Towards a Big Data Taxonomy
v. 30_September_2013

Bill Mandrick, PhD

Data Tactics Corp

wmandrick@data-tactics.com

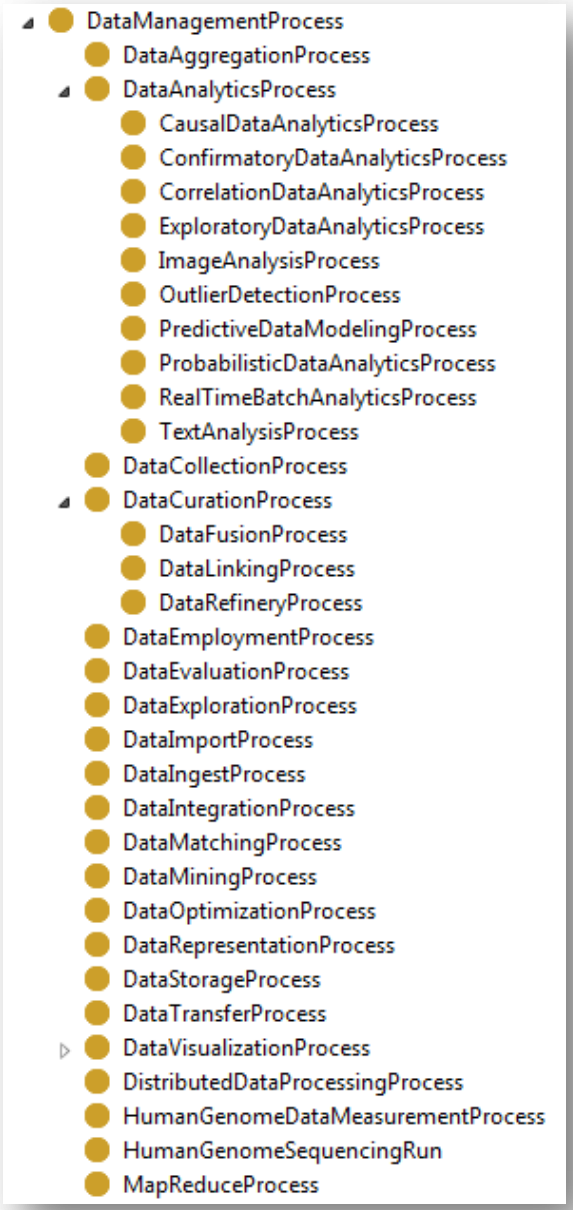
Scientific Taxonomies Represent

- Types of Processes (e.g. Data Management Process)
- Types of Continuant Objects
  - Physical Objects (e.g. Disk Drive, Building, etc.)
  - Information Content Entities (e.g. Measurement, etc.)
- Types of Characteristics
  - Qualities
  - Roles
- Relationships
  - Between Processes
  - Between Objects
  - Between Characteristics

Towards a Big Data Taxonomy

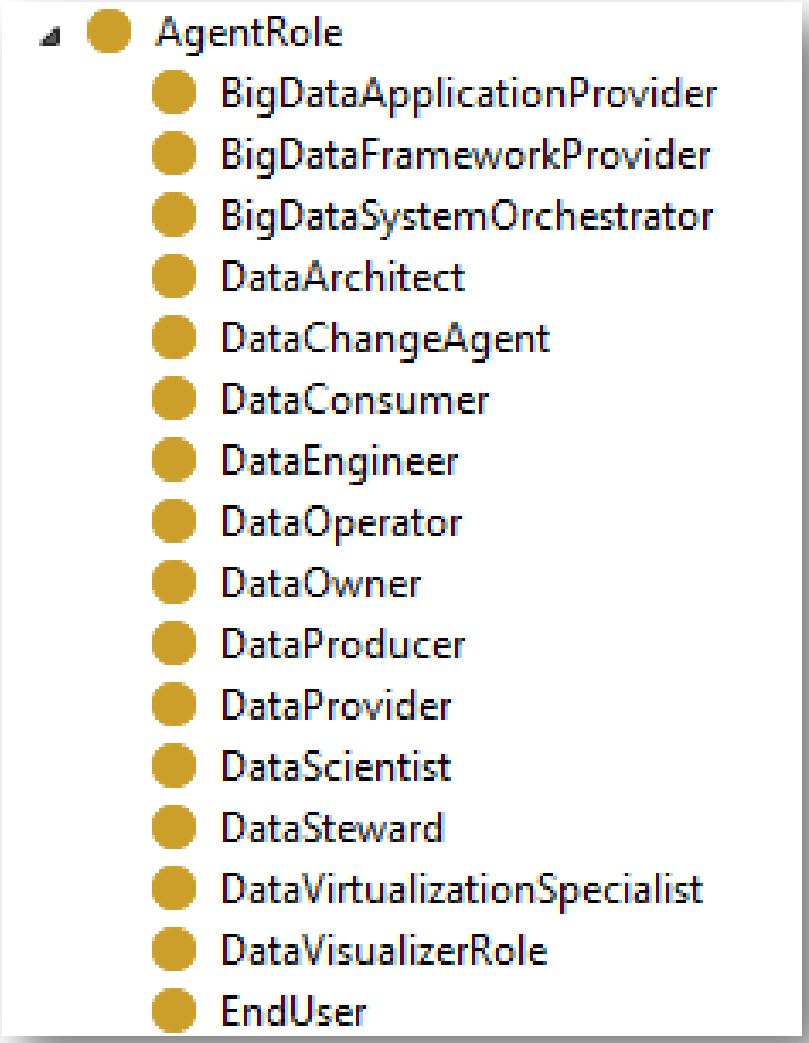
- Big Data Related Processes
  - Data Analytics Process
  - Data Curation Process, etc…
- Big Data Related Agent Roles
  - Big Data Application Provider
  - Big Data Framework Provider, etc…
- Big Data Related Information Bearing Entities
  - Computer Disk
  - Printed Paper Copy, etc…
- Big Data Related Information Artifacts
  - Big Table
  - Data Directory, etc…

Data Management Processes



Each of these Data Management Processes can be decomposed into (sub)processes

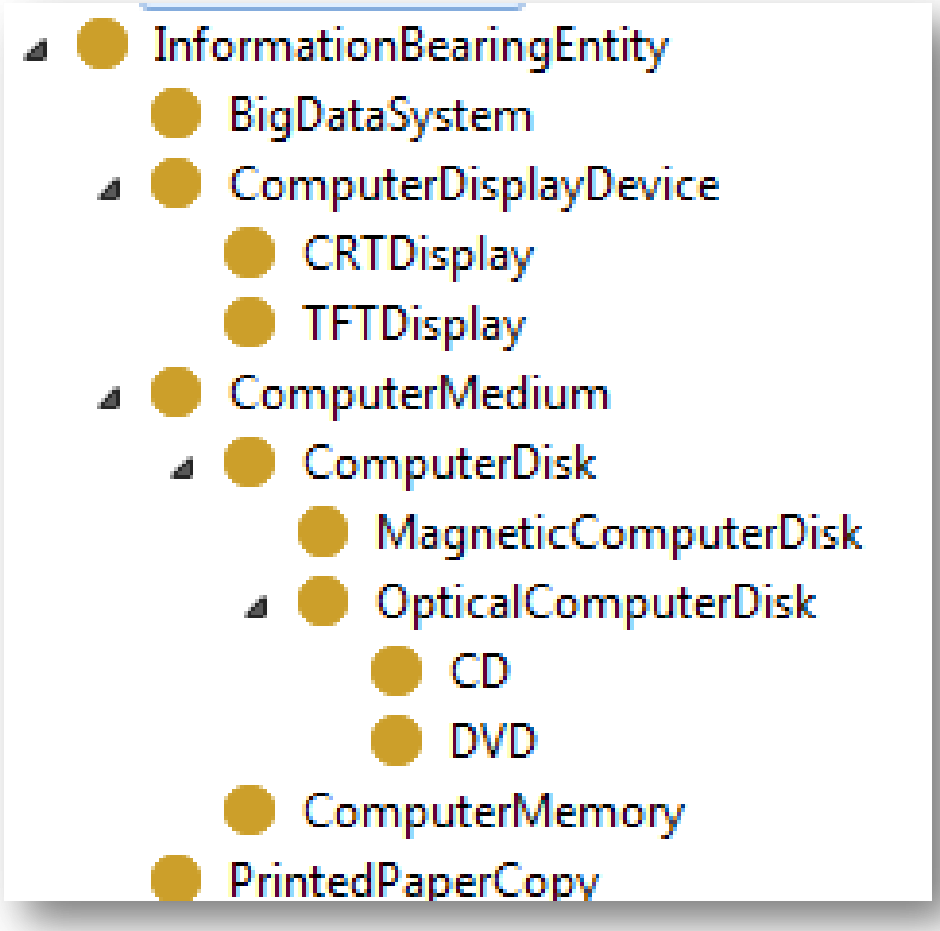
Big Data Agent Roles



The Roles of Persons and Organizations

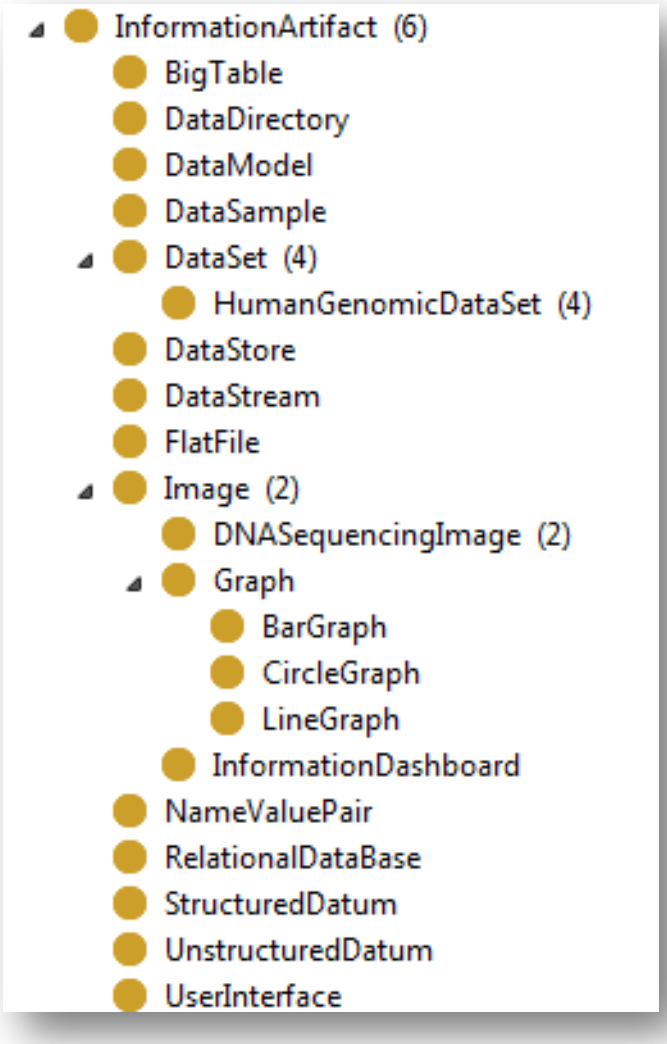
An Organization or Person can be in multiple roles at the same time, e.g. Data Producer and Data Consumer

Information Bearing Entities



Information Bearing Entities (IBE's) are the Physical Bearers of Big Data and Information

Information Artifacts

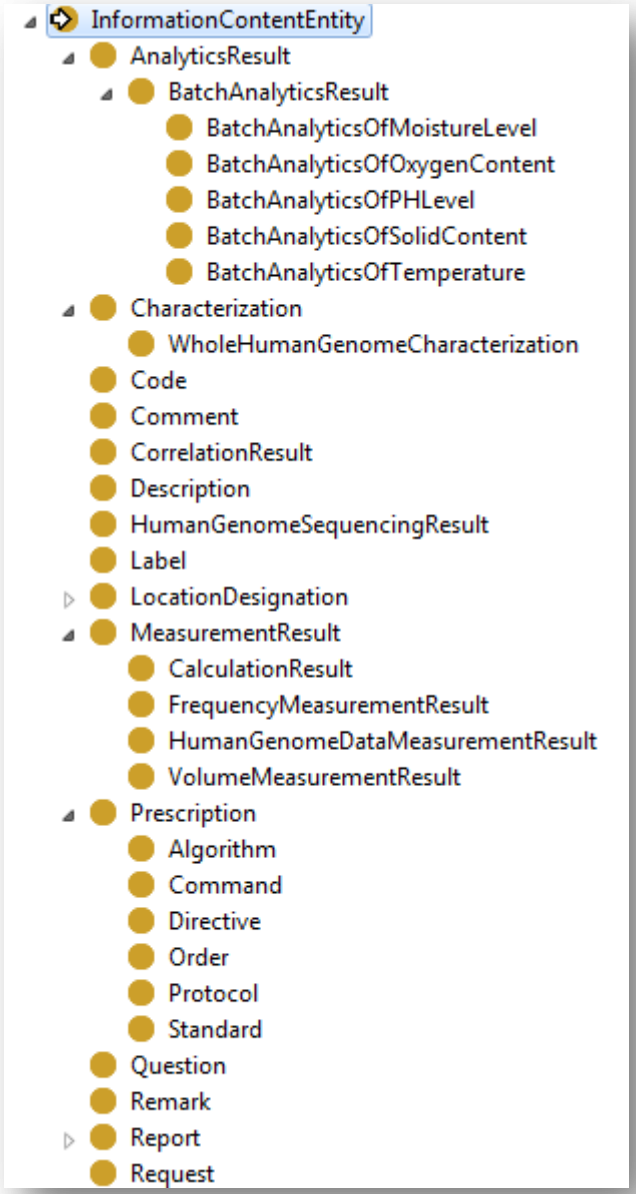


An Information Artifact is a composite of

i) some Physical Bearer and

ii) the Information Content Entity

Information Content Entities

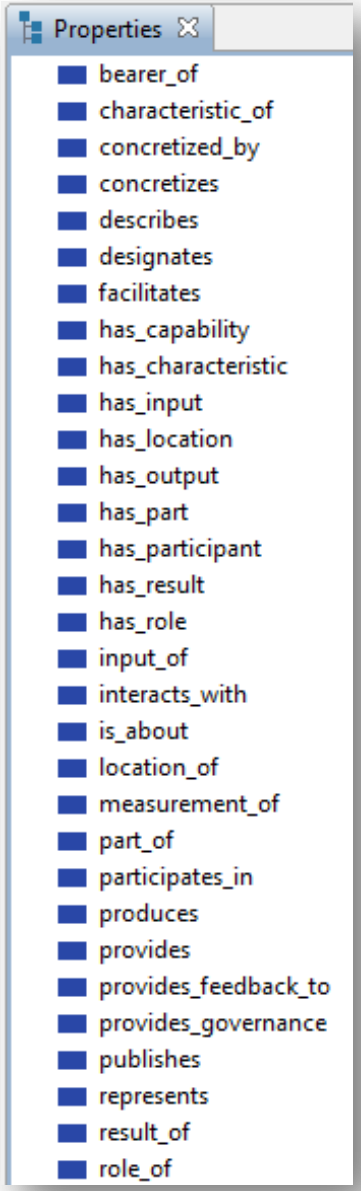


An Information Content Entity (ICE) is the product of some Information Producing Activity such as Analyzing, Measuring, Describing, Commanding, etc.

Information Content Entities are concretized by the Information Bearing Entities, but they are distinct from them.

Information Content Entities are the ***meaning*** that results from some Information Producing Activity.

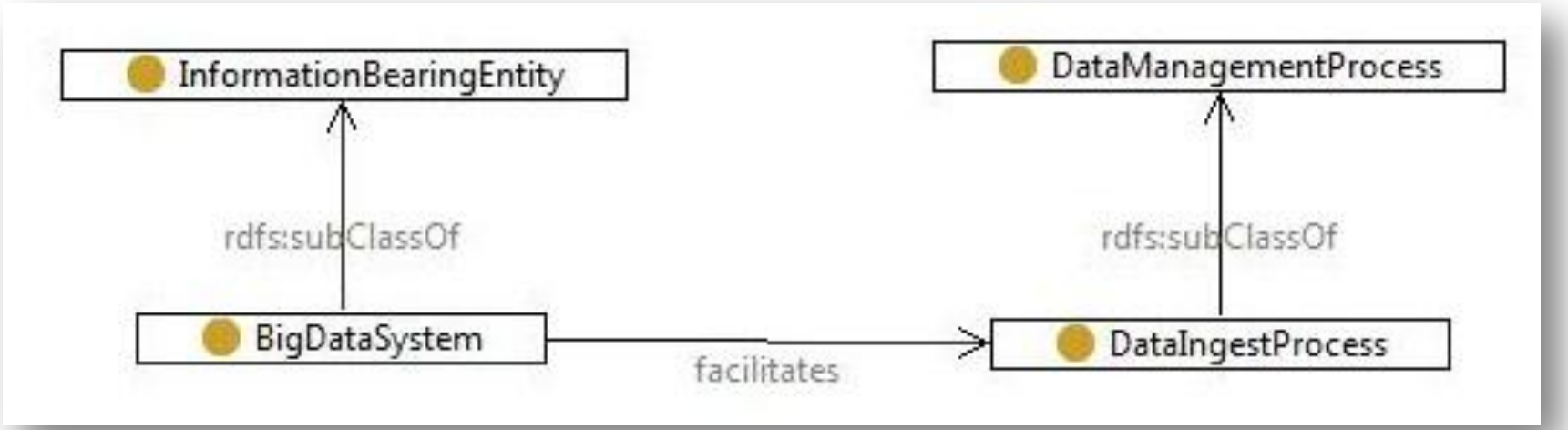
Standardized Relations



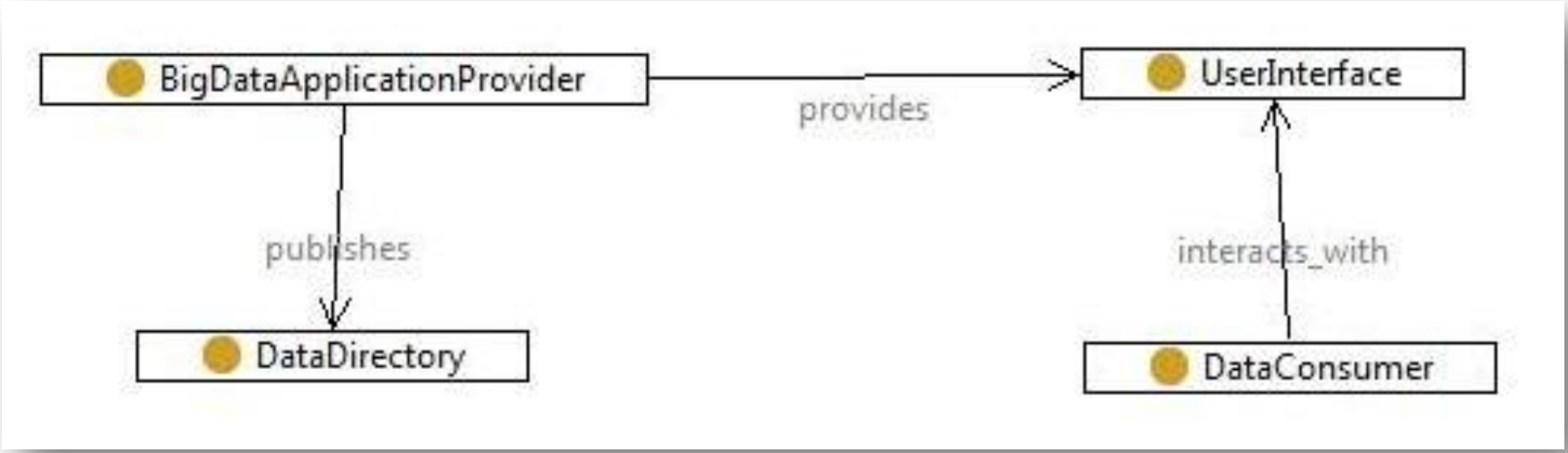
The use of standardized relations facilitate reasoning (inferencing) about the Big Data domain.

See next few slides for examples of basic inferencing about the Big Data Domain.

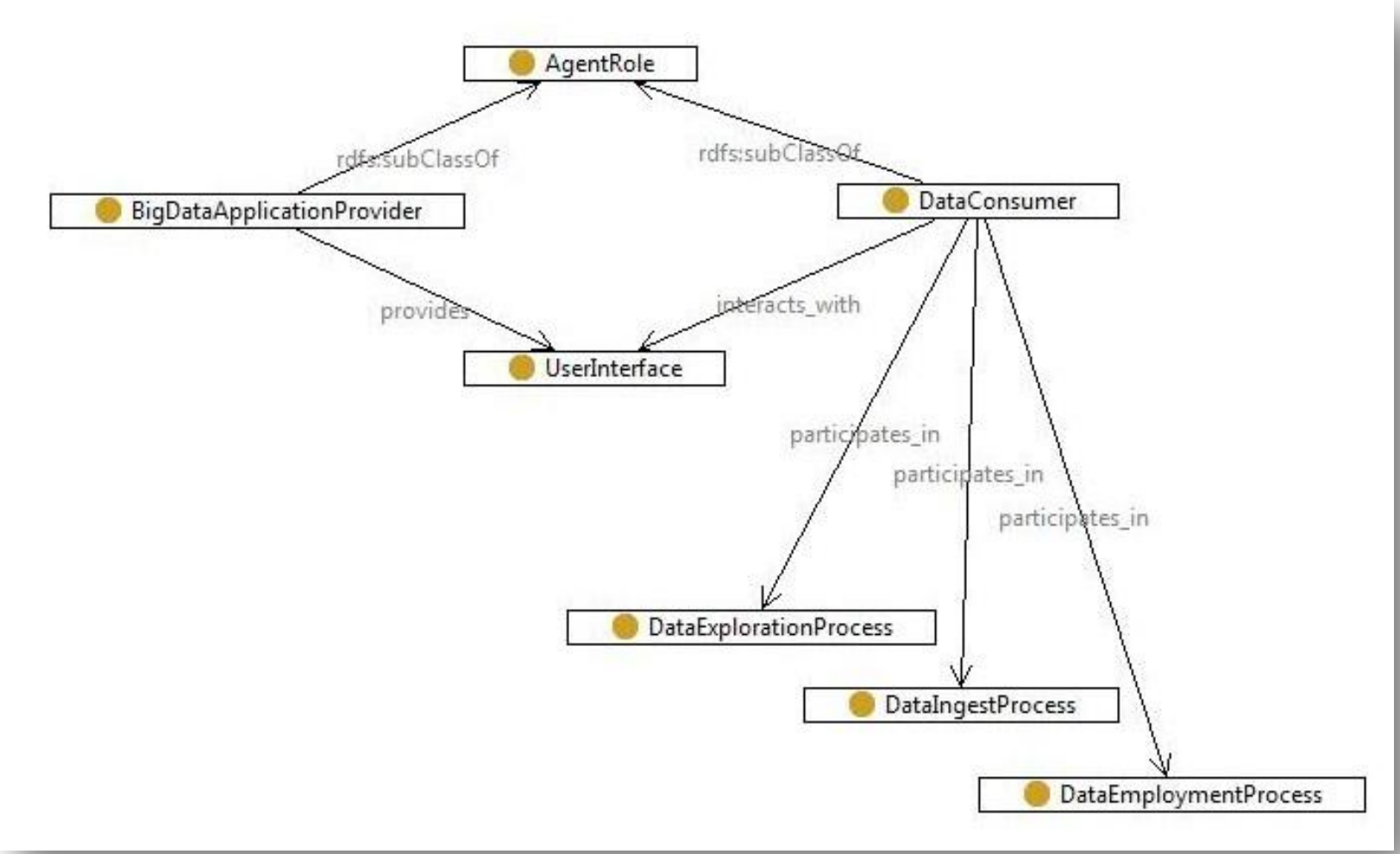
Data Systems and Data Management



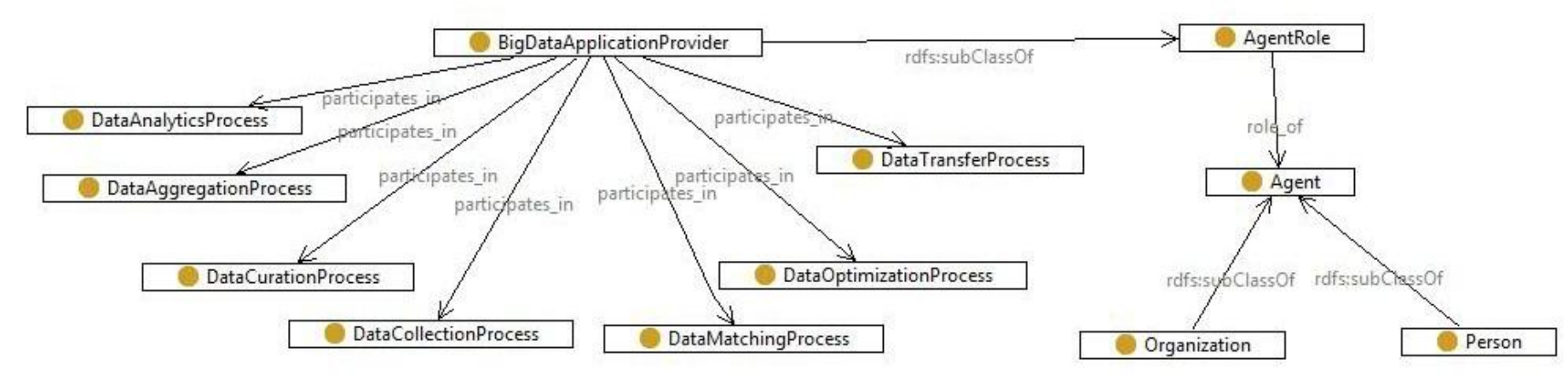
Big Data Application Provider



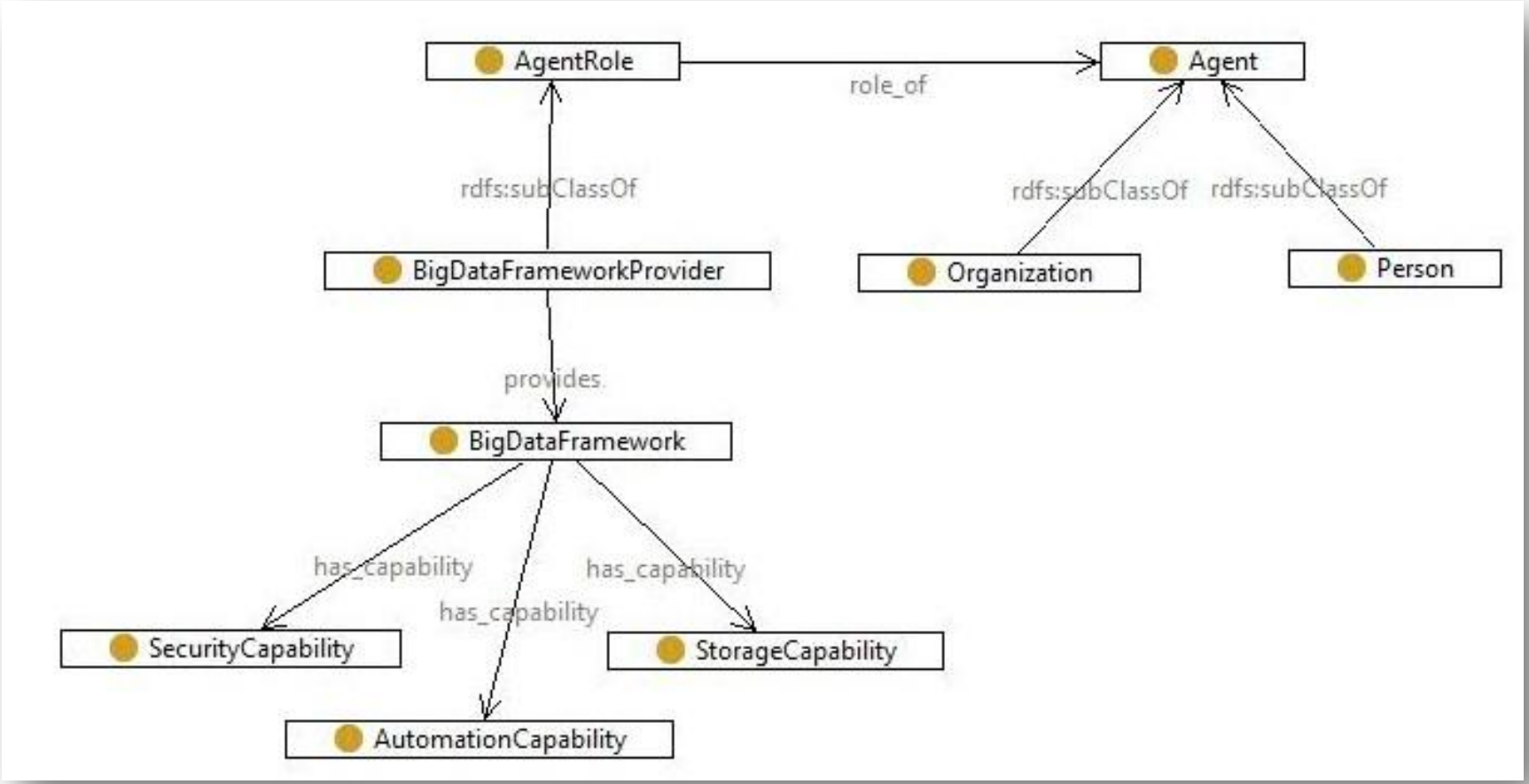
Big Data Application Provider



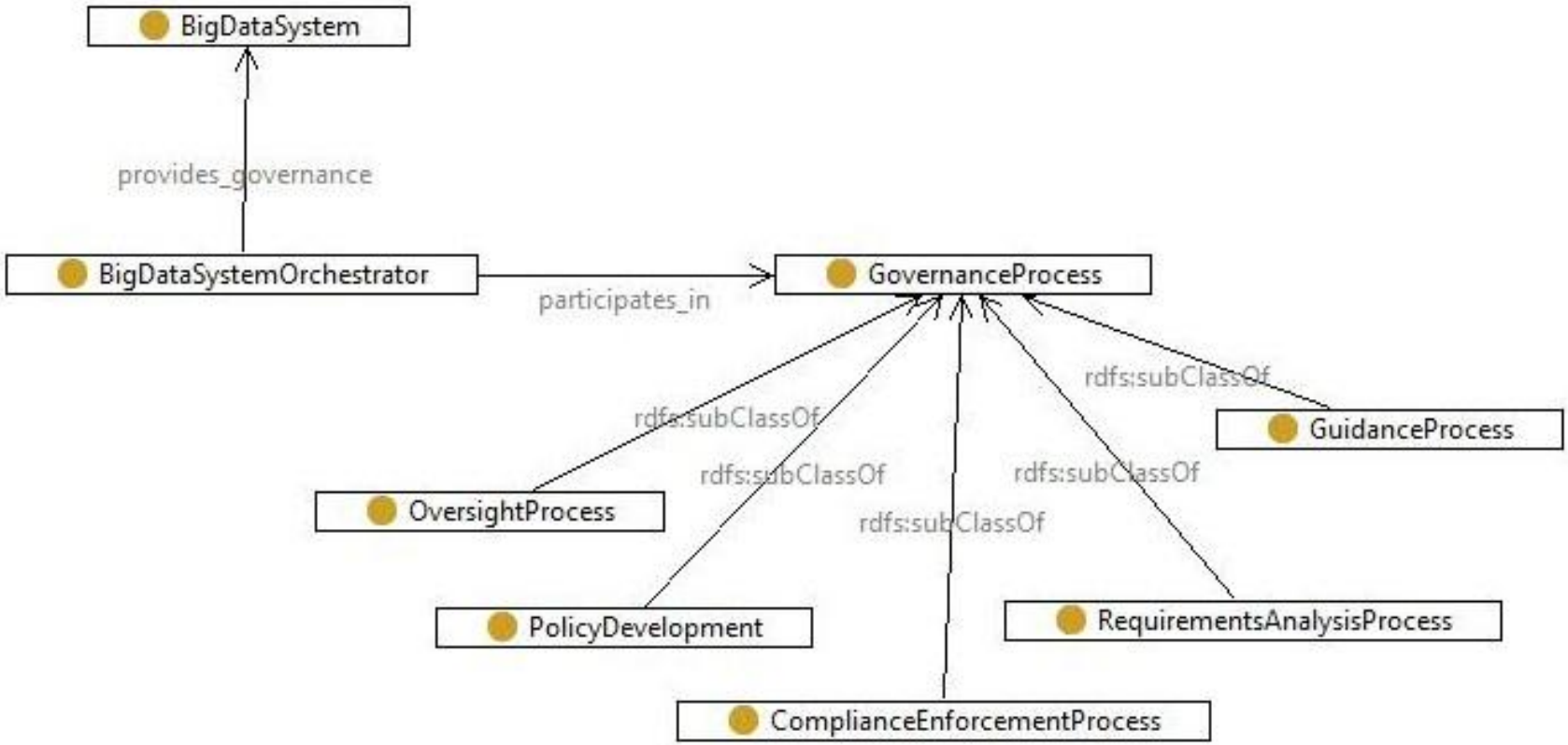
Big Data Application Provider



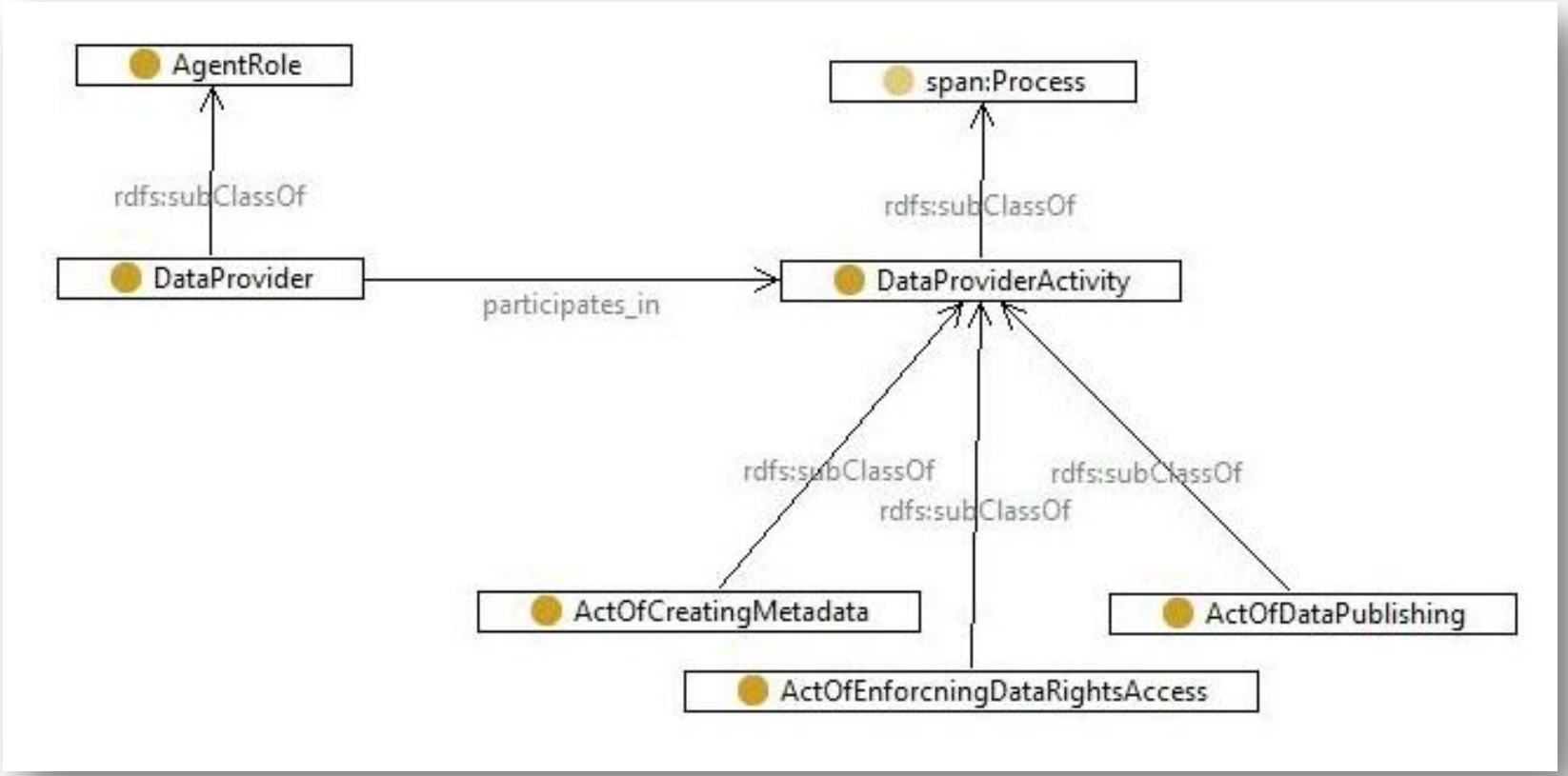
Big Data Framework Provider



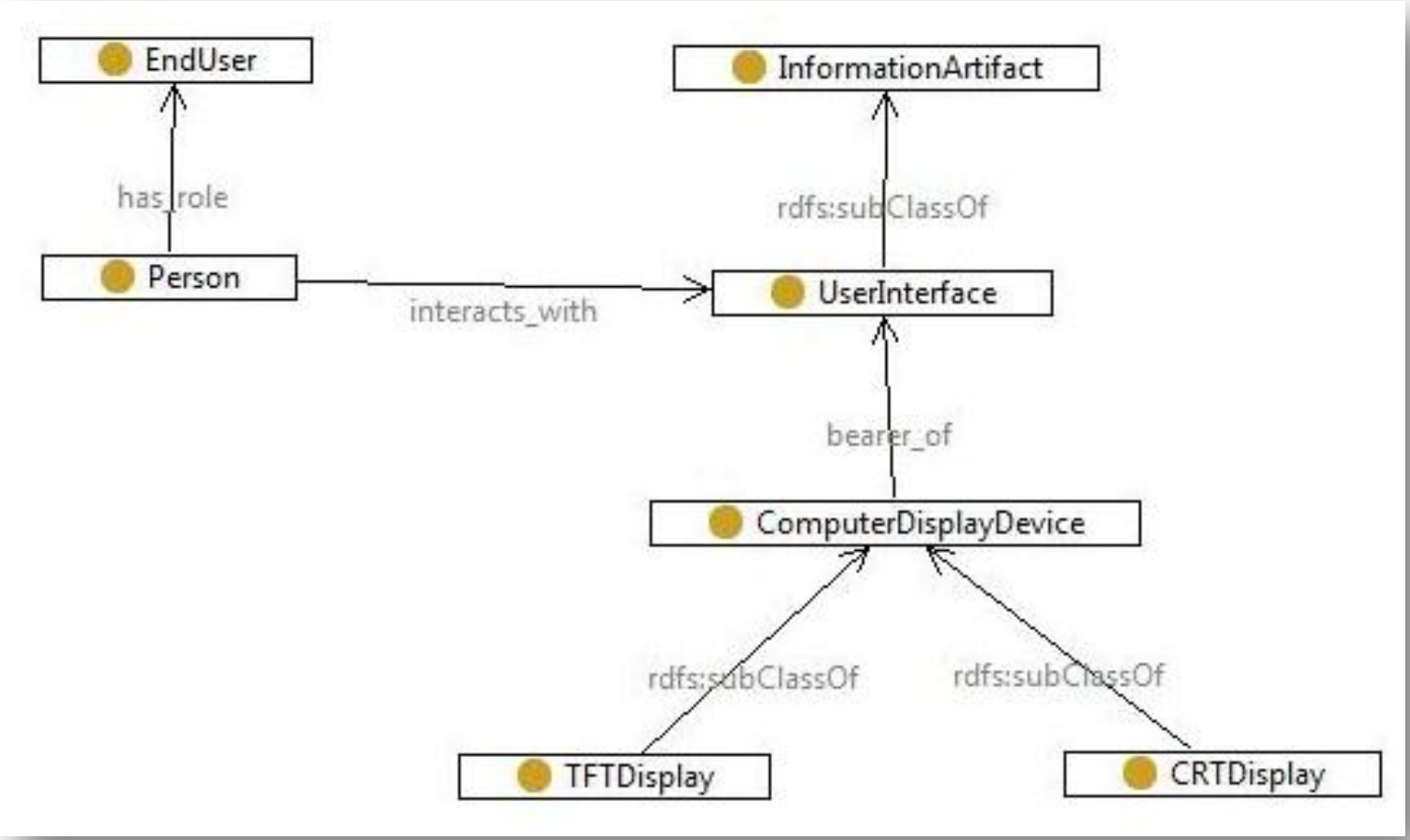
System Orchestrator



Data Provider Activities



User Interface



Conclusion

- This method can be done for any part of the Big Data Taxonomy
- Need SME input for various areas/domains
- Need to add definitions in owl
- Need to expand set of standardized relations
- Link ***instances*** to the taxonomy (e.g. actual data sets, batch analytics data samples, etc.)

Questions or Comments
Bill Mandrick at:
wmandrick@data-tactics.com