Mediator Group

Completeness

Why is completeness so important?

- What answers do users expect?
  - Fullest answer set.
  - Some answer set.
  - Some indication of fullness with answer.
- Do we need a mediator?
  - Selling point of R-GMA is that we have mediation.

When might tuples be inaccessible to a query?

- Network/computer failures. *(Can't control)*
- Non streaming stream producers. *(Can redesign)*

What might users want when they pose a query?

- Source of a particular tuple.

Provenance

- Archivers built *on the fly*, tailored to answer a particular query with a time to live.
- In memory archivers to improve performance.
- Answers to complex queries when schema can change rapidly.

Querying Across VOs

- One common registry across all VOs?
  - Easier for the mediator.
- VO specific registries?
  - More scalable.
  - **Mediator needs access to each registry.**

Problems with Current Producers

- There exists static data and streaming data.

  Should static data have a measurement timestamp?
- Users want to connect existing databases.
- Existing names are confusing, e.g.
  - Database Producer: doesn't publish a database.
  - Archiver: doesn't archive data.
- Choice of producer types is bewildering:

  Do I pick a latest, history or stream producer?
- Difficult to mediate with current producers.

Proposed Solution

- Three Primary Producers:
  - Stream Producer
    - Insert() for streaming.
    - Snapshot table for latest queries.
    - Retention period for history queries.
  - Database Publisher **urgent!**
    - No insert method.
    - “Pull” data from database.
    - Publishes some or all of the database.
  - Canonical Producer (as is).