

## **Overcoming the Technical and Policy Constraints that Limit Large-Scale Data Integration**

The National Academies  
Lecture Room, 2100 C Street, NW  
Washington, DC

August 19-20, 2009 (Wednesday-Thursday)

### **AGENDA**

#### August 19, 2009 (Wednesday)

8:00 AM	Continental breakfast available	
8:30	Chair's Opening Remarks	Michael Stonebraker
<b><u>Data Integration Stretch Goals, Technical Needs, and Policy Issues: Views From Various Domains</u></b>		
9:00	Geospatial Data	Keith Clarke, UCSB
9:20	Life Sciences	Carl Kesselman, USC
9:40	Physics	Tim Frazier, LLNL
10:00	Astronomy	Alex Szalay, Johns Hopkins
10:20	Earth Sciences	Tom Karl, NOAA
10:40	Research Libraries	Clifford Lynch, CNI
11:00	Break	
11:20	Open Discussion	
Noon	<b><u>Working Lunch: Agency Perspectives</u></b>	
	About 10-15 minutes apiece from sponsors and other agencies	
	James St. Pierre, NIST	
	Michael Marron, NIH	
	Ed Seidel, NSF	
1:20	<b><u>State of the Art in Data Integration -- Structured Data</u></b>	
1:20	Data Federations	Laura Haas, IBM
1:40	Data type conversion and ETL technology	Lee Scheffler, IBM
2:00	Automatic conversion	Michael Siegel, MIT
2:20	The SciDB approach	David Maier, Portland State University

2:40	Linked, Open Data	Orri Erling, OpenLink
3:00	Microsoft Approach to Data Conversion	Phil Bernstein, Microsoft
3:20	Break	
3:40	<b><u>State of the Art Data Integration Solutions: Semi-structured Data</u></b>	
3:40	Google Approach	Alon Halevy, Google
4:00	Yahoo Approach	Raghu Ramakrishnan, Yahoo
4:20	Map/Reduce/Hive/Pig paradigm	Amr Awadallah, Cloudera
4:40	<b><u>Perspectives</u></b>	
4:40	Policy Perspective	Mike Nelson, Georgetown
5:00	Policy Perspective	Chris Greer, NITRD
5:20	Lessons from a Large-scale Information Integration Ecosystem	Michael Brodie, Verizon
5:40	Business Perspective	Josephine Cheng, IBM
6:00	Organization of Day Two: Charge 3 participants to present at start of Day Two with their takes on the problem	
6:30	Working Dinner	

August 20, 2009

8:30	<b><u>Open Brainstorming</u></b>
	Three speakers at 10 minutes each
9:00	Identify 3 areas for detailed and moderated discussion
9:15	<b><u>Topic Area One</u></b> (Moderator TBD)
10:00	Break
10:30	<b><u>Topic Area Two</u></b> (Moderator TBD)
11:15	<b><u>Topic Area Three</u></b> (Moderator TBD)
Noon	Adjourn Public Workshop
12:30	Planning Committee Lunch/Executive Session (Room 148)
	Outlining the Summary Report and Next Steps
3:00	Planning Committee Adjourns