

Distributed Personal Data Management

Ken Rimey

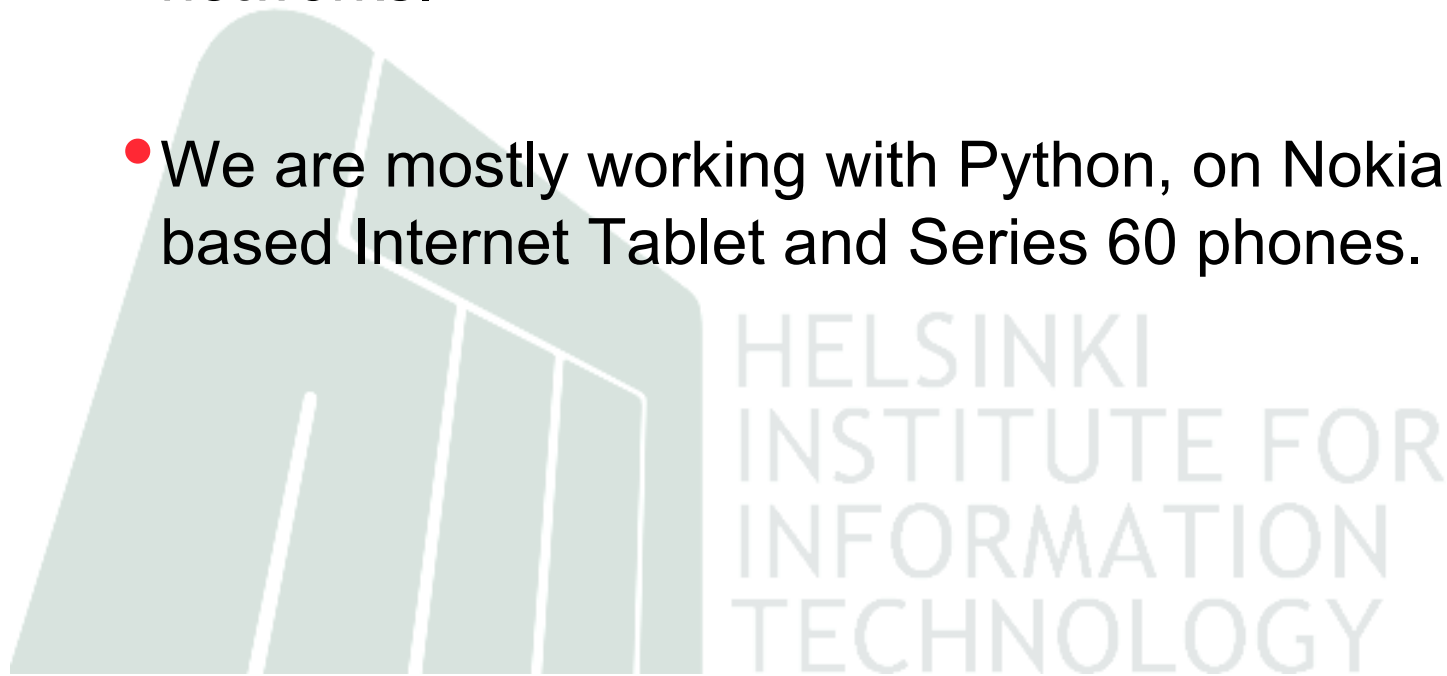
- Small software technology research group, ~5 people.
- Pragmatic middleware development driven by rough engineering prototypes of end-user applications.
- Projects:
 - Services for All
 - ePerSpace II
 - PDIS (2003-2004)

Services for All (S4ALL)

- ITEA project aiming “to enable user-centric services that are easy to create, share and use.”
 - End user can create compositions (orchestrations) of web services using the S4ALL Composer.
 - He can publish these or execute them directly on a mobile device.
 - Reference to component web services is via semantic descriptions (OWL-S).
- Partners: Consortium led by Alcatel. Other Finnish partners are Nokia and Capricode.
- Duration: 2005-1H2007
- Volume at HIIT: 4 researchers

HIIT Focus in S4ALL

- First prototype of a handheld Composer (linear compositions, explicit URLs, execution on phone).
- Tools for exposing native phone APIs.
- Prototyping of new kinds of services for local wireless networks.
- We are mostly working with Python, on Nokia's Linux-based Internet Tablet and Series 60 phones.



ePerSpace II

- Proposal planned for FP6 IST call on “Networked Audio Visual Systems and Home Platforms” (20.12.2005).
- ePerSpace = “Towards the era of personal services at home and everywhere.”
- Partners: Consortium led by France Telecom.
- Duration: 2006-2007
- Tentative volume at HIIT: 2 researchers



HIIT Focus in ePerSpace II

- Distributed data management for the digital home:
 - How will consumers manage their files on multiple handheld devices, personal computers, and home/car entertainment boxes?
 - Build on the update-anywhere replicated XML database technology developed in the PDIS project.
- Objective: To develop a conceptual model for end-user management of replicated data on diverse networked devices that is powerful, easy to learn, and practical to implement.

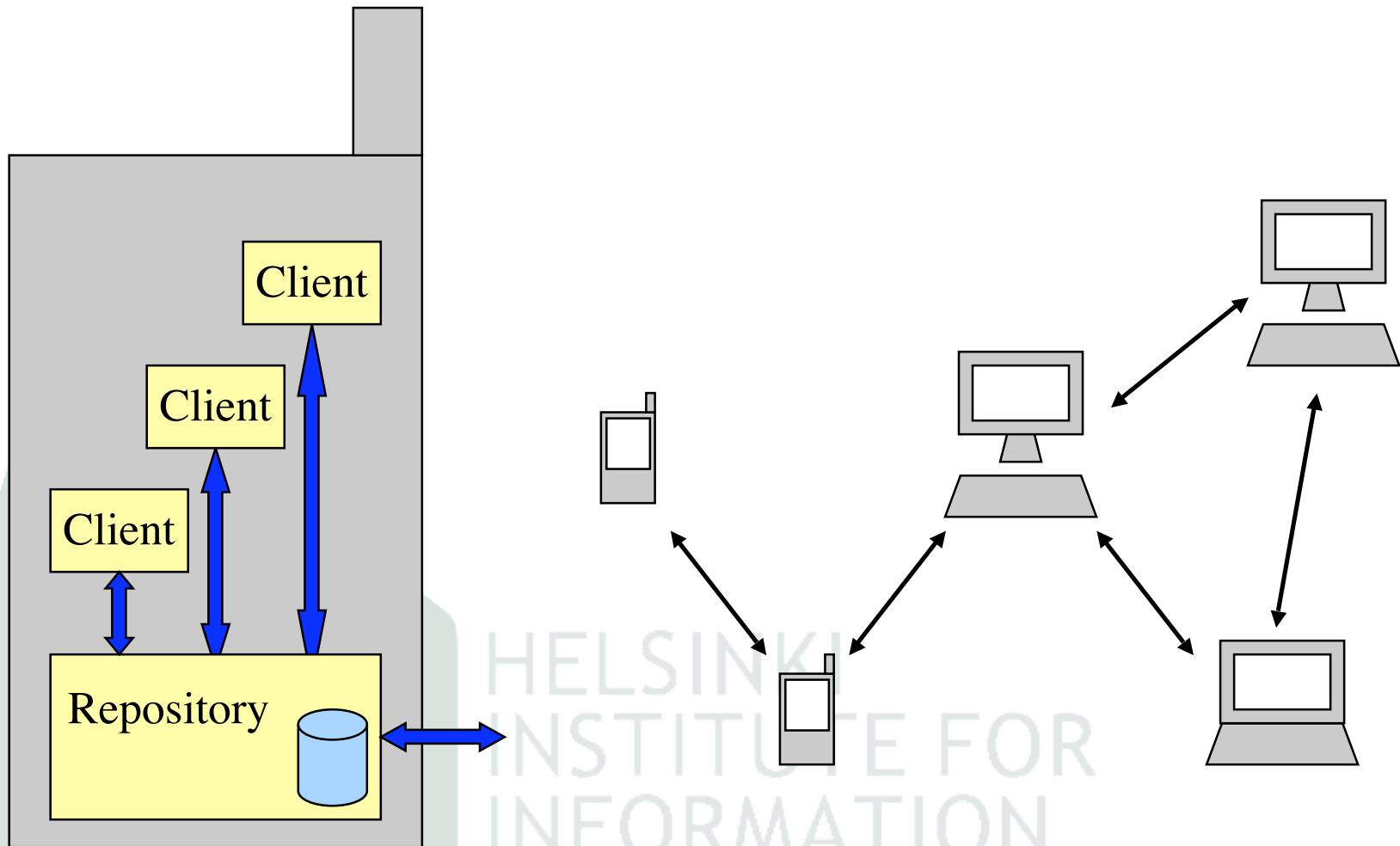
Motivation

- Home life increasingly involves digital media files, shared personal databases (Family Information Management), and occasionally office-type documents.
 - Home and personal electronics are acquiring huge storage capacities.
 - Local wireless networking is increasingly viable between devices, and not just for Internet access.
 - Replicating data on several devices ensures its availability and protects it from loss.
 - Storage is a simple generic function that a wide variety of devices can support in a compatible way.
- Managing one hard drive is hard enough. Managing a network of them is even harder!

Personal Distributed Information Store (PDIS)

- Project sponsored by Tekes and several companies (including Nokia) during 2003-2004.
- Next-generation data synchronization: Enable people to store copies of their data on several devices, and several computers, and to keep these copies in sync.
 - Share PIM data between devices and people.
 - Store digital media metadata for managing media files across devices.
- State-transfer, update-anywhere replicated XML database: Epidemic change propagation allows any device to be synchronized with any other at any time.

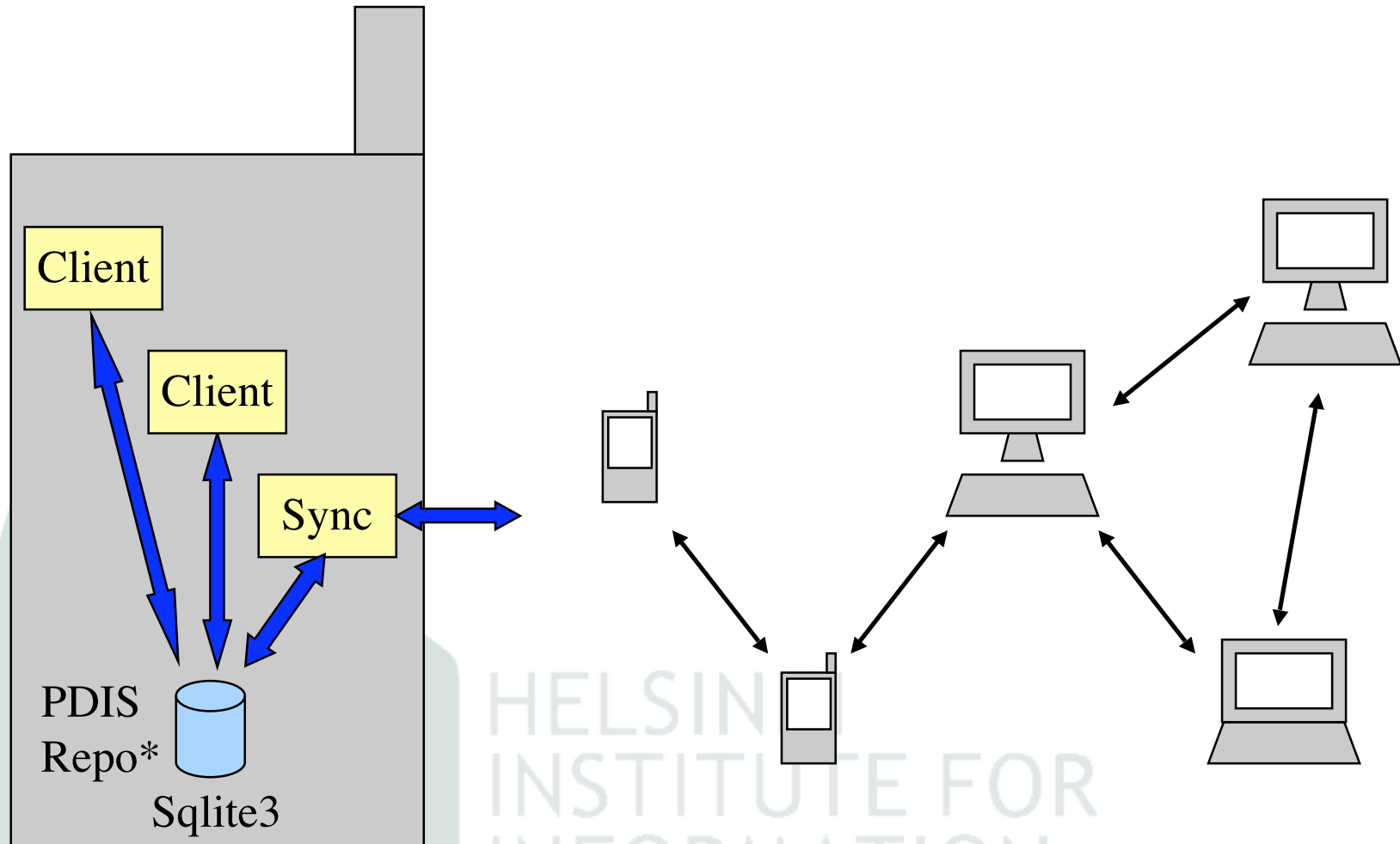
PDIS Architecture



Key Ideas in PDIS

- Utopian viewpoint: All repositories must maintain PDIS versioning information. PDIS is not a sync protocol for legacy databases.
- A PDIS repository stores large collections of small, well-formed XML documents.
- *Live queries* combine lazy evaluation of XPath queries with change notification.
- Records/documents (contact, calendar entry, etc.) are atomic for purposes of synchronization.
- Continuously propagate new versions instead of requiring repeated point-to-point syncs.
- PDIS defines a versioning scheme, so synchronization is just copying of versions between two repositories.
- Conflict resolution is independent of synchronization.

Future PDIS Architecture



*Base data model will be multivalued dictionaries.
XML support will be handled by client libraries.