

Crosscloud

Let Users Control Their Data
... and We All Benefit

Sandro Hawke (sandro@hawke.org, [sandhawke](#))
MIT Decentralized Information Group
19 November 2013

About this Talk

- This talk is about *planned* research and development
 - ◆ I'll describe the problem space
 - ◆ and our research approach
 - ◆ but: **no results yet, just ideas**
- (Oddly) I'm not a computer scientist
 - ◆ I'm a software developer
 - ◆ Lots of work in standards at W3C
 - ◆ Esp. Semantic Web standards
 - ◆ Tim Berners-Lee is the PI for this work
- We're look for an RA starting Jun 2014

The Basic Idea

→ Decentralized Software

- ◆ The functionality people are used to
- ◆ The interfaces they're used to
- ◆ BUT no central authority running it all

→ We see it in email, the web, irc

- ◆ You don't need permission to set up a website
- ◆ Not entirely black and white (cf SOPA)
- ◆ nearly everything else is centralized

→ Big Questions

- ◆ What value does decentralization provide?
- ◆ How can it possibly work?
- ◆ Who will benefit?

Decentralization

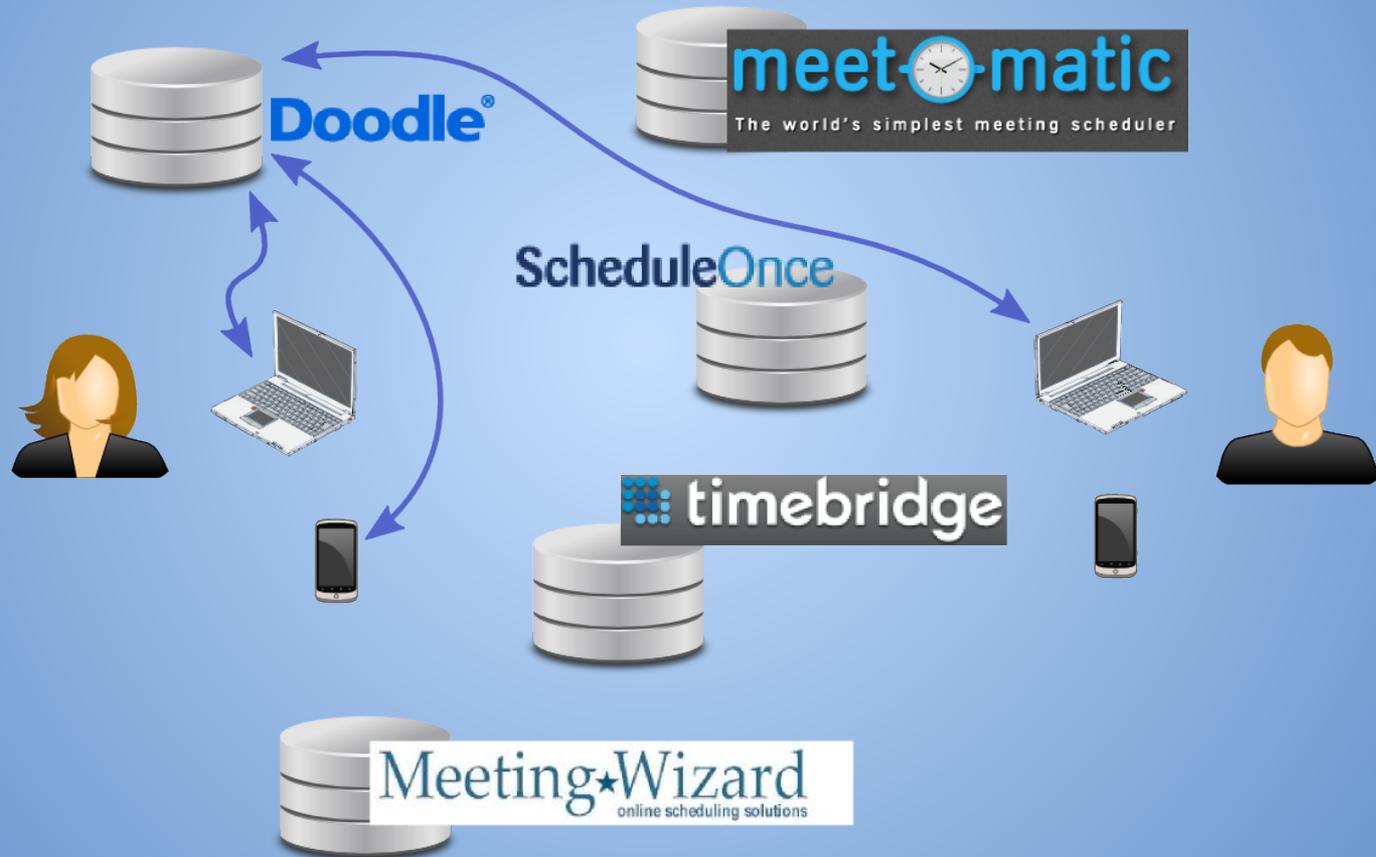
- Gives users more choice
- Lowers market barriers to entry
 - ◆ good for competition
 - ◆ bad for market leaders
- Fosters innovation

Social applications are natural monopolies because of network effects.

Crosscloud will end that.

Scheduling a Meeting

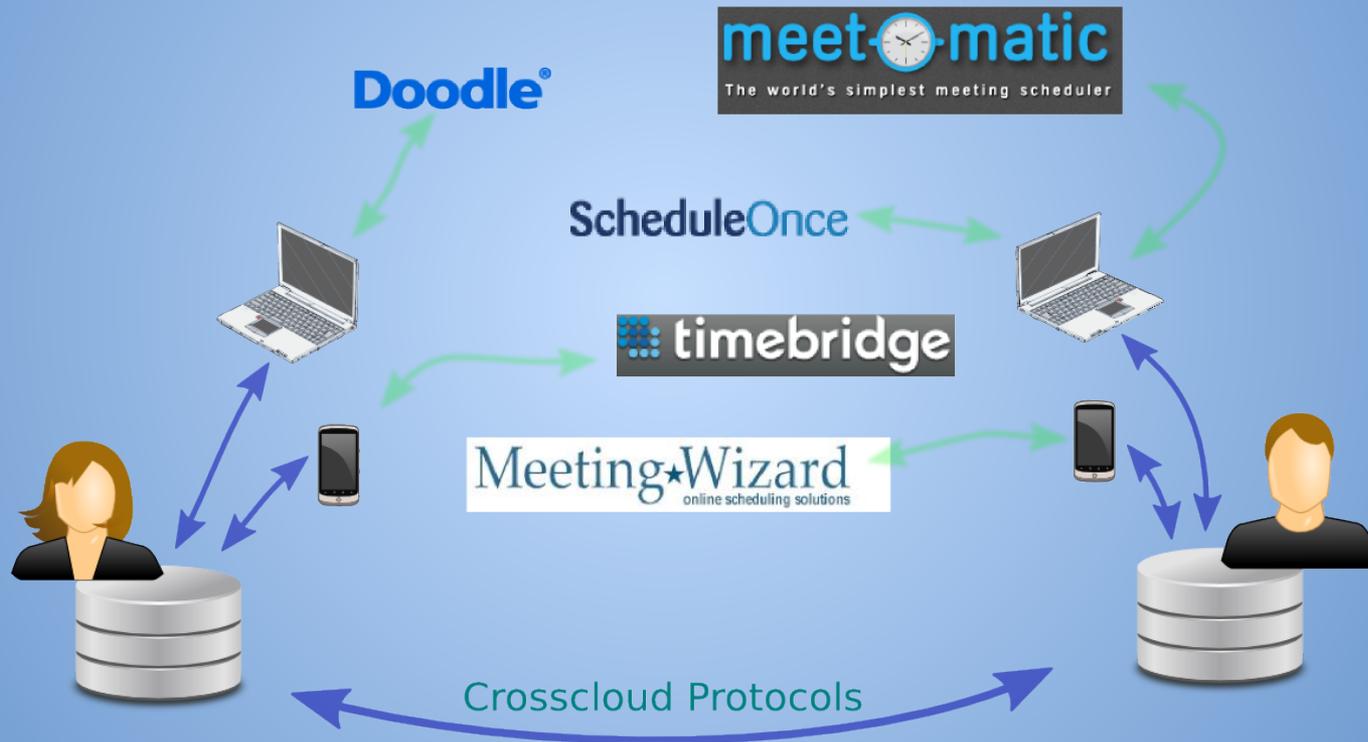
Scheduling a meeting today (without Crosscloud)



Everyone invited has to use the same vendor!

With Crosscloud

Scheduling a meeting -- With Crosscloud



You remain in control of your data.
You are free to pick different software and service providers.

Crosscloud

- Software stores the data under user control
- Protocols make it all look like one big db
- A bit like:
 - ◆ everyone has a machine-readable blog
 - ◆ your scheduling choices appear on yours
 - ◆ ... with access control
 - ◆ and client software scoops it all up

Overview

1. What's wrong with today's systems
2. What would we like to see?
3. How this might work (technically)
4. How this might work (commercially)
5. The Project
6. Q&A

“Social Software”

→ We’re using the term broadly

Computer systems which provide additional value by involving additional users.

- ◆ multiuser systems
- ◆ groupware
- ◆ social apps
- ◆ “social machines” (sociam project)

Some Social Software

imdb (1990), wikipedia (2001)

ebay (1995), craigslist (1995)

slashdot (1997), reddit (2005)

epinions (1999), yelp (2004)

flickr (2004), youtube (2005), vimeo (2004)

okcupid (2004), match (1995), eharmony (2000)

linkedin (2003), facebook (2004)

twitter (2006), instagram (2010), vine (2013)

blogger (1999), tumblr (2007), google+, pinterest

aim (1997), ym (1998), gtalk (2005) [irc (1988)]

meetup (2002), evite (1998)

second life (2003), world of warcraft (2004)

google pagerank (1998), google docs

farmville (2009), words with friends (2009)

sourceforge (1999), github (2008)

What's Wrong With Social Software Today?

- Not used by the people you want
- Changes unexpectedly
- Users have minimal influence/control
- Not available to every Internet user
 - ◆ so it can't be used by governments
 - ◆ or social organizations (eg churches)
- Segmented markets (even when widely used)
- Unwanted ads
- Privacy issues
- Reliability issues (for some applications)
- High barrier to entry (Network Effect)
stifles innovation

Rephrased as a checklist

- is it used by the people you want?
- is the UI consistent enough for you?
- do you have enough control over the UI?
- is it available for all internet users?
- does it have 100% of the market?
- are the ads, if any, okay for you?
- does it sufficiently protect your privacy?
- is it reliable enough for you?
- does it evolve/improve as well as it would with open competition?

What About New Systems?

- Health records
 - ◆ shared with patient, caregivers, research, with appropriate controls
- Experimental data
 - ◆ along with complete provenance for reported results
- Meeting/Calendar Sharing
 - ◆ Why can't my dentist and I schedule electronically?
Why can't W3C groups?
- Government Services
 - ◆ budget data, legislation, EPA
 - ◆ disaster relief

Why Haven't These Appeared?

- Insufficient revenue stream
- Insufficient willingness to cede control to one vendor
- Small solutions fail to reach critical mass

What Developers Need

- Simpler development process
 - ◆ Even with great cloud hosting, scalable backend operations are a challenge
- Responsibility for code, not operations
 - ◆ Small developers don't want to be running ops, maintaining user data 24x7 for years
- A practical way to respect the rights of the users
- A way to share users, forming an *aggregate* critical mass

Overview

1. What's wrong with today's systems
2. What would we like to see?

<< we are here >>

3. How this might work (technically)
4. How this might work (business)
5. The Project
6. Q&A

Part 3: Technology

(How Might This Work?)

- Everything an app stores, it stores as Linked Data on the user's site (in a *pod*)
- Good access control
- A backlink indexing system
- A way to subscribe to updates
- Data conversion between competing vocabs
- Social resolution for conflicting information

Example:

Endorse This Talk



- Imagine there's a database of talks
- This talk is in it, described in Linked Data
- You have an app for browsing talks
- You can endorse a talk
 - ◆ “like”, “thumbs up”, “+1”, give it a star
- You can see who else has endorsed it
- You can see what someone has endorsed

The Talk, in LD

A simple RDF database at <http://example.org/talks> might include:

```
<http://example.org/talks/talk55>  
  cal:begins "2013-11-19T06:00Z"^^xs:dateTime;  
  cal:location <http://id.mit.edu/rm32-155>;  
  dc:creator <http://www.w3.org/People/Sandro/data#Sandro_Hawke>.
```

Your Endorsment

- Your TalksBrowser app shows you that data
- When you click “endorse”, it POSTS this triple:

`<http://example.org/talks/talk55>`

`eg:endorsedBy <http://alice.userpods.com/me>.`

to `http://talksbrowser.alice.userpods.com`

Wave Hands

That's as far as the “classical” Semantic Web vision goes.

- There's data about the talk on one site
- There's an endorsement on another
- “somehow” they come together and it all works (handwave, handwave)
- In the Human Web, we could use a search engine to maybe find endorsements.
(pagerank!)

Where to store that endorsement?

At example.org/talks	At talksbrowser.alice.userpods.com
Very easy to find.	
	Alice can retract it
	Alice can keep it private
	Alice's followers can see it
	3rd party apps can use it (competition! innovation! control!)
	Alice's identity is connected to endorsement

How to make it findable?

3rd party indexing of the backlinks (“data registration”)

```
<http://example.org/talks/talk55>  
  cal:begins "2013-11-19T06:00Z"^^xs:datetimestamp;  
  cal:location <http://id.mit.edu/rm32-155>;  
  dc:creator <http://www.w3.org/People/Sandro/data#Sandro_Hawke>;  
  datareg:index <http://rdfindex.com>.
```

Registration

→ When Alice's pod server receives that triple...

`<http://example.org/talks/talk55>`

`eg:endorsedBy <http://alice.userpods.com/me>.`

... it dereferences those three URLs, looks for `datareg:index` declarations, and then

→ It notifies those index servers!

Querying

- Now, when Bob wants to know who has endorsed the talk, his app:
 - ◆ fetches the data about the talk
 - ◆ notices the `datareg:index` declaration
 - ◆ asks the given index about other data
 - ◆ queries those data sources

Thus, all the (public, registered) endorsements can be found

- Do this efficiently and securely at scale will be a challenge. (cf “Research”)

Real Time?

- Bob would like to see the star appear, as soon as Alice clicks 'endorse'
- Doable, just needs a notification protocol
 - ◆ something like PubSubHubbub

(some Research, and Standards Work here)

Multiple Vocabularies?

- Alice used eg:endorsedBy
- Bob is looking for fb:likes

- Someone needs to publish a vocabulary mapping.
 - ◆ owl:sameAs, other OWL
 - ◆ RIF?
 - ◆ N3?
 - ◆ JavaScript?

- Then register that mapping
(some Research here, too.)

Conflicting Information (Edits)

- Alice posts that snacks will be served
- Bob posts the snacks will not be served
- What does Charlie see?
- Maybe it depends on:
 - ◆ which came first
 - ◆ if Alice says she's correcting Bob's post
 - ◆ Charlie's relationships to Alice and Bob

(Research!)

Technical Summary

- All data in user pods
- With access control
- index to backlinks
- change notification
- vocabulary mapping
- allowing edits, social resolution for inconsistencies

Part 4: Business (Driving Adoption)

- Growth Strategy
- Developers
- Users
- Pod Providers

How To Grow

→ It's a platform, multi-party ecosystem

Android (JUST AN ANALOGY)	Crosscloud
Google Android Division	Crosscloud Project Team
Users of Mobile Devices	Users of Social Software
Android App Developers	Crosscloud App Developers
Hardware Manufacturers	Pod Providers

- Start narrow, with “customers” who love you
- ◆ For a platform, “customers” includes the whole ecosystem

Likely Developers

- Lone coders, seeking impact
 - ◆ No need for ops team
- Governments, non-profits
 - ◆ Universal reach
 - ◆ Not ceding control to a private corp
- Small Projects within Enterprises
 - ◆ Steers around red tape (no “service”)
- Highly Competitive Folks
 - ◆ Looking to take on Youtube, Facebook, etc
- Ethical Developers
 - ◆ People who want to respect their users

Unlikely Developers

→ Folks looking for a big exit from owning their market

Crosscloud apps don't lock in the users, and don't own the network effect.

More of a conventional business, where you have to keep the customer happy.

So: crowdfunding instead of VCs?

More Unlikely Developers

- Folks building their business on user data
 - ◆ hugely popular approach right now

If you have a legitimate use for it, ask the user for permission.

(They'll tell their pod server to let you see it)

Likely Users

→ People who love the idea

- ◆ Internet Activists
- ◆ Hobbyist Developers

→ People who actually benefit

- ◆ Folks working in a diverse social context
(where you can't get everyone to use the same tools)
 - W3C Participants
 - Relief Workers

Pod Providers

- A bit like email providers
 - ◆ Might be a server in your basement
 - ◆ Might be like gmail
(your choice)
- We'll do this for now & provide software
- Might be startups
- Existing cloud service providers
 - ◆ Dropbox, AWS, Facebook, Google
- Folks with vast customer base
 - ◆ Verizon, Mastercard, Experian, Post Office
- Folks with strong user connection
 - ◆ Banks, Local Gov't, Schools, Employers

Part 5: The Research Project

Structure

Hypotheses

Next Steps

Project Structure

- Contemplating since 2009
- Fundraising since early 2012
- Funded, July 2013
 - ◆ Knight Foundation “Prototype” grant
 - ◆ NSF EAGER grant
 - ◆ QCRI CSAIL grant, plus QCRI personnel (3-5 people, ~2 years, ramping up)

- PI, postdoc, RA, coders + QCRI

- Evolving Connection to W3C

Broad Technical Hypothesis

- Can one build full-featured decentralized apps, using RDF linked data?
- ◆ performance
 - ◆ scaling within vocab
 - ◆ scaling across vocabs
 - ◆ security

plan: build demo apps & test them

Results may be mixed: maybe only some kinds of apps will be practical.

Developer Adoption Hypotheses

- Some people will love developing for Crosscloud

Plan: try to recruit developers

- Eventually: for every cool non-Crosscloud app, there will be a Crosscloud equivalent.

plan: wait and see? talk to developers.

User Adoption Hypotheses

- Early Crosscloud apps will be viral (>5%/wk) in some communities, with people getting each other to sign up
plan: encourage viral apps
- In time, some apps will be mega-hits, breaking into the mainstream
- In time, nearly everyone get an account (like email)
- Eventually, on average, folks will be willing to pay >\$10/yr
plan: charge money early & often (potential conflict with Research policies)

Next steps

- demo apps (validate basic concept)
- toolkit for very eager developers (validate developer enthusiasm)
- aim for some viral apps (validate user's willingness to adopt)
- ... while collaborating, and remaining flexible, since standards require a wide array of stakeholders

Conclusion

- Research is cool
- There are some hard problems to solve
- Everyone can help
- The results could help a lot of people

Any questions?

Adoption Questions

What would it take for you to develop your apps for Crosscloud?

What would it take for you to use a Crosscloud app?

What would it take for you to pay \$5/mo for a pod server?