

## Beyond Hypertext: A Web of Dynamic, Structured Data Sources"

The Web has been an incredibly effective means of disseminating information, primarily in the form of interlinked HTML documents. Unfortunately this Web is geared towards presenting information to users -- not towards facilitating data flow among the producers, curators, and integrators of data. In this talk I describe our ongoing efforts to design a global-scale "Web of structured data and data sources" where data can flow along links.

Our work revolves around the Orchestra system, which provides a means of linking data sources, such that data and updates may be translated and propagated along the links. Some sources may contain incompatible or inconsistent data -- not only due to variations in data freshness, but also due to different *\*beliefs\** among sources. Such conflicts must be managed by considering freshness and authoritativeness. Orchestra develops techniques for tracking data usage and derivation (provenance), and for using provenance and the authoritativeness of the sources to help manage consistency. Orchestra's query layer, Q, learns about authoritativeness, given user feedback about the relevance of query answers.

I will wrap up the talk by briefly describing our ongoing work to build the successor to Orchestra, Concerto, which *\*makes recommendations\** about data and sources likely to be of interest to the user, given the overall data usage patterns in the system.

This is joint work with Todd Green, Grigoris Karvounarakis, Nicholas Taylor, Partha Pratim Talukdar, Marie Jacob, Val Tannen, Fernando Pereira, and Sudipto Guha.