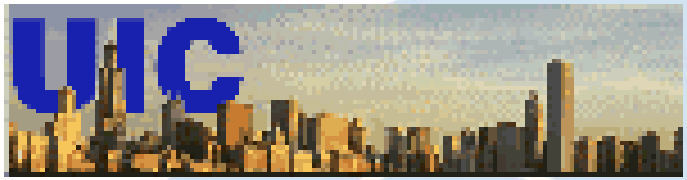


open data

Using DataSpace Archives to Support Long Term Stewardship of Remote and Distributed Data

Robert Grossman, Steve Eick, David Hanley, Xinwei Hong & Parthasarathy Krishnaswamy

www.ncdm.uic.edu

NASA/IEEE MSST 2004
12th NASA Goddard/21st IEEE Conference on
Mass Storage Systems & Technologies
The Inn and Conference Center
University of Maryland University College
Adelphi MD USA
April 13-16, 2004

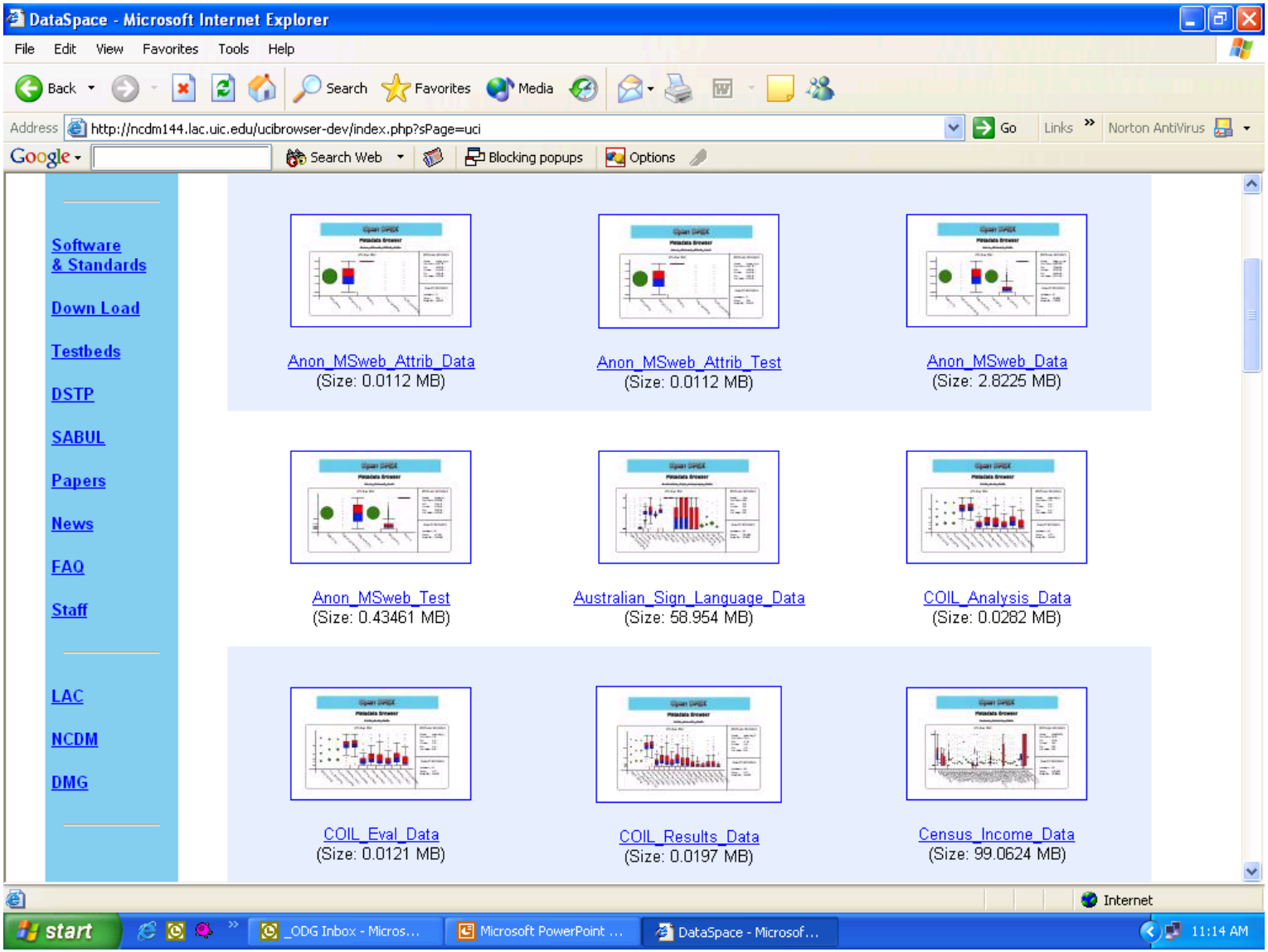
NASA

IEEE COMPUTER SOCIETY

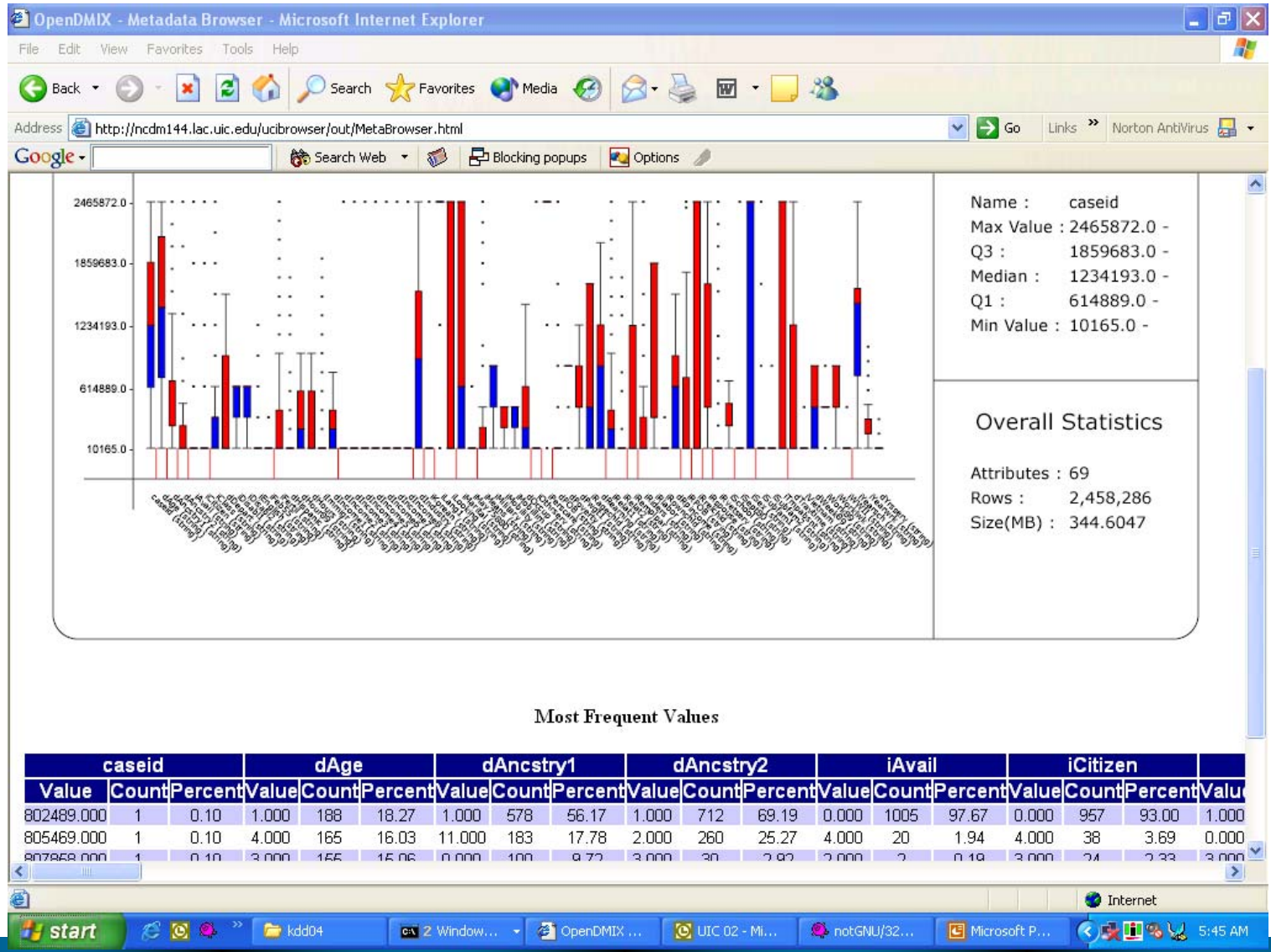
DataSpace Archives

- Use web services' WSDL and UDDI for *discovery* of data
- Use Data Set Icons and DataSpace services for *exploration* of remote data
- Use XML/SOAP for *transporting small data* and metadata
- Use specialized high performance web services (SOAP+) for *transporting large data*
- Use DataSpace distributed keys for *integrating* data
- Provide *long term storage* of data using Internet Backplane Protocol (IBP)

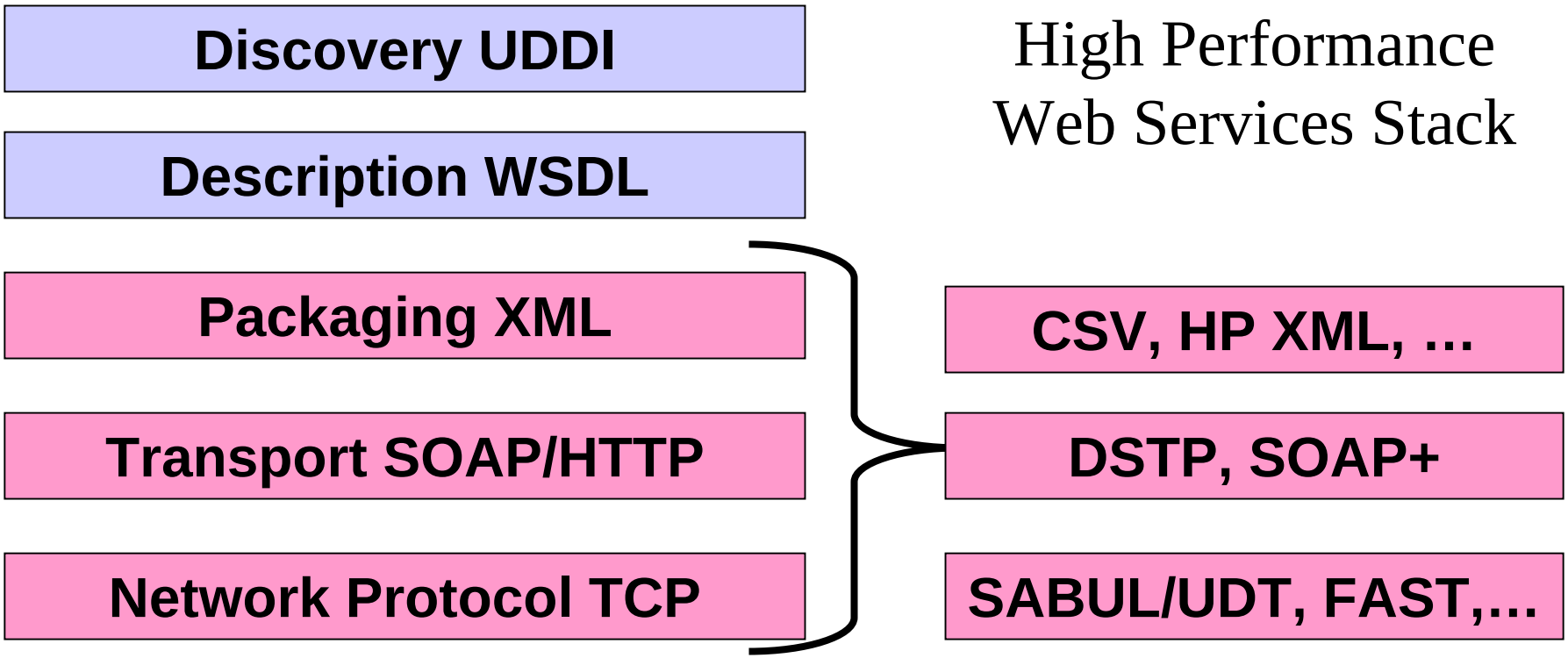
Example: UCI KDD Archives



Example: UCI KDD Census Data Set



High Performance Web Services (SOAP+)



Data Transfer using XML/SOAP vs. XML/SOAP+

Number Records	XML/SOAP (sec)	SOAP+ (sec)
10,000	0.65	0.21
50,000	177	0.72
150,000	673	125
375,000	3078	301
1,000,000	21121	823