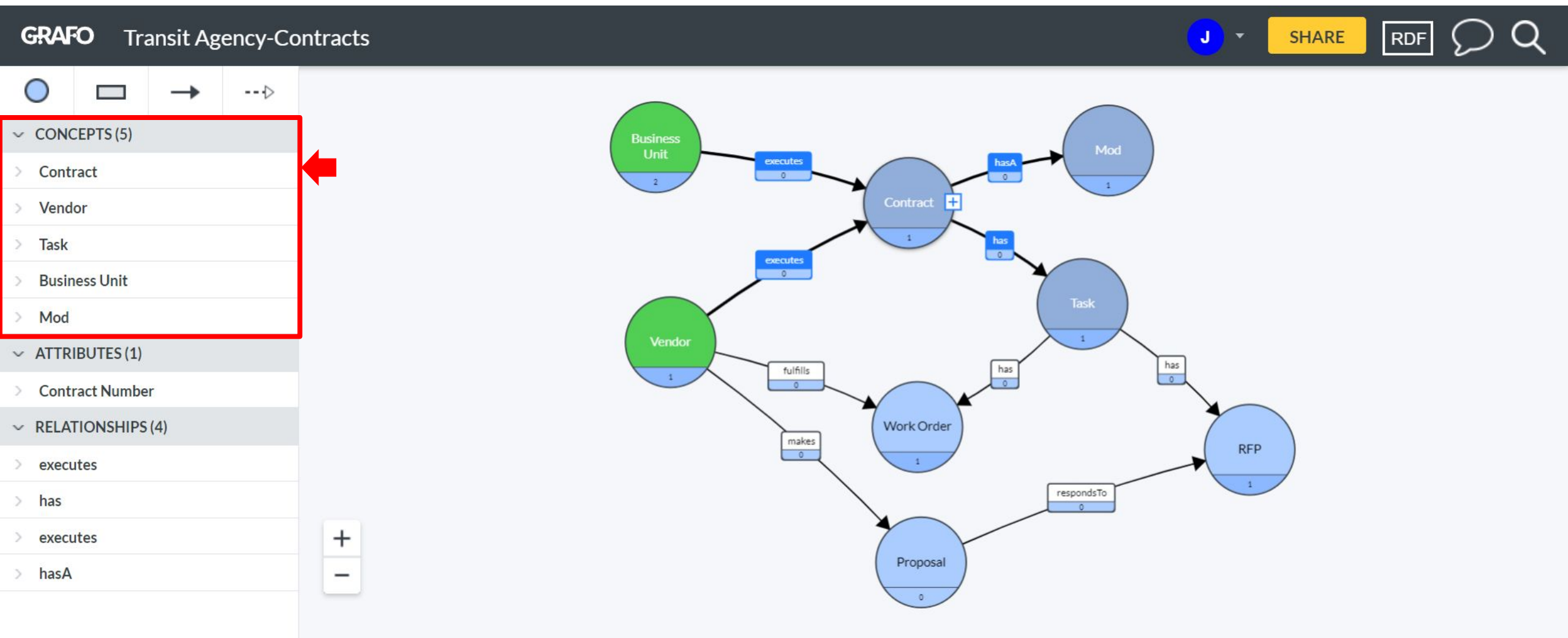
- Employee ID (BI#/LBI#)
- Contract#/Mod# (RFP#)
- Line-Station Code/Bus Route
- Train Car#/Bus#
- SOP# (P/I#, Policy#)
- Task Order/Job Code#/Work Order#



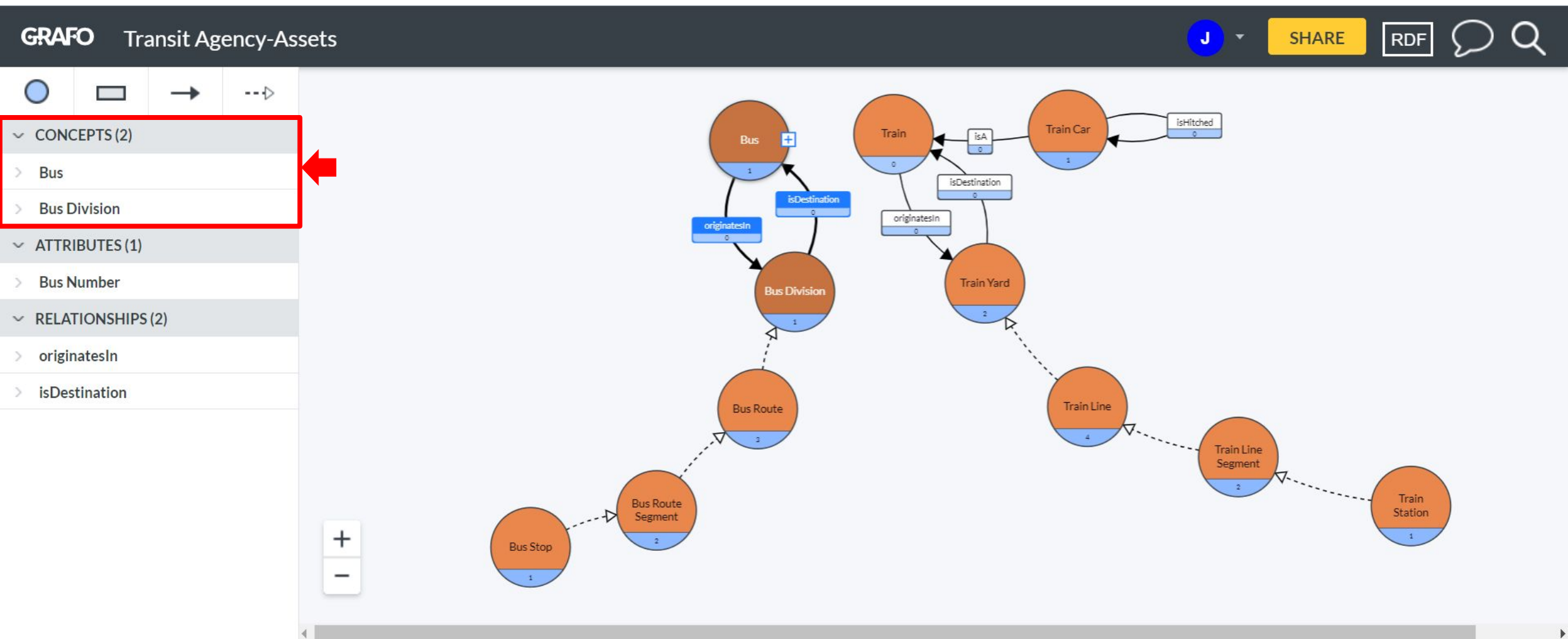
Graph of relationships between employees and policies



Graph of relationships between vendors and contracts



Graph of relationships among transportation assets



RDF behind the transportation asset graph

The screenshot shows the GRAFO editor interface for a Transit Agency. The main area displays an RDF schema with the following classes and their definitions:

```
153      rdfs:isDefinedBy <http://www.gra.fo/schema/untitled-ekg> ;  
154      rdfs:label "Train Station Identifier" .  
155  
156 #####  
157 # Classes  
158 #####  
159  
160 ### http://www.gra.fo/schema/untitled-ekg#Bus  
161 gf:Bus rdt:type owl:Class ;  
162      rdfs:subClassOf [ rdt:type owl:Restriction ;  
163                      owl:onProperty gf:originatesin ;  
164                      owl:someValuesFrom gf:BusDivision  
165                      ] ;  
166      rdfs:isDefinedBy <http://www.gra.fo/schema/untitled-ekg> ;  
167      rdfs:label "Bus" .  
168  
169  
170 ### http://www.gra.fo/schema/untitled-ekg#BusDivision  
171 gf:BusDivision rdt:type owl:Class ;  
172      rdfs:subClassOf [ rdt:type owl:Restriction ;  
173                      owl:onProperty gf:isdestination ;  
174                      owl:someValuesFrom gf:Bus  
175                      ] ;  
176      rdfs:isDefinedBy <http://www.gra.fo/schema/untitled-ekg> ;  
177      rdfs:label "Bus Division" .  
178  
179  
180 ### http://www.gra.fo/schema/untitled-ekg#BusRoute  
181 gf:BusRoute rdt:type owl:Class ;  
182      rdfs:subClassOf gf:BusDivision ;  
183      rdfs:isDefinedBy <http://www.gra.fo/schema/untitled-ekg> ;  
184      rdfs:label "Bus Route" .  
185  
186  
187 ### http://www.gra.fo/schema/untitled-ekg#BusRouteSegment  
188 gf:BusRouteSegment rdt:type owl:Class ;
```

Red boxes highlight the URIs for each class, with red arrows pointing to the label "URI":

- URI: `http://www.gra.fo/schema/untitled-ekg#Bus`
- URI: `http://www.gra.fo/schema/untitled-ekg#BusDivision`
- URI: `http://www.gra.fo/schema/untitled-ekg#BusRoute`
- URI: `http://www.gra.fo/schema/untitled-ekg#BusRouteSegment`

Summary of path to a transit agency knowledge graph

- Started as a project to digitize legacy paper records
- Generated separate metadata specs for each business unit
- But applied common naming conventions as much as possible
- Created a concordance of all the metadata attributes
- Generated examples of knowledge graphs that could be applied enterprise-wide

Knowledge graph vs. BPM

- The goal of a business process model is to specify a repeatable business process.
- A knowledge graph is about specifying named entities with attributes that facilitate the reliable retrieval of things.

Questions



Joseph Busch, Principal Consultant,
jbusch@taxonomystrategies.com

