

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](#)

Sustainability Metaphors

"A gardener's work
is never at an end."

-- John Evelyn
(1620-1706)



Kent Beck ✓

@KentBeck

Follow

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.

10:57 AM - 28 May 2019

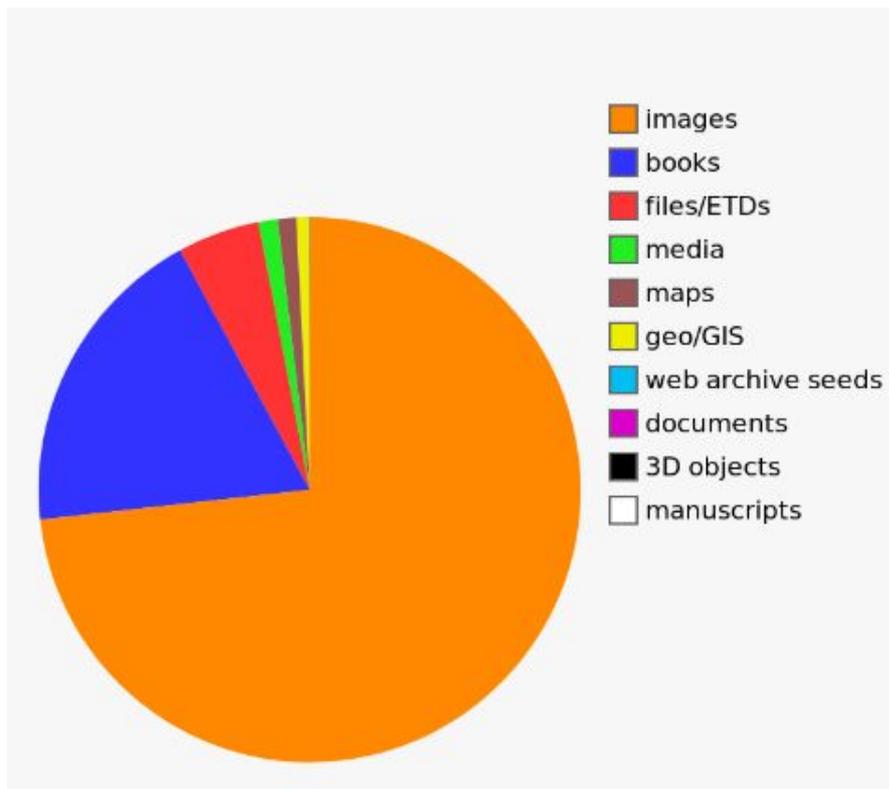
88 Retweets 226 Likes



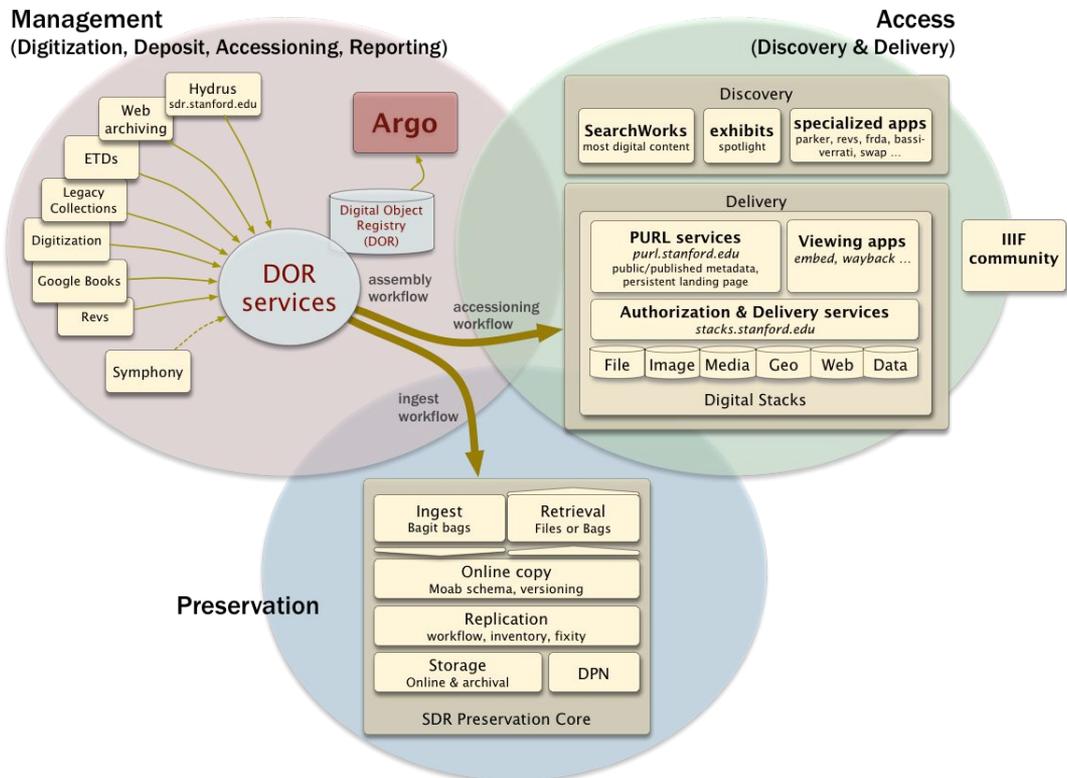
SDR in a Nutshell

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

SDR Content Types



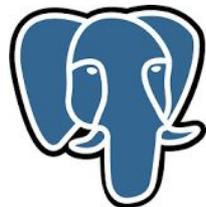
SDR Architecture



SDR Applications

- Hydrus
- ETDs (private, sorry!)
- Pre-Assembly
- Argo
- + background workers (“robots”)

SDR Technology



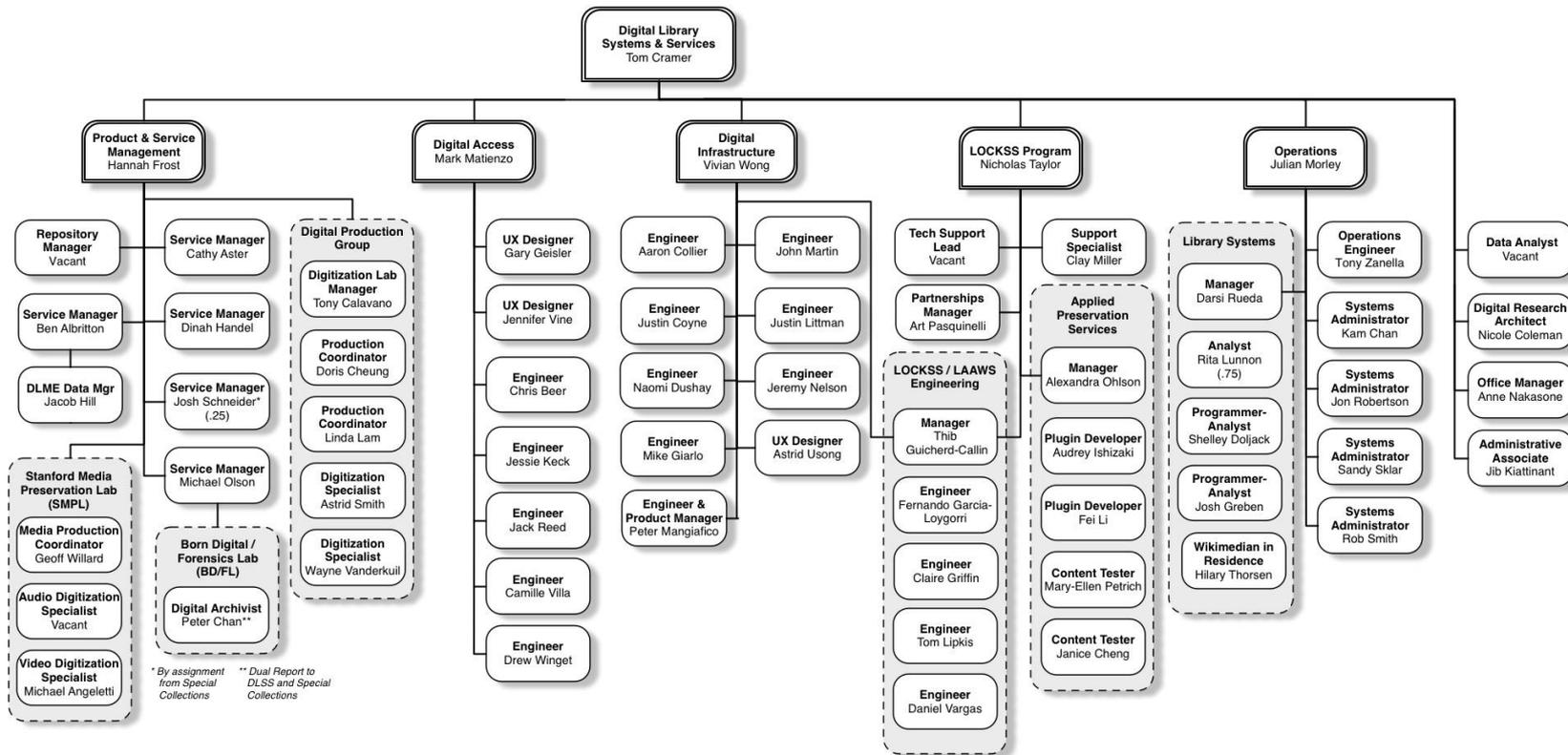
SDR Nuts & Bolts

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

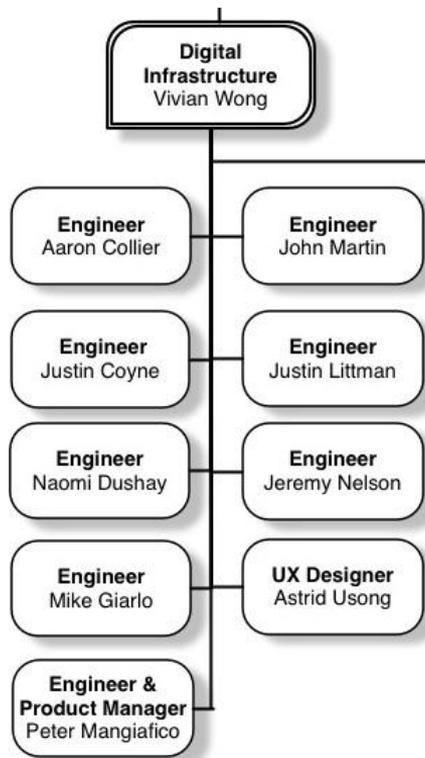
Who Runs SDR?

- **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?



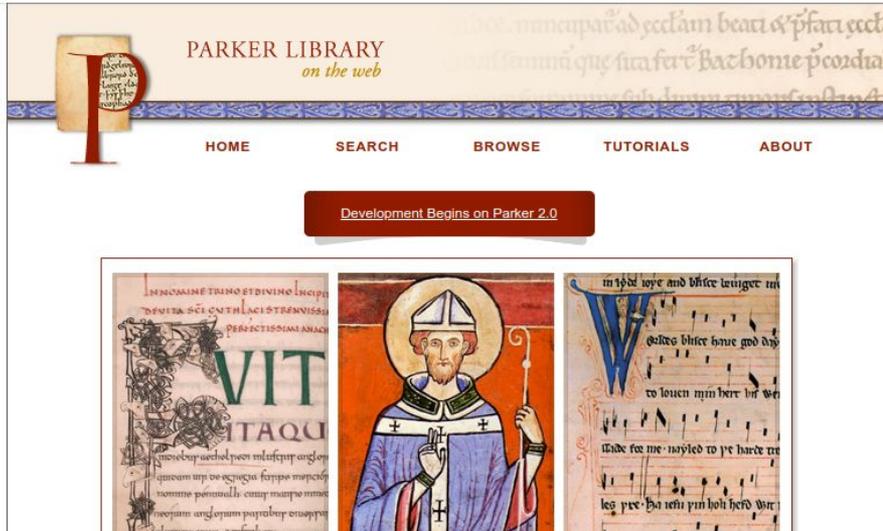
Who Runs SDR?



HOW DID WE
GET HERE?



Shiny Boutique Sites



Shiny Boutique Sites

PARKER LIBRARY
on the web

HOME SEARCH BROWSE TUTORIALS ABOUT

Development Begins on Parker 2.0

VIT
TAQU

W

Stanford LIBRARIES English | Italiano Feedback

Bassi-Veratti Collection

Home Collection highlights Background Content inventory About

Bassi-Veratti Collection

Search

COLLECTION HIGHLIGHTS More highlights

About Laura Bassi (5 Items)

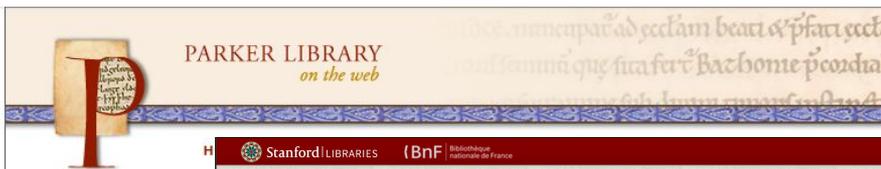
Poems in honor of Laura Bassi (9 Items)

Laura Bassi's awards and medals (8 Items)

About the collection

The online Bassi-Veratti Collection is a multi-year collaboration of the Stanford University Libraries, the Biblioteca Comunale dell'Archiginnasio, Bologna, Italy, and the Istituto per i Beni Artistici, Culturali e Naturali della Regione Emilia-Romagna, to produce a digital version of the archive of the influential woman scientist, Laura Bassi. Part of the rich collections of the Archiginnasio, the collection contains not only materials about Laura Bassi, but also those of her husband, scientist Giuseppe Veratti, their children, and his family. [more »](#)

Shiny Boutique Sites



Feedback
Bibliothèque nationale de France

French Revolution Digital Archive

A collaboration of the Stanford University Libraries and the Bibliothèque nationale de France

HOME
ARCHIVES PARLEMENTAIRES
IMAGES
ABOUT

French Revolution Digital Archive

SEARCH

+ options

Search ▾

Limit your search

- Political timeline >
- People >
- Document type >
- Medium (images) >
- Genre (images) >
- Artist (images) >
- Collector (images) >
- Volume (AP) >

About the collections

The *French Revolution Digital Archive (FRDA)* is a multi-year collaboration of the Stanford University Libraries and the *Bibliothèque nationale de France (BnF)* to produce a digital version of the key research sources of the French Revolution and make them available to the international scholarly community. The archive is based around two main resources, the *Archives parlementaires* and a vast corpus of images first brought together in 1989 and known as the *Images de la Révolution française*.

Archives parlementaires / Parliamentary archives

The *Archives parlementaires* is a chronologically-ordered edited collection of sources on the French Revolution. It was conceived in the mid 19th century as a project to produce a definitive record of parliamentary deliberations and also includes letters, reports, speeches, and other first-hand accounts from a great variety of published and archival sources. Because of copyright limitations, *FRDA* contains the *AP* volumes covering the years 1787-1794. The text of these volumes has been marked up using TEI so that speakers, places, dates, and terms in the published index can be easily found. Users can see both scanned images of the *AP* pages or just the texts.

Images de la Révolution française / Images of the French Revolution

The *Images* are composed of high-resolution digital images of approximately 14 000 individual visual items, primarily prints, but also illustrations, medals, coins, and other objects, which display aspects of the Revolution. These materials were selected, mainly from the collections of the *Département des Estampes et de la photographie*, but also from other BnF departments, and include thousands of images for the important collections entitled *Hennin* and *De Vinck*. Detailed metadata exists for the images, so that researchers can search by artist, subject, genre, and place.

Feedback
Bibliothèque nationale de France

Bassi-Veratti Collection

Home
Collection highlights
Background
Content inventory
About

COLLECTION HIGHLIGHTS

About Laura Bassi
(5 Items)

Poems in honor of Laura Bassi
(9 Items)

Laura Bassi's awards and medals
(8 Items)

More highlights

About the collection

The online Bassi-Veratti Collection is a multi-year collaboration of the Stanford University Libraries, the *Biblioteca Comunale dell'Archiginnasio*, Bologna, Italy, and the *Istituto per i Beni Artistici, Culturali e Naturali della Regione Emilia-Romagna*, to produce a digital version of the archive of the influential woman scientist, Laura Bassi. Part of the rich collections of the *Archiginnasio*, the collection contains not only materials about Laura Bassi, but also those of her husband, scientist *Giuseppe Veratti*, their children, and his family. [more »](#)

Contributing to OSS but not using much



hyku

Staff turnover

Bad development practices



Come, all ye sons of labor
Who wish to change your lot,
Who're spunk enough to travel
Beyond your native cot:
Come, leave the crowded cities
Where work is coartage,
And come with us to settle
In western Kansas' plains.
— from "Immigration Song"

If you want to go far, go together.



Tight coupling to Fedora 3



**eXtensible
M
L** Markup Language

The logo for eXtensible Markup Language (XML) is displayed. It features the word "eXtensible" in black, with the "X" inside a blue hexagon. Below it, the word "Markup" is in black, with the "M" inside a yellow-green hexagon. Below that, the word "Language" is in black, with the "L" inside a pink hexagon. The three hexagons are arranged in a descending staircase pattern from left to right.



Complex systems

What keeps us up at night

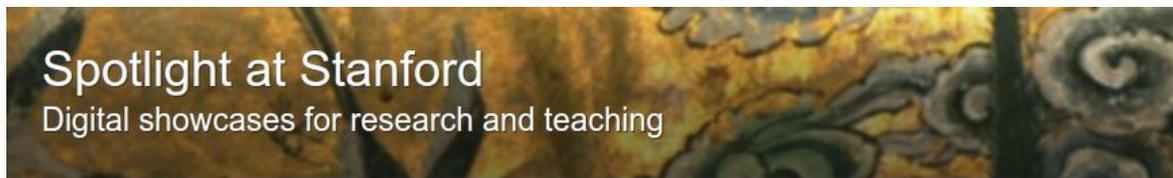


Better development practices

Prioritizing maintenance

- Maintenance work cycles
- Patch parties
- Weekly patch day
- First responder

Consolidate



- All
- Arts & Humanities
- Government & Society
- Maps & Geography
- Science & Engineering
- Stanfordiana



Stanford University Piano Roll Archive



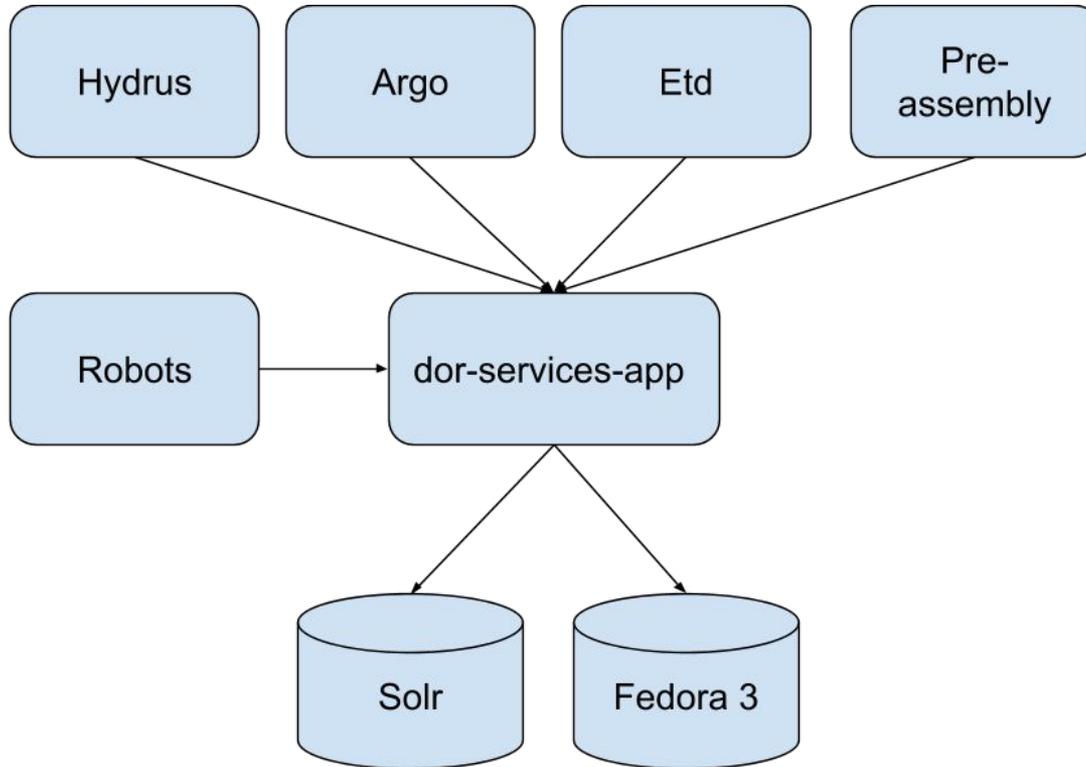
Parker Library On the Web



Transcontinental Railroad

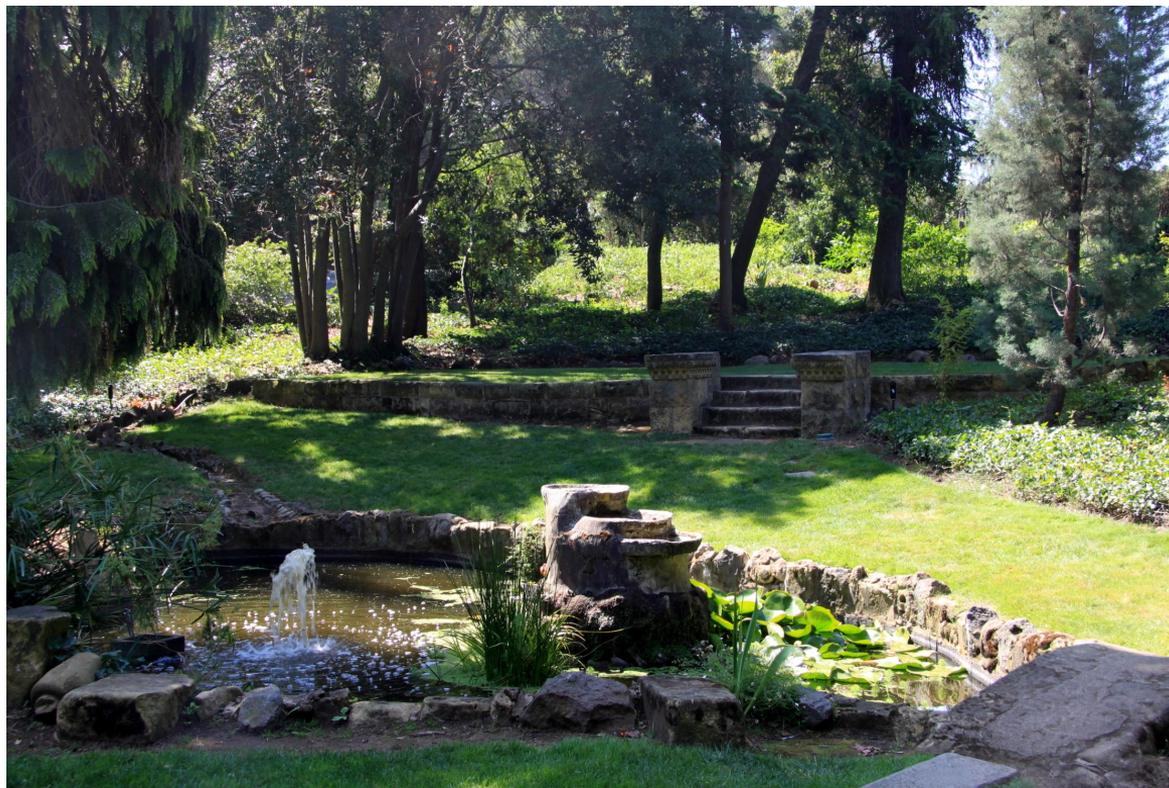
**Come back into the
OSS fold**

Refactoring to decouple/isolate F3



WARNING

In Summary...



Questions?

Justin Coyne <jcoyne85@stanford.edu>

Mike Giarlo <mjgiarlo@stanford.edu>