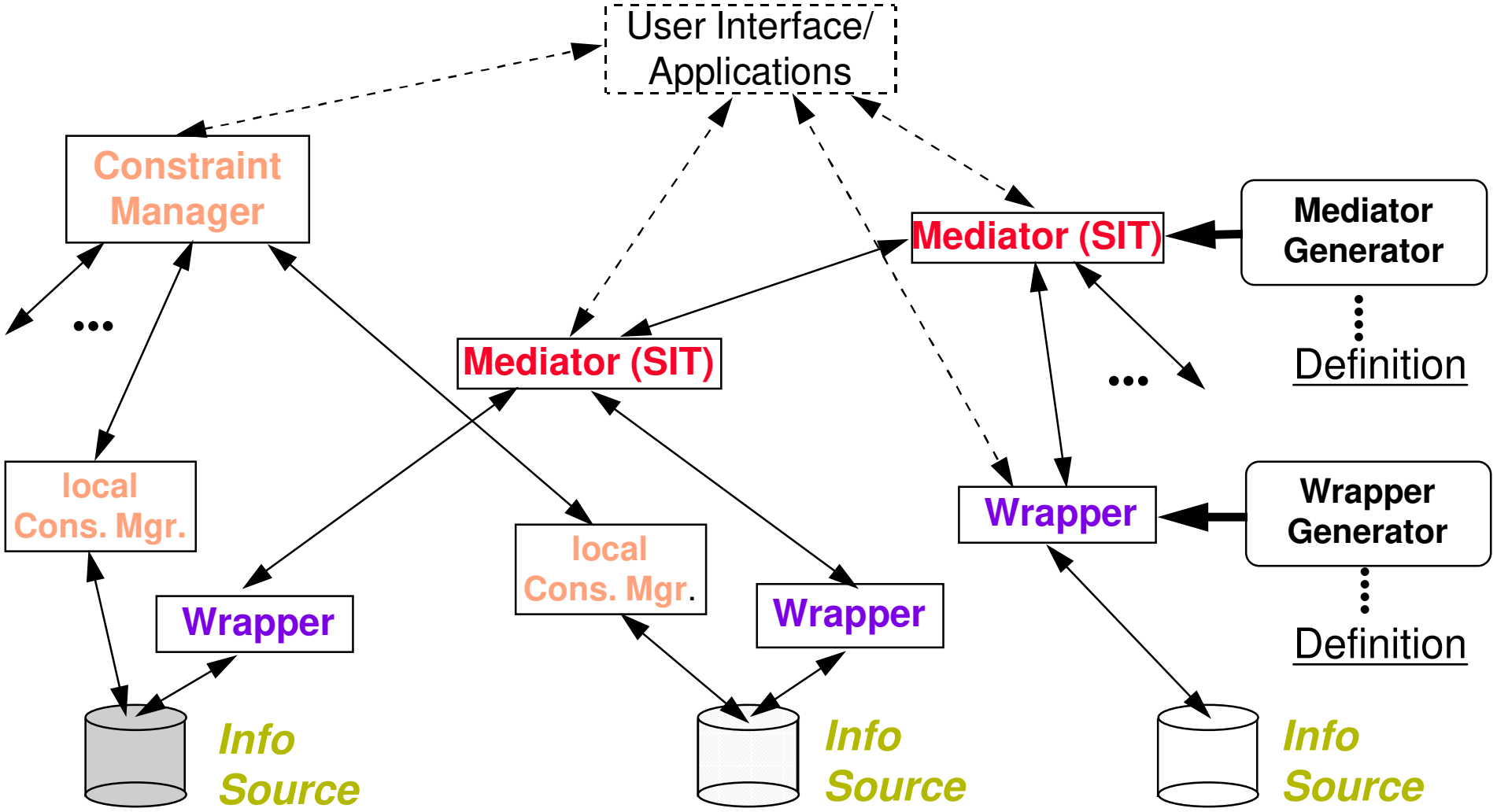
Stanford *TSIMMIS* Project



Goals/Payoff

- Object Exchange Model (OEM) for representing dynamic, unstructured data
- *Access* to *diverse sources* (structured and unstructured) such as database, legacy, file, and e-mail systems
- Powerful unified and configurable *user interface* for browsing sources locally or remotely (via internet)
- Rapid generation of *wrappers* and *mediators* (for semantic integration and transformation of information), allowing simpler integration of new sources
- New database system (LORE) for caching OEM objects in wrappers and mediators for improved efficiency

Technical Challenges

- Flexible, object-oriented data/knowledge model that is expressive yet simple
- High level description language for wrappers and mediators that is easy to use
- Generation of efficient wrappers and mediators based on user specification
- Intelligent browsing of unknown and often unstructured information
- Constraint management across heterogeneous sites

Schedule and Milestones

- *A: Sept. '94*: Object Exchange Model (OEM) and preliminary testbed
- *B: Sept. '95*: Toolkit for automatically generating translators, Mosaic-Based Information Explorer (MOBIE), OEM repository (LORE)
- *C: Feb. '95*: Wrapper generator toolkit, definition of mediator specification language
- *D: Sept. '96*: Mediator generator toolkit, constraint management toolkit, complete testbed demonstration

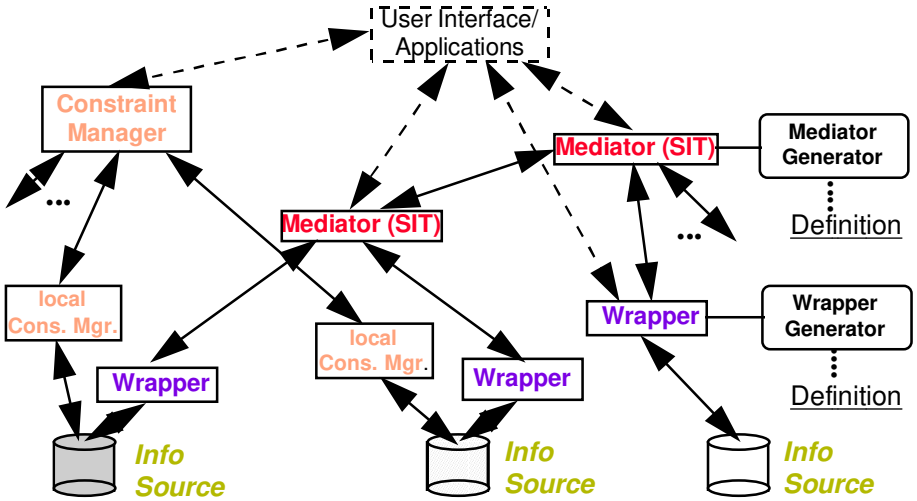
Background

- Traditional database systems (both relational and object oriented) focus on structured information and require a pre-defined schema
- Many sources, on the other hand, contain semi-structured information, or dynamic information whose schema is changing
- To cope with such information, many sources use a self-describing or tagged data model. Examples: Lotus Notes, tagged files, frames, e-mail, SGML, HTML
- TSIMMIS unifies these models and provides a sound query language. Furthermore, tools for integrating such sources with traditional databases are developed

Technical Approach

- Build upon existing and proven database technology (e.g., query languages, query optimization), adapting to a dynamic, self-describing information model
- Implement query *normalization* capabilities into wrappers, so they can accept queries the underlying sources cannot support directly
- Develop a powerful logic-based mediator specification language that can describe the necessary semantic transformations
- Allow users to explore sources through WWW browsers
- Demonstrate technology developed on real sources that include legacy, relational database, and file systems

Stanford *TSIMMIS* Project



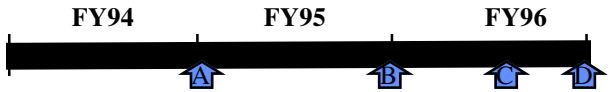
Goals/Payoff

- Object Exchange Model (OEM) for representing dynamic, unstructured data
- *Access* to *diverse sources* (structured and unstructured) such as database, legacy, file, and e-mail systems
- Powerful unified and configurable *user interface* for browsing sources locally or remotely (via internet)
- Rapid generation of *wrappers* and *mediators* (for semantic integration and transformation of information), allowing simpler integration of new sources
- New database system (LORE) for caching OEM objects in wrappers and mediators for improved efficiency

Technical Challenges

- Flexible, object-oriented data/knowledge model that is expressive yet simple
- High level description language for wrappers and mediators that is easy to use
- Generation of efficient wrappers and mediators based on user specification
- Intelligent browsing of unknown and often unstructured information
- Constraint management across heterogeneous sites

Schedule and Milestones



- *A: Sept. '94*: Object Exchange Model (OEM) and preliminary testbed
- *B: Sept. '95*: Toolkit for automatically generating translators, Mosaic-Based Information Explorer (MOBIE), OEM repository (LORE)
- *C: Feb. '95*: Wrapper generator toolkit, definition of mediator specification language
- *D: Sept. '96*: Mediator generator toolkit, constraint management toolkit, complete testbed demonstration

Recent Accomplishments

- **OEM**: Defined simple but powerful model for exchange of information among heterogeneous sites
- **LORE**: Designed and implemented a persistent store for OEM objects, including a powerful SQL-like language **LOREL**
- **Wrappers & Mediators:** Built a collection of wrappers and mediators for accessing heterogeneous bibliographic sources, including a legacy system (FOLIO), a relational database (SYBASE), and a file system
- **MOBIE**: Implemented a configurable system for browsing and exploring OEM objects through a standard World Wide Web interface like Mosaic or Netscape
- **Wrapper Generator:** Designed and implemented a preliminary translator generator system
- **Constraint Management Toolkit**: Designed and implemented toolkit for managing constraints over heterogeneous sites

Publications and Documentation

- "The TSIMMIS Project: Integration of Heterogeneous Information Sources." In *Proceedings 100th Anniversary Meeting, Information Processing Society Japan*, October 1994, Tokyo, Japan, pp. 7-18. URL=http://www-db.stanford.edu,/pub/chawathe/1994/tsimmis-overview.ps
- "Object Exchange Across Heterogeneous Information Sources." In *IEEE International Conference on Data Engineering*, March 1995, pp. 251-260. URL=http://www-db.stanford.edu/pub/papakonstantinou/1994/object-exchange-heterogeneous-is.ps
- "Integrating and Accessing Heterogeneous Information Sources in TSIMMIS." In *Proceedings of AAAI Spring Symposium on Information Gathering*, AAAI, March 1995, pp. 61-64. URL=http://www-db.stanford.edu/pub/garcia/1995/tsimmis-abstract-aaai.ps
- "The TSIMMIS Approach to Mediation: Data Models and Languages." In *NGITS-95 (Next Generation Information Technologies and Systems)*, Naharia, Israel, June 1995. URL=http://www-db.stanford.edu/pub/garcia/1995/tsimmis-models-languages.ps

Publications cont'd

- "Flexible Constraint Management in Loosely Coupled Distributed Databases." Technical Report, Stanford University. URL=http://www-db.stanford.edu/pub/chawathe/1994/cm.ps
- "Information Translation, Mediation, and Mosaic-Based Browsing in the TSIMMIS System." In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, San Jose, California, May 1995, page 483. URL=http://www-db.stanford.edu/pub/hammer/1994/mobie-demo-proposal.ps
- "Querying Semistructured Heterogeneous Information." Submitted for publication. URL=http://www-db.stanford.edu/pub/quass/1994/querying-full.ps
- "Medmaker: A Mediation System Based on Declarative Specifications." Submitted for publication. URL=http://www-db.stanford.edu/pub/papakonstantinou/1995/medmaker.ps
- "A Query Translation Scheme for Rapid Implementation of Wrappers." Submitted for publication. URL=http://www-db.stanford.edu/pub/papakonstantinou/1995/querytran.ps

Publications cont'd

- "Answering Queries Using Templates with Binding Patterns." In *Proceedings of the 14th ACM PODS*, San Jose, California, May 1995, pp. 105-112. URL=http://www-db.stanford.edu/pub/rajaraman/1994/limited-opsets.ps