OKN Breakout: Horizontal

Design for Evolution

Our philosophy: at all points, system should enable gradual adoption, with low costs to publishers and users

That means:

- No heavy syntax (e.g., avoid heavy reification)
- No required coordination between publishers
- Limited novel languages

Design Decisions

Endpoints offer a query interface, not HTTP-style file transfer.

Data dumps are possible, not required

Data model: “Triples++”. Triples with defaultable extensions for time, space as well as “metadata attributes” like provenance, trust scores

Time, space often make sense, but not for certain important domains

Query model allows traditional record retrieval, aggregates, etc

All queries and results are processed in “best effort” manner, similar to HTML

Moving Parts

Data Publishers provide query endpoints

Integrators are subtype of Data Publishers: they draw on other Data Publishers

Leaderboard tracks data access, publicizes successful datasets, encourages schema convergence

Each part needs new software. The necessary software isn't scientific research, but perhaps best built in research-friendly environments

Giri: The Browser Analogue

Browsers provide different interfaces to users

The general-purpose Siri (Giri) is one way to access endpoints. It’s an NL interface to the universe of data

Unlike Siri, can be pointed at data-offering endpoints

Use Case

Patient has Mutation A, Disease B. Dr asks: what other bad things can happen?

Patient has Mutation A, Diseases B and C. Dr asks: what drugs are suitable?

We need:
1. A natural interface for medical providers, and
2. A structured database of medical knowledge from researchers

Need to kick off virtuous ecosystem of improving data access, data publishing

Community TODO List

Software:

- Giri (browser equivalent)
- Query server (httpd equivalent)
- Integrator Service (search engine equivalent)

Research Questions:

- User-friendly Data Integration, Entity Resolution
- Trustworthy Trust Scores