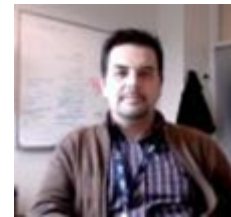
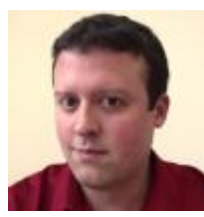


# Refactoring the eagle-i and VIVO Ontologies to Create the ISF

Shahim Essaid, Jon Corson-Rikert, Brian Lowe,  
Melissa Haendel, Carlo Torniai



# Background

- The eagle-i and VIVO ontology teams recognized in 2009 that in eagle-i there were people attached to resources, and in VIVO, people wanted to represent research resources
- Several joint meetings were held during the ARRA funding period (2009-2012)
- Both software applications are ontology-driven in addition to storing data as RDF

# 2011 Joint ICBO Poster

- Recognized the benefit of aligning under a common upper ontology (BFO)
- Demonstrated existing overlap while recognizing a need to more closely examine similarities and differences
- Acknowledged issues in common, including the need for shared instances and identifiers on people, organizations, and resources

# Overview of ontology changes

- Refactoring ERO and VIVO ontologies
  - Adopting BFO as an upper ontology and new OBO relationships
    - being more explicit about processes, role, and relationships
  - Adopt the VCard model to group contact attributes associated with a person, e.g., from a given affiliation
  - Shared instances and vocabularies
- Supporting clinical expertise
  - clinical encounter module
  - expertise measurement module

# Aligning with the BFO upper ontology

- ERO used BFO from the start and VIVO had a BFO mapping from 2011
- The current work brings VIVO closer to other BFO based ontologies, including ERO
  - *e.g.*, Ontology of Biomedical Investigations (OBI) and the Information Artifact Ontology (IAO)
- Sharing a common upper ontology simplifies the reuse of classes and enables better semantic links to existing ontologies and data
- Alignment involves properties, not just

# The BFO enforces high-level modeling principles for representing

- Objects that continue through time
- Their qualities, roles, functions, etc.
- How they interact with each other in processes
- Location and temporal entities

# Refining an approach for reifying relationships

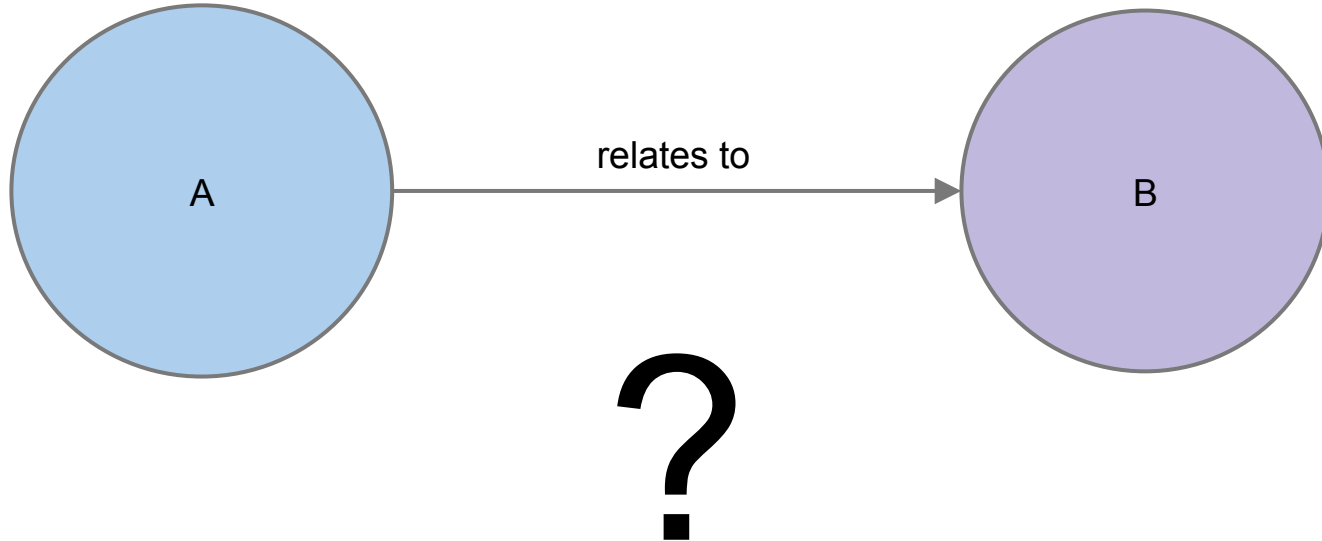
- There is a need for modeling relationships without implying ongoing interactions (i.e. processes)
  - For example, one person may be assigned as the mentor of another person (a relationship)
- The BFO model supports roles in the processes of actual mentoring interactions
- The ISF extends the BFO model with a high-level "Relationship" class to capture the ongoing process-independent relationships between entities

# Relationship examples include ...

- **Positions** relating a person to an organization with a title and related roles over a period of time
- **Grants** relating a PI, a funding source, the administering department, and projects supported
- **Credentials, degrees**, etc. showing one person recognized in a specific way in a specific context

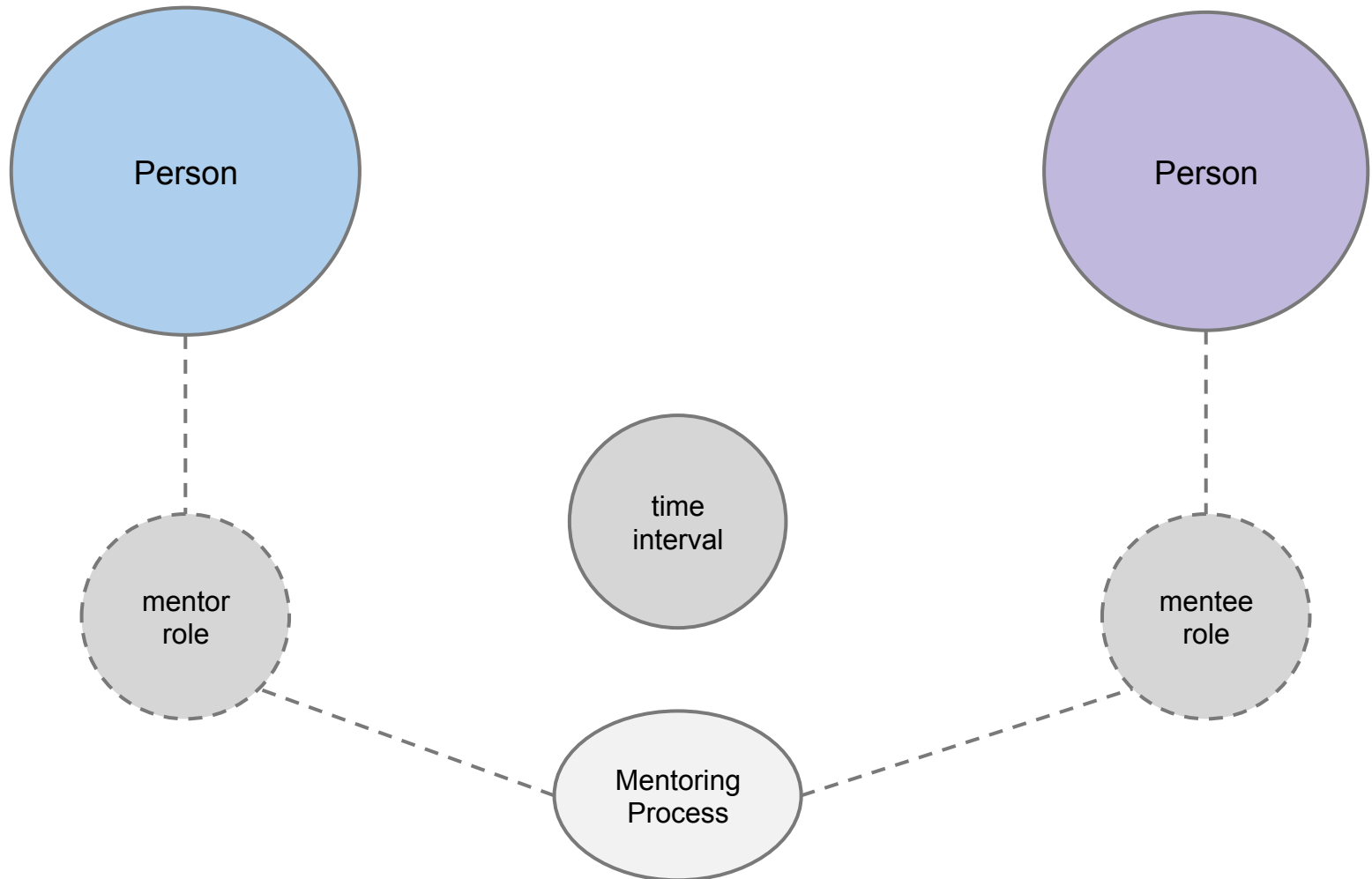


# Binary RDF relations fall short

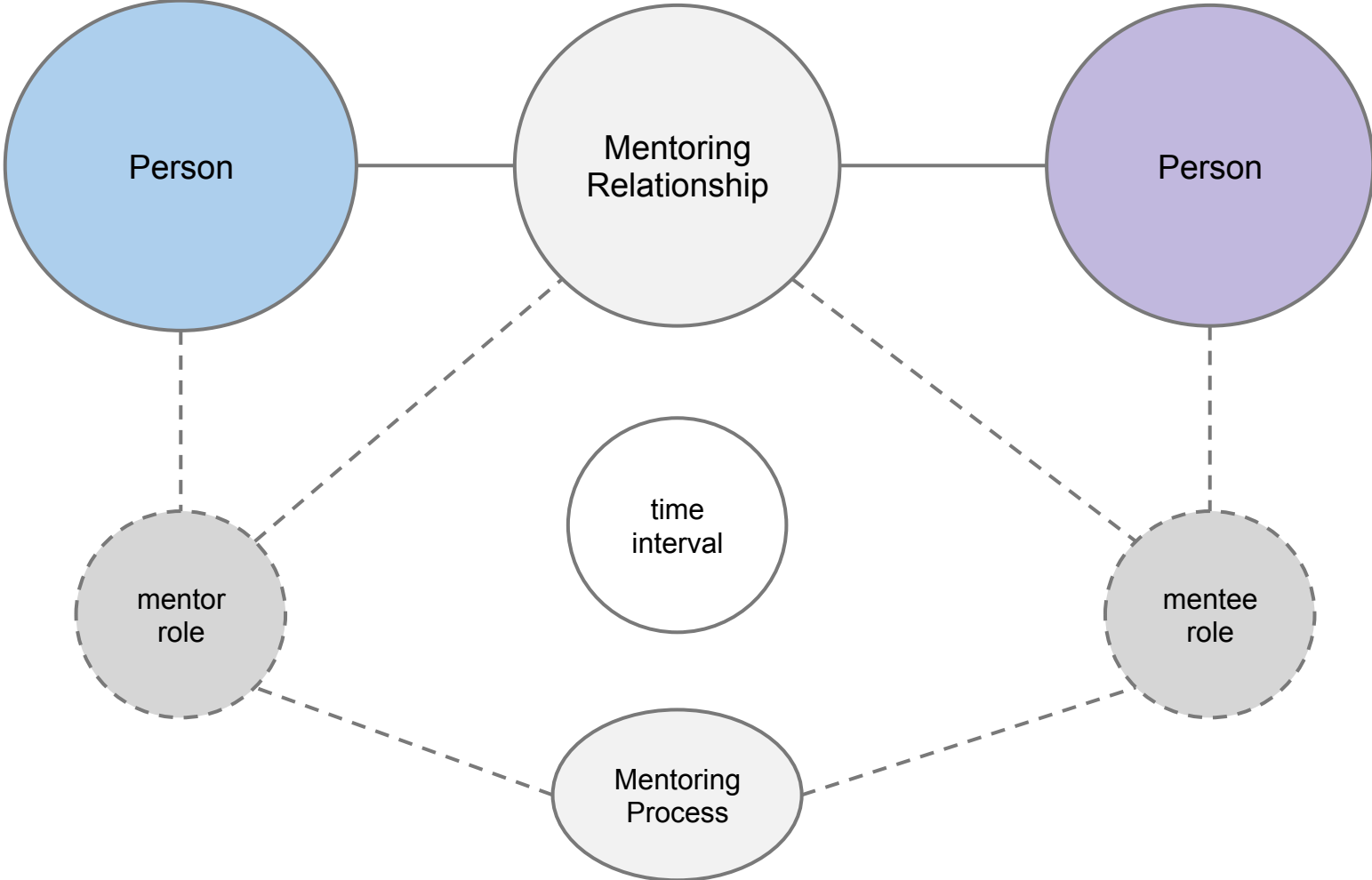


- For what time period?
- With what role?
- With any additional attributes?
- Who is asserting this relationship and in what context?

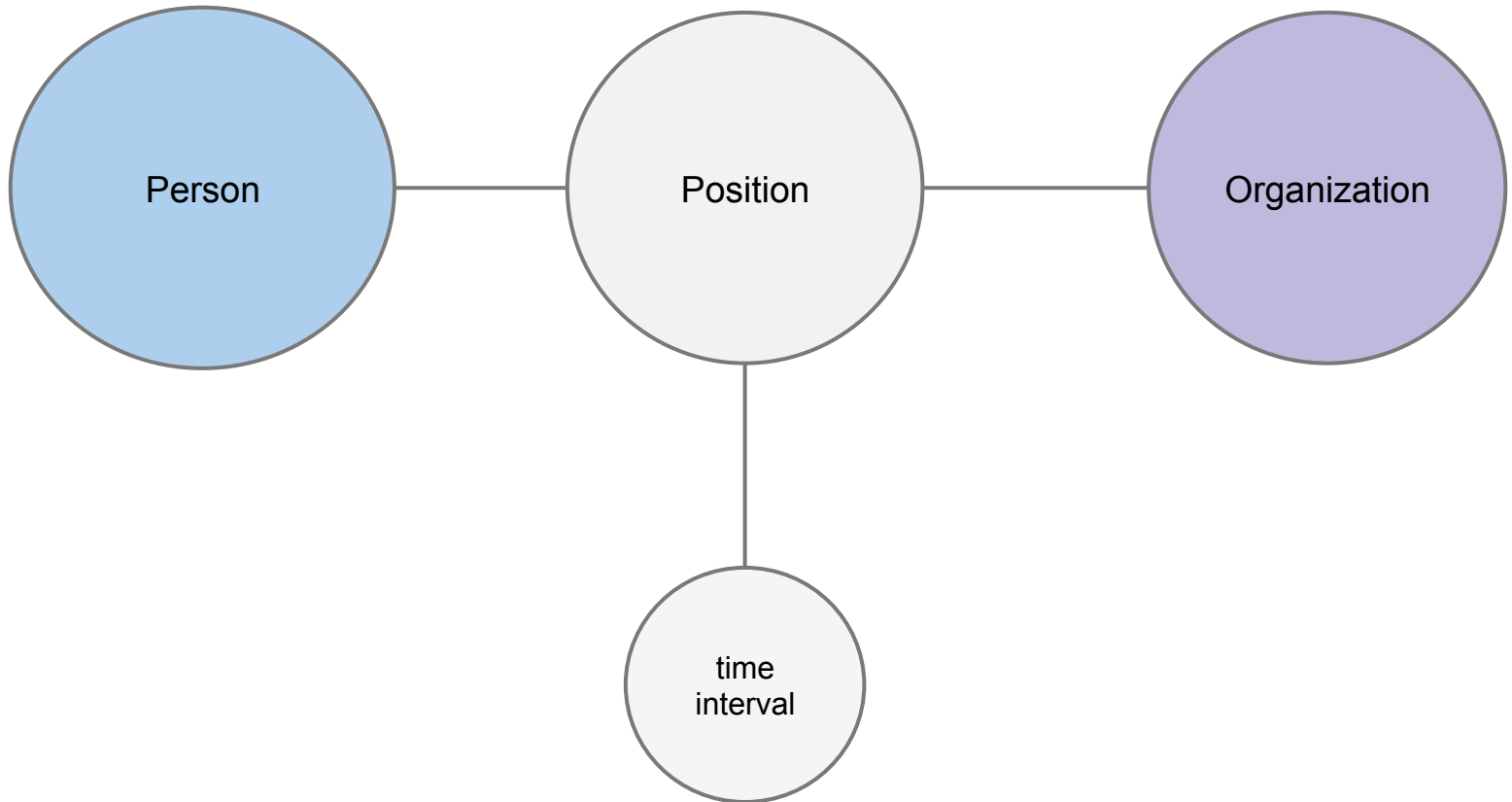
# The BFO pattern for roles in processes



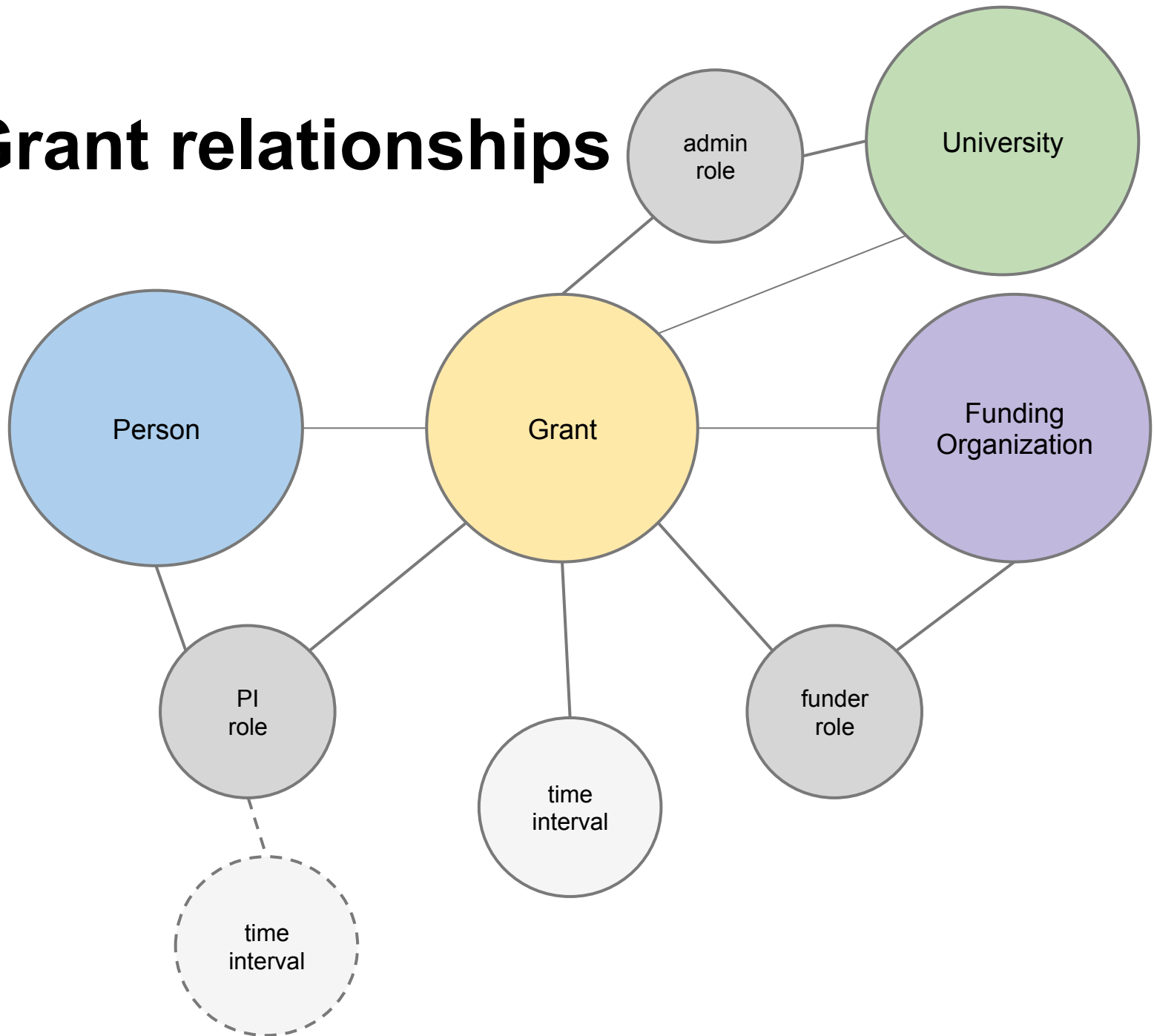
# The additional Relationship pattern



# Position relationships



# Grant relationships



# Using a common set of OWL properties

- Adopting the obo-relation ontology from <http://code.google.com/p/obo-relations>
- This is an intermediate (and BFO oriented) effort for developing a shared set of OWL properties
- BFO 2.0 might cause another change for this set of properties, but there are many issues to be resolved
- VIVO will be adopting them and ERO will migrate to the new identifiers in order to have a shared set in the ISF

# Standardizing an approach for "vocabularies"

- Based on the SKOS OWL vocabulary (a concept model)
- Both VIVO and ERO will migrate to this new vocabulary model
- An example:
  - A person having their own instance of a PhD degree (in actual existence), vs.
  - being a candidate for the "PhD degree" concept (a reference to the type of a potential degree)
  - the identifiers could be the same but their "logical" meaning would be different (OWL punning)
- A core set of vocabularies will be maintained alongside the ISE but end users could add to or

# Vocabulary examples

- ICD9, CPT, and other medical coding systems
- Degree types
- Credential and Award types
- Date/time resolution specifications
- Document statuses

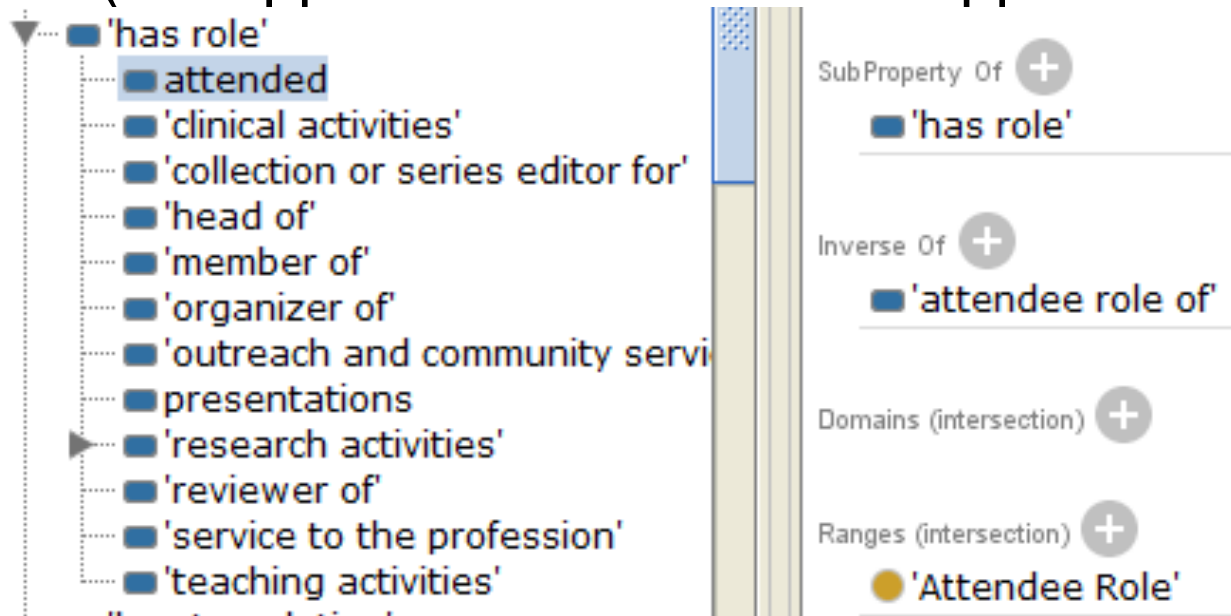


# ISF-based shared instance data

- There is no existing repository for reusable instance identifiers
- The ISF will provide a set of instances (OWL named individuals) with URL identifiers
  - Medical credentialing organizations and programs (instances of organization)
  - Medical specialty boards
- The goal is to provide a sharable set of instances and their URL identifiers
- These are instances of ISF classes, as opposed to instances of SKOS concepts in the vocabulary model

# Removing several object properties

- Several properties are logically redundant but were intended to support the VIVO application and RDF queries
- The ISF ontology will not include them but they could be added as application specific extensions where needed (or supported as labels in an application)



# Adjusting class definitions

- Several defining axioms were noted to be application oriented in nature, or too constraining

The screenshot displays a class hierarchy in a software tool. On the left, a tree view shows the following classes: Project, Relationship, 'Advising Role', Authorship, 'Research Opportunity', Role, 'Attendee Role', 'Clinical Role', 'Editor Role', 'Leader Role', 'Member Role', 'Organizer Role', 'Outreach Project', and 'Presenter Role'. The 'Project' class is selected. On the right, the 'SubClass Of' list for 'Project' is shown, containing the following defining axioms:

- 'contact information' **only** Literal
- 'date/time interval' **only** 'Date/Time Interval'
- 'domestic geographic focus' **only** 'Geographic Region'
- 'funding provided via' **only** Agreement
- 'geographic focus' **only** 'Geographic Region'
- 'international geographic focus' **only** 'Geographic Region'
- description **only** Literal
- participant **only** Role
- produces **only** 'Information Resource'
- webpage **only** URLLink

# Migrating to numerical URLs

- Using numerical identifiers ( "BFO\_0000050" vs. "part of") is a recommended practice but:
  - It is an obstacle for humans reading RDF data
  - It makes writing RDF queries less intuitive
  - Existing applications and queries will have to adjust
- **Benefits**
  - Increase the interoperability with other Linked Open Data datasets
  - Adhere to good development principles (labels can change while identifiers stay constant)
  - Reuse existing OBO relationships

# Data migration

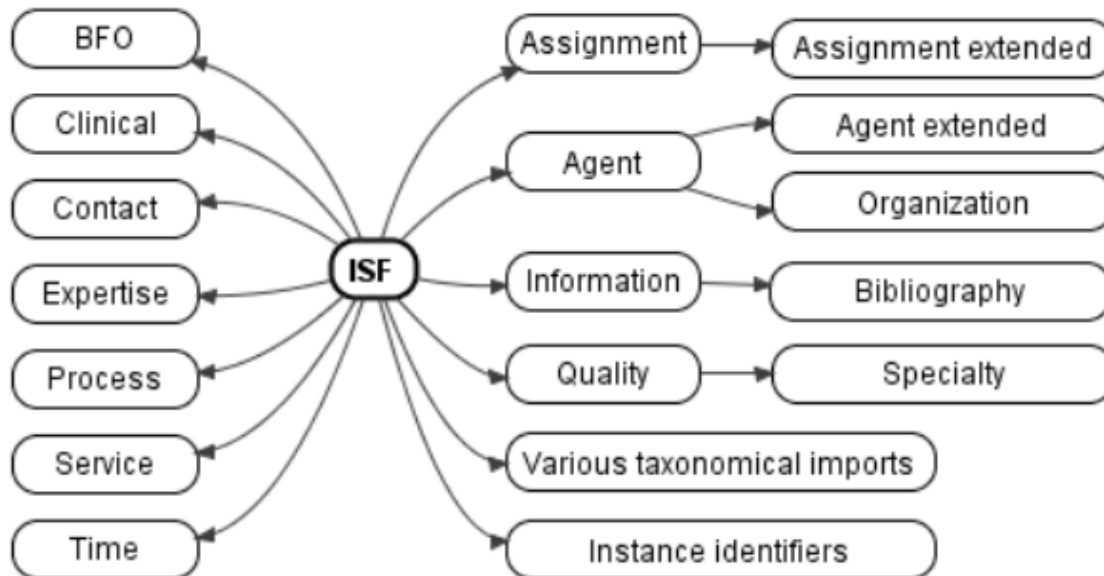
- VIVO and eagle-i software platforms are being updated to ISF in upcoming releases
  - existing data will be migrated as part of routine application upgrade processes
  - manual review will only be required in cases where existing types or properties are split
- Other tools that presently output VIVO RDF will have the option to adapt natively or batch convert exported data from VIVO 1.5 to VIVO 1.6 (ISF)
- ISF data is linked open data (LOD) and applications written generically for LOD should tolerate ISF additions, ignoring what is not expected

# ISF beta release

- Expected for the end of April, with draft documentation
- Beta release files will still reflect the existing VIVO and ERO files (with refactored content) and few new files or modules
- A single top level OWL file will import and show the full ISF
- The ISF will also be packaged in modules for the final release to simplify reuse of smaller components
  - modules are derived from the ontology and may overlap where convenient in a manner similar to database views
- Feedback on desired modules would be very helpful

# Proposed content modules

- Modules provide a simplified and self contained view of part of the ISF
- Each module will have a visual diagram and an example use case with sample RDF data
- Written documentation will suggest how the module should be applied or extended



# Integration plans

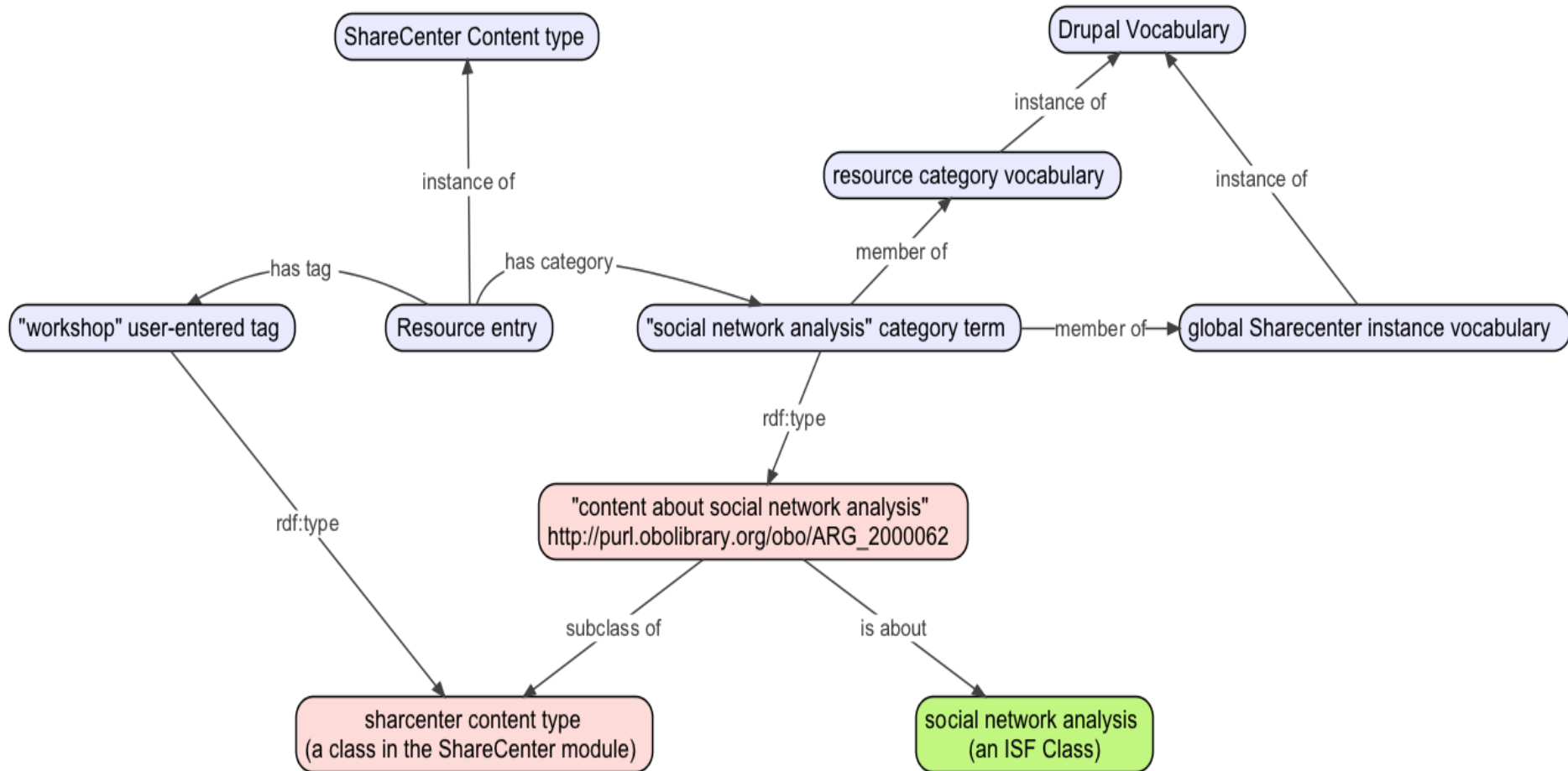
- The eagle-i and VIVO applications have started the integration of the ISF
- This effort will involve validation and testing of the beta release to help prepare for the final release
- This will be a slow and ongoing process but we hope to have access to ISF compliant data soon even if it is not directly generated by their corresponding applications



# ShareCenter integration

- ShareCenter integration was accomplished by defining a Drupal tagging vocabulary
- Each tag can relate the tagged content to one or more ISF classes through OWL definitions
- This approach provides a level of indirection and avoids the inclusion of "tagging" classes in the core ISF ontology
- The RDF representation of the Drupal tags, and the ShareCenter ISF-based module, provide the links in the generated RDF data

# Sharecenter integration



# Summary

- A beta release by the end of April
- The ontology integration effort led to:
  - more structural changes in VIVO as compared to ERO
  - new URLs for many properties
  - adopting a reified relationship pattern and using it consistently
  - adding a clinical aspect
  - developing a better model for shared vocabularies and instances
- Application and data integration is ongoing
- The goal is to have a final release by August

**Questions & comments?**