



Annual Conference of ITA

ACITA 2009

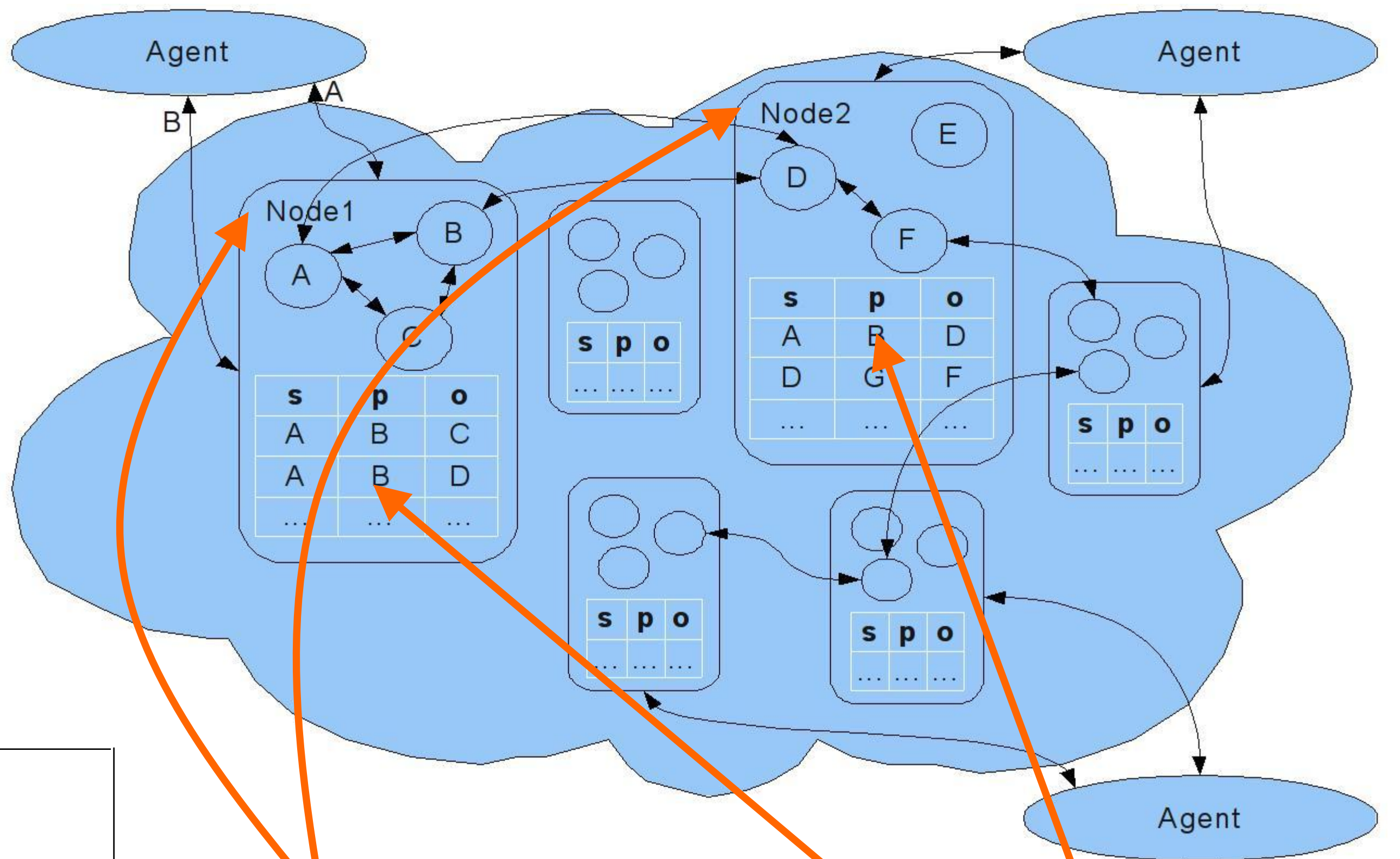


GIDS: Global Interlinked Data Store

Gareth Jones (IBM UK), Dave Braines (IBM UK), Paul Smart (Southampton), Jie Bao (RPI), Trung Dong Huynh (Southampton)

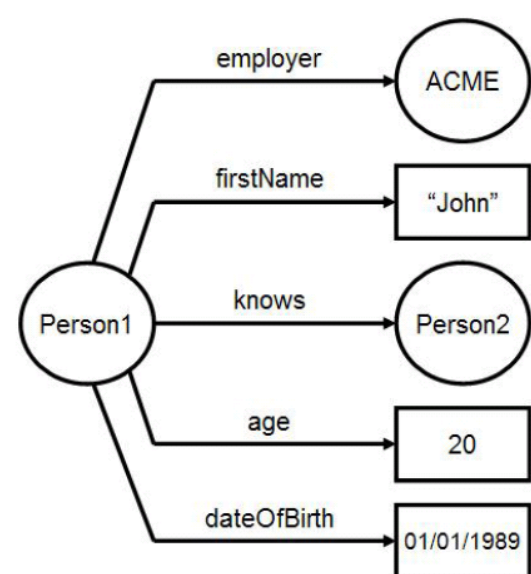
A new technique for semantic data access (RDF) in support of the Semantic Web and Linked Data Web...

Enables data to be easily stored *directly on the network* through simple network requests, to be retrieved in the same way with a *known response format* to facilitate subsequent automated processing, and enables each entity referenced by a triple to *store a separate copy* of that triple to facilitate improved navigation.



- Node 1 defines and stores information related to three entities (A, B, C)
- Node 2 defines and stores information related to three entities (D, E, F).
- The triple of particular interest (A, B, D) is stored against Node1 and Node2 because Node1 defines entities A and B, whereas Node2 defines entity D.

Triples are statements with a **subject**, **predicate** and **object** which can be literal or non-literal values.

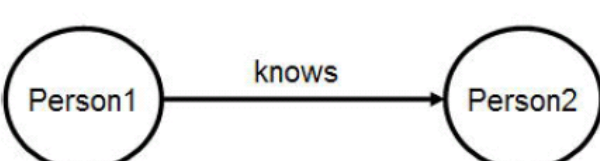


An example entity, comprised of numerous triples

Entities are collections of triples which each share the same non-literal subject value.



An example triple with a literal value



An example triple with a non-literal value

Core capabilities:

- A defined response format
Specifically RDF in the sample implementation
- Direct triple access interface
Via resolution of the dereferenceable URIs
- Multiple storage locations
At each URI location specified within each triple

Advantages:

- Data visibility
Dereferenceable URI usage, coupled with data distribution greatly increase the opportunity for discovery of data through linkages
- Data access
Potential entry points to the data from any of the three locations in the data triple
- Data ownership
Data is stored locally to the entities which are referenced within the data
- Data redundancy
Storage of multiple copies of the same data triples provides some level of natural redundancy

Disadvantages:

- Efficiency
Multiple storage of data potentially increases number of requests to retrieve all known facts
- Reliability & trust
Distributed network environment (WWW) is inherently unreliable and potentially untrustworthy
- Privacy
The multiple potential storage locations for each triple can lead to concerns about privacy.h

Assumptions:

- Dereferenceable URIs
Building upon the notions of the linked data web, all URIs should be dereferenceable and return a defined response format
- Anonymous / blank nodes*
These cannot be supported by the GIDS since they do not have universally unique identifiers.
- Inference / entailment support*
Not appropriate for an RDF storage layer

* Note that these capabilities are assumed to be delegated to the application or agent using the GIDS

