Sustaining a (Large-Scale) Repository Architecture

Behind the Scenes of the Stanford Digital Repository

Justin Coyne (@j_coyne) & Mike Giarlo (@mjgiarlo)
Stanford Libraries
Stanford University
Stanford, California
Why We’re Here
Sustainability, Defined

“Software has the interesting property that when you make changes to it, either fixing bugs or adding features, the structural quality of the code will typically decline with every change. Especially when there is time-pressure, and there always is… But changing software hastily is self-defeating. Because lower quality code is more difficult to change.

“This leads to a vicious cycle of hasty changes that decrease quality which increases time pressure even further. Clearly, this way of evolving software is not sustainable. So for software, ‘sustainable’ actually means ‘evolvable.’

“The ISO standard 25010 for software product quality uses the term ‘maintainable’ exactly for this quality characteristic of being easy to change.”

-- Joost Visser, Sustainable Software with Agile
"A gardener's work is never at an end."

-- John Evelyn (1620-1706)

Software development is a leaky rowboat. Behavior changes are rowing--making progress toward a goal, however dimly glimpsed. Structure changes are bailing--not progress in a measurable sense but absolutely necessary for progress.
SDR in a Nutshell

- Born in 2006
- ~2.2M objects
- >1K collections
- 25 distinct workflows
- >500TB (and growing)
SDR Content Types
SDR Applications

- **Hydrus**
- ETDs (private, sorry!)
- **Pre-Assembly**
- **Argo**
- + background workers (“robots”)
SDR Technology
SDR Nuts & Bolts

- ~30-40 codebases
- ~1M lines of code
- ~60-70 virtual machines (staging & production)
Who Runs SDR?

- **Stanford University**
  - Top 10 (in US) research budget (> US$1.6B)
  - Top 5 (in US) endowment (> US$26B)
  - ~17K students (undergrad. + grad.)
  - > 2K faculty
  - 17 Nobel laureates

- **Stanford Libraries**
  - > 9.5M physical volumes
  - ~400 staff
  - > US$20M endowment
Who Runs SDR?
Who Runs SDR?

- Digital Infrastructure
  - Vivian Wong

- Engineer
  - Aaron Collier
  - Justin Coyne
  - Naomi Dushay
  - Mike Giarlo

- Engineer
  - John Martin
  - Justin Littman
  - Jeremy Nelson

- UX Designer
  - Astrid Usong

- Engineer & Product Manager
  - Peter Mangiafico
SDR Current State
Shiny Boutique Sites
Shiny Boutique Sites
Shiny Boutique Sites
Contributing to OSS but not using much
Staff turnover
Bad development practices
If you want to go far, go together.
Tight coupling to Fedora 3
Complex systems
What keeps us up at night
Better development practices
Prioritizing maintenance

• Maintenance work cycles
• Patch parties
• Weekly patch day
• First responder
Consolidate
Come back into the OSS fold
Refactoring to decouple/isolate F3
In Summary...
Questions?

Justin Coyne <jcoyne85@stanford.edu>
Mike Giarlo <mjgiarlo@stanford.edu>