

Privacy, Cost, and Availability Tradeoffs in Decentralized Online Social Networks

Amre Shakimov, Landon Cox

Duke University

Alexander Varshavsky, Ramón Cáceres

AT&T Labs

17 August 2009

Online Social Networks (OSNs) are enormously popular



Allow people to share information such as personal profiles, photos, and messages with members of their social networks

Privacy issues with current OSNs

- Services are centralized
 - Hold data for millions of users in a single administrative domain
 - Vulnerable to large-scale privacy breaches
- Terms of service often grant providers rights to user data
 - Provider may display and distribute data in any way it sees fit
 - Advertising-driven business models create incentives to share data with third parties in ways that may diminish user privacy
- Prominent privacy violations have already been seen
- Public awareness of privacy issues is growing

Other issues with centralized OSNs

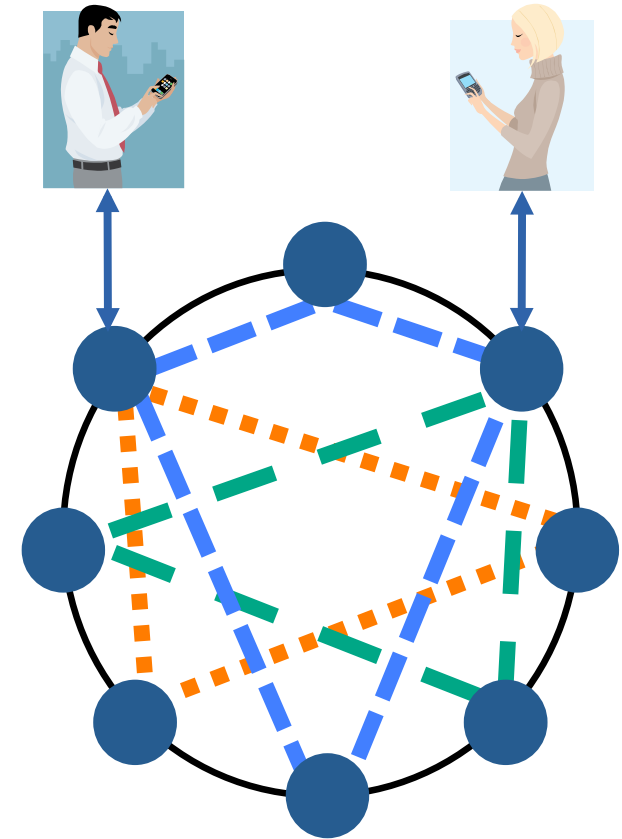
- May shut down and leave users without access to their data
- May not provide users their data in an open format
- May not be extensible in ways that users want
- May not scale

“Are all of these circumstances and conversations going to push the social web over the edge, toward a more distributed and less centralized model?”

New York Times, 11 August 2009

Vis-à-Vis distributed OSN framework

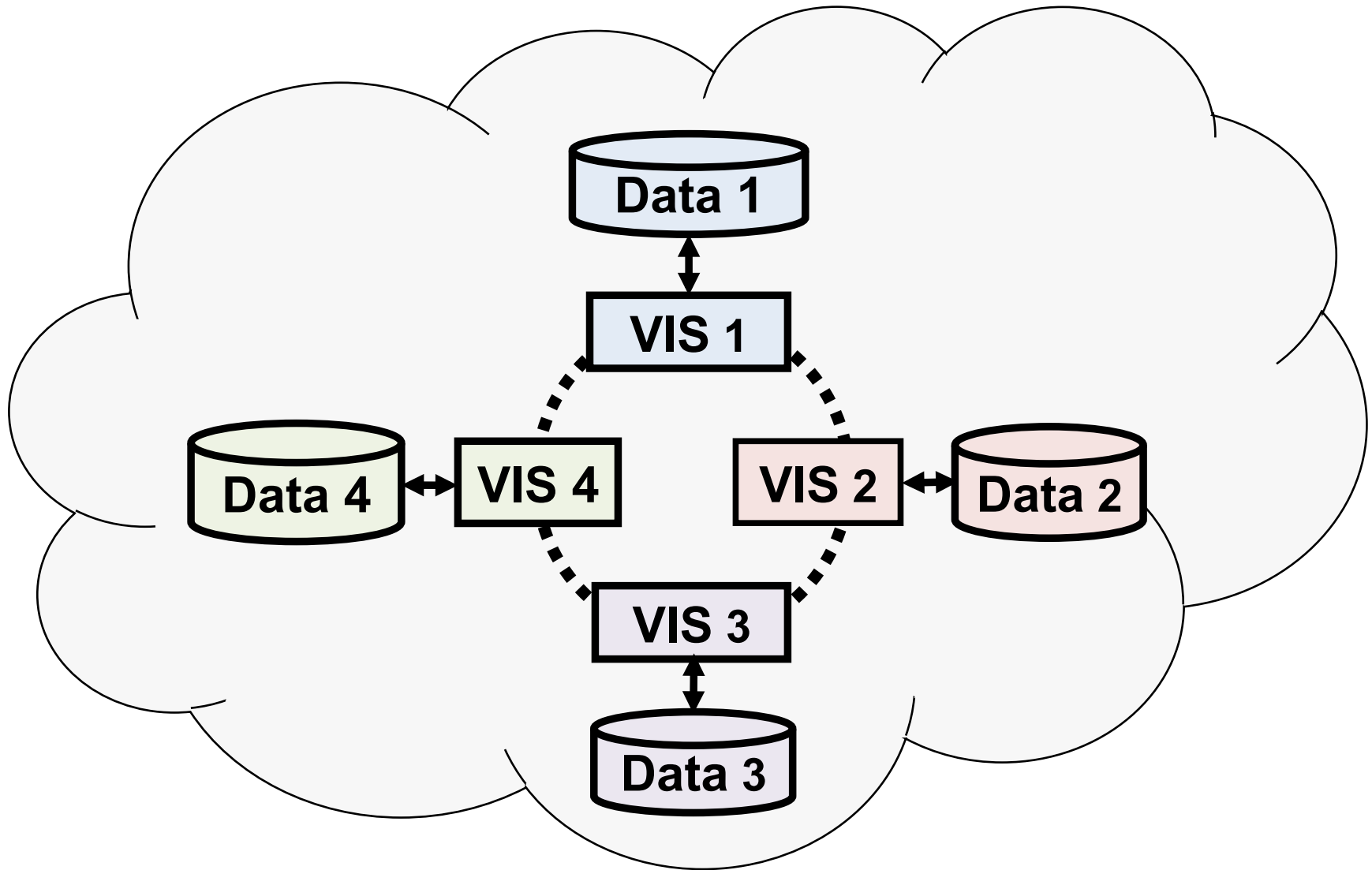
- Each person keeps his private data on his own Virtual Individual Server (VIS)
- VISs can be virtual machines
- Private data is distributed across many administrative domains
- VIS owners maintain rights to data
- VISs self-organize into decentralized overlay networks, one overlay per social group with which VIS owners wish to share data
- Mimics privacy expectations and trust relationships of offline social networks



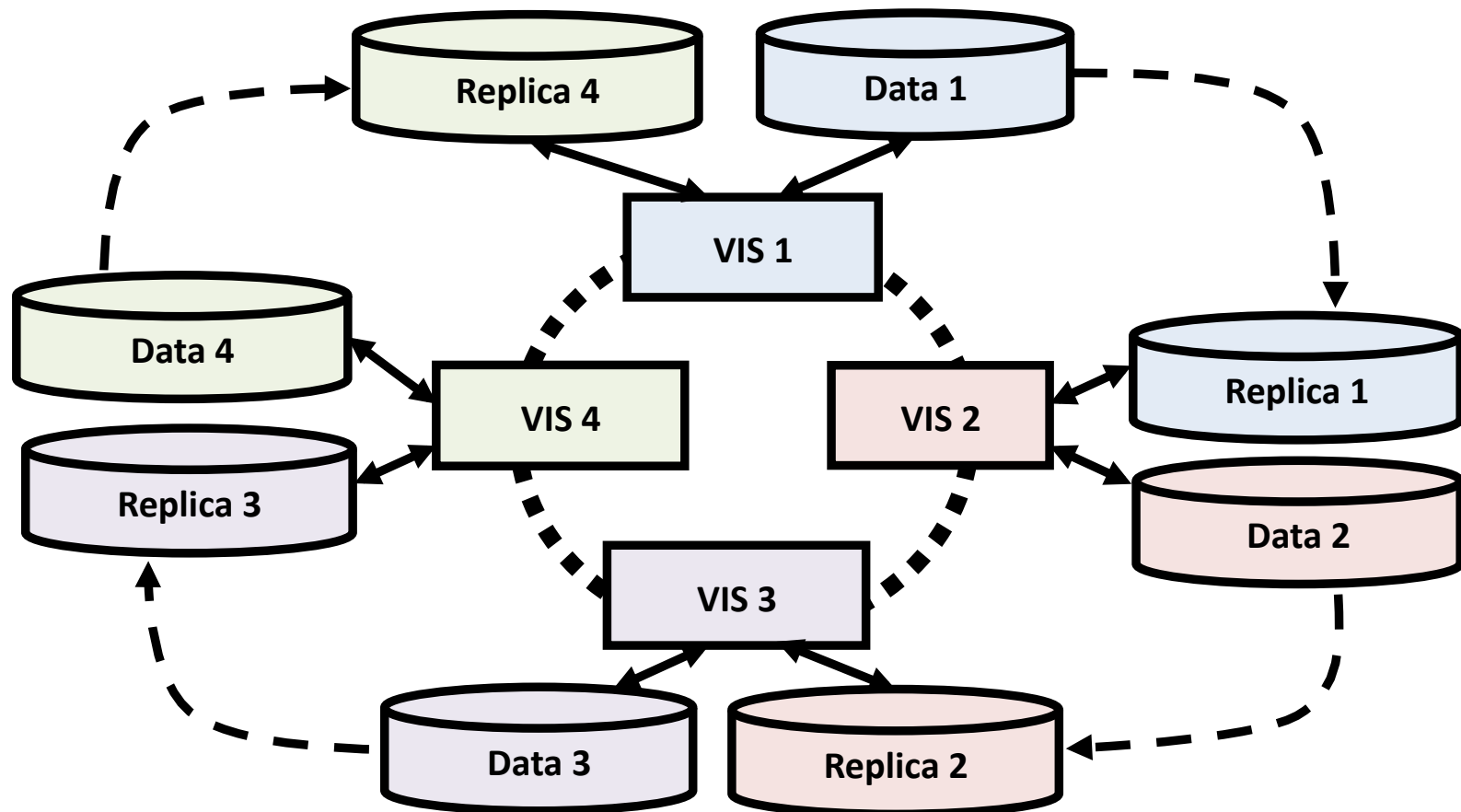
Three distributed OSN approaches

- Cloud-based decentralization
- Desktop-based decentralization
with socially informed replication
- Hybrid of cloud- and desktop-based

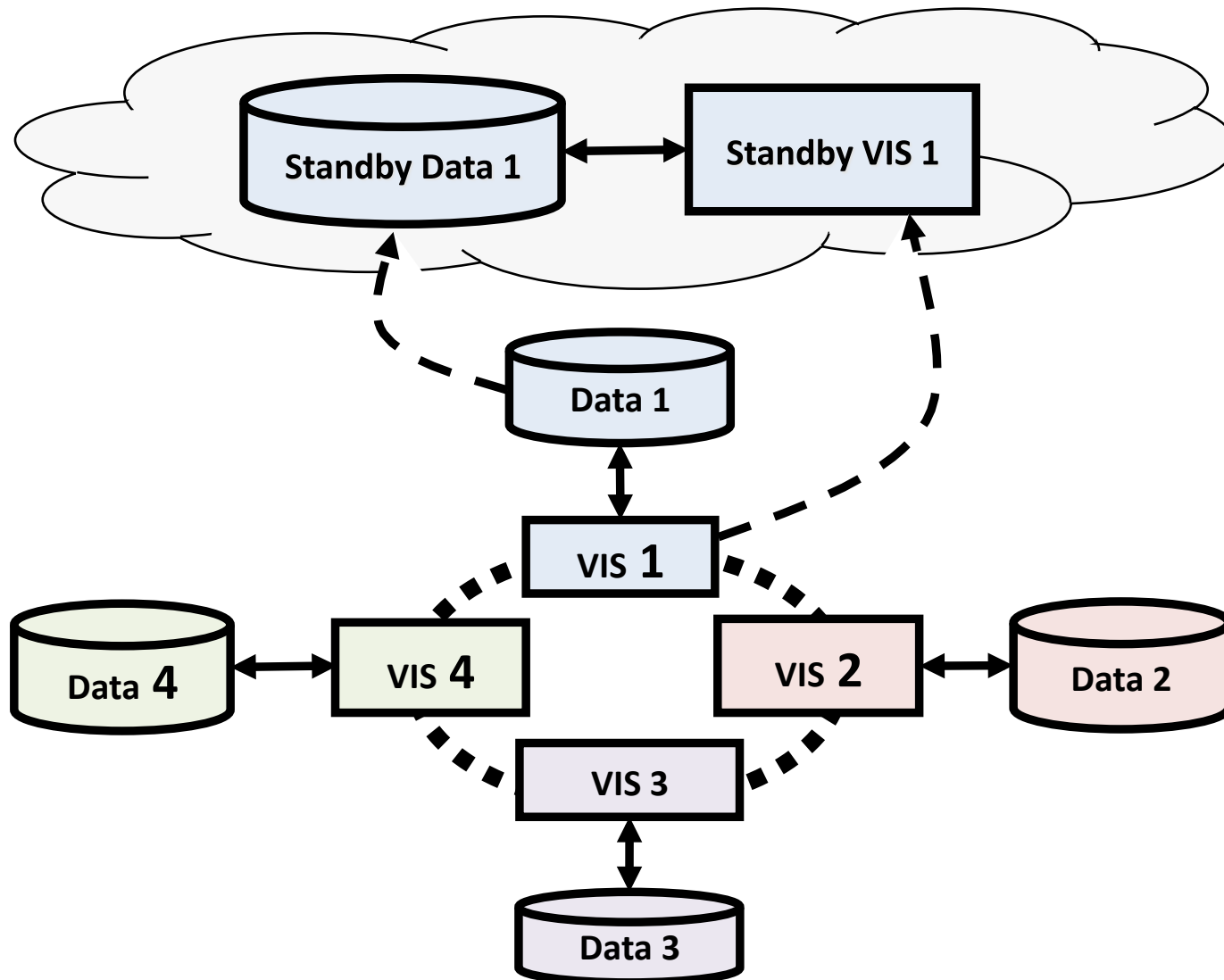
Cloud-based decentralization



Desktop-based decentralization with socially informed replication



Hybrid decentralization



Vis-à-Vis architecture

- Two tiers of distributed hash tables (DHTs)
 - Top tier contains one DHT for the Meta Group
 - Bottom tier contains one DHT per social group
- Supports a wide variety of groups
 - Open or restricted, public or secret, ...
- Enables scalable operations
 - Create, join, leave, insert, query, ...
- Framework for many popular OSN features
 - Suggest friends, plug in third-party apps, ...

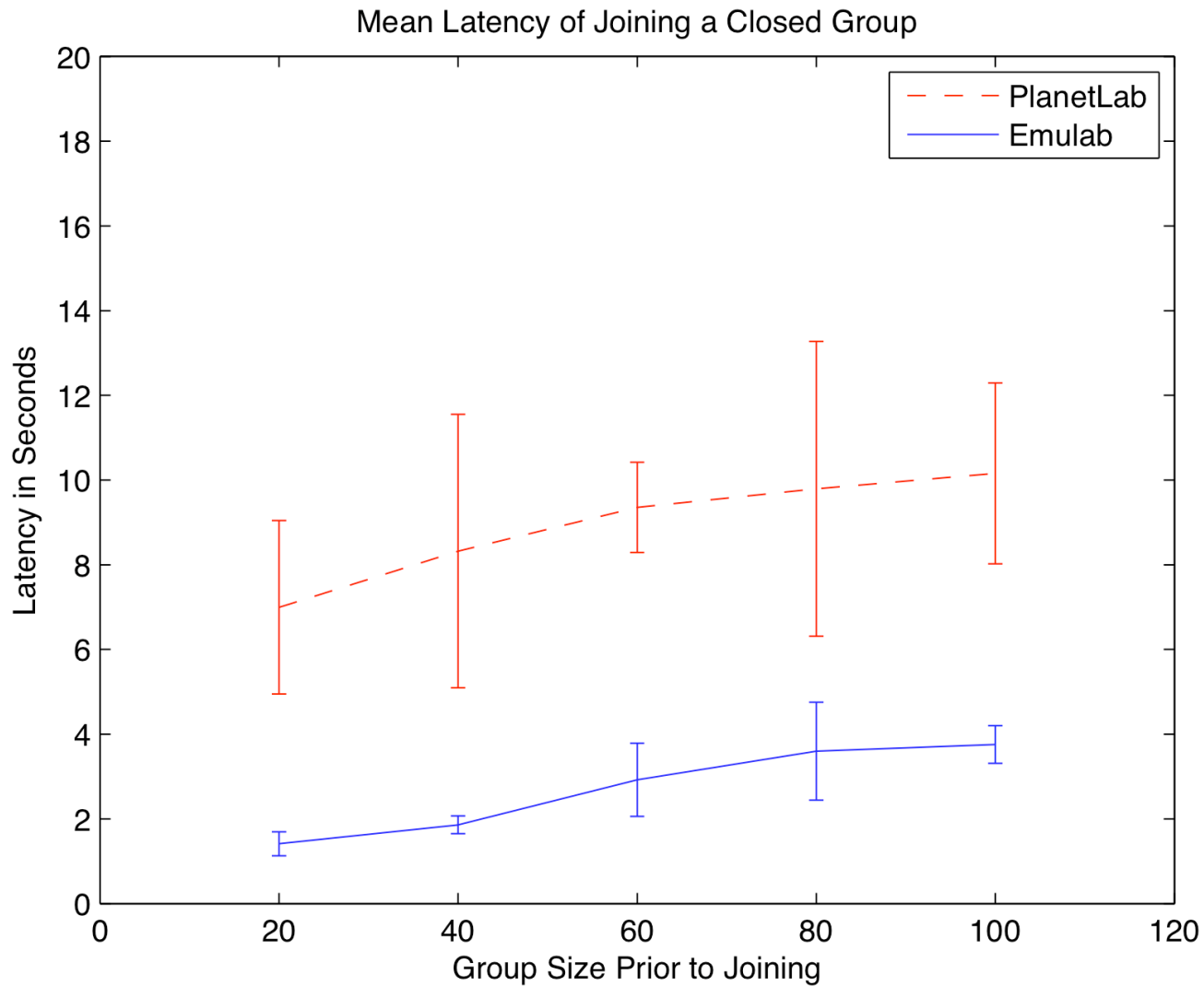
Vis-à-Vis prototype

- OSN software
 - Pastry for basic DHT functionality
 - Scribe for multicast over DHTs
 - Additions to support groups and Meta Group
- Base virtual machine software
 - Full server software stack: Linux, Apache, MySQL, ...
- VISs deployed at Amazon EC2, Emulab, PlanetLab, Duke University, and AT&T Labs

Locations of 120 VISs on PlanetLab



Latency to join a restricted group



Many uses of a VIS besides OSN

- Trusted resource-rich proxy for mobile devices
 - Saves battery, bandwidth, storage, processing on devices
 - Many possible applications
 - Participatory sensing
 - Synchronization and backup
 - Location-based services
 - Other applications...
- Trusted online presence for VIS owners
 - Web server
 - Email server
 - Other services...
- Helps preserve owner privacy across all these uses
- Amortizes its cost across all these uses

Related Work

- Distributed OSNs
 - PeerSoN [Buchegger & Datta, 2009]
 - P2P OSN [Cutillo, Molva & Strufe, 2009]
 - Ego [Amick & Ypodimatopoulos, 2009]
 - DiSo (open-source project)
- Hide information from centralized OSNs
 - NOYB [Guha, Tang & Francis, 2008]
 - flyByNight [Lucas & Borisov, 2008]

Summary

- Current OSNs suffer from privacy and other problems
- Vis-à-Vis is a decentralized OSN framework based on VISs
 - Distributes data across many administrative domains
 - Gives people ownership and control over personal data
- Three approaches to distributed OSNs based on VISs
 - Cloud-based: high availability, high cost
 - Desktop-based with socially informed replication: low cost, complex replication policies to achieve high availability
 - Hybrid: high availability, low cost, moderate complexity?
- Important to explore alternatives as public awareness of privacy issues grows and cost of computing drops