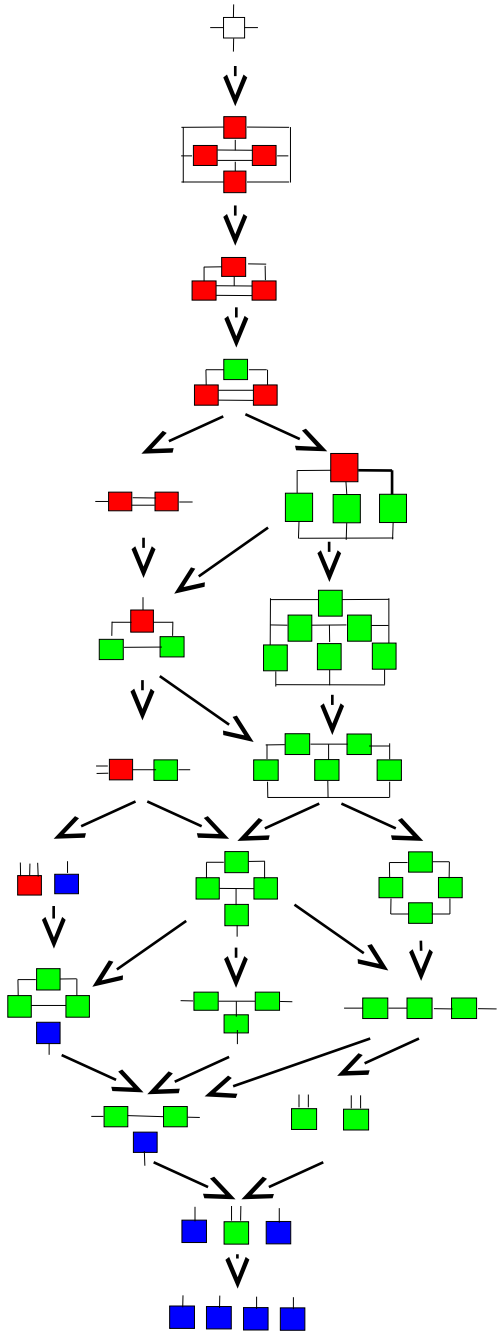
An Open-Source Python Package for OCCAM

CS Capstone Presentation

2019-01-24

Joe Fusion

Systems Science Program, PSU



Who we are

Systems Science Program

(Harder House)

Joe Fusion

Martin Zwick

and...
the occam-ra
open source team!

What is OCCAM

- a software implementation of Reconstructability Analysis

What is Reconstructability Analysis

a.k.a. discrete multivariate modeling (DMM)

- a probabilistic graphical modeling method, related to logistic regression and Bayesian networks
- a modeling method that is excellent at:
  - handling qualitative data: healthcare data, gene sequences, survey data, demographics
  - finding nonlinearities and multivariate interactions with few assumptions
  - producing *interpretable* models that are easy to understand & explain
  - data mining – searching through many models

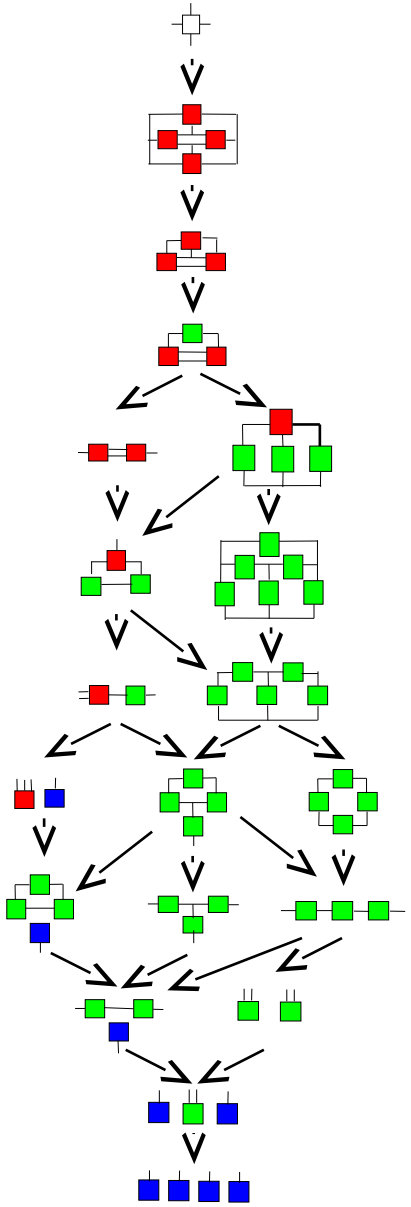
This is the *lattice of structures* for a 4-variable system.

It shows 20 general structures, or 114 specific structures with variable rearrangement.

For 5 variables: 180 and 6,894

For 6: 16,143 and 7,785,062

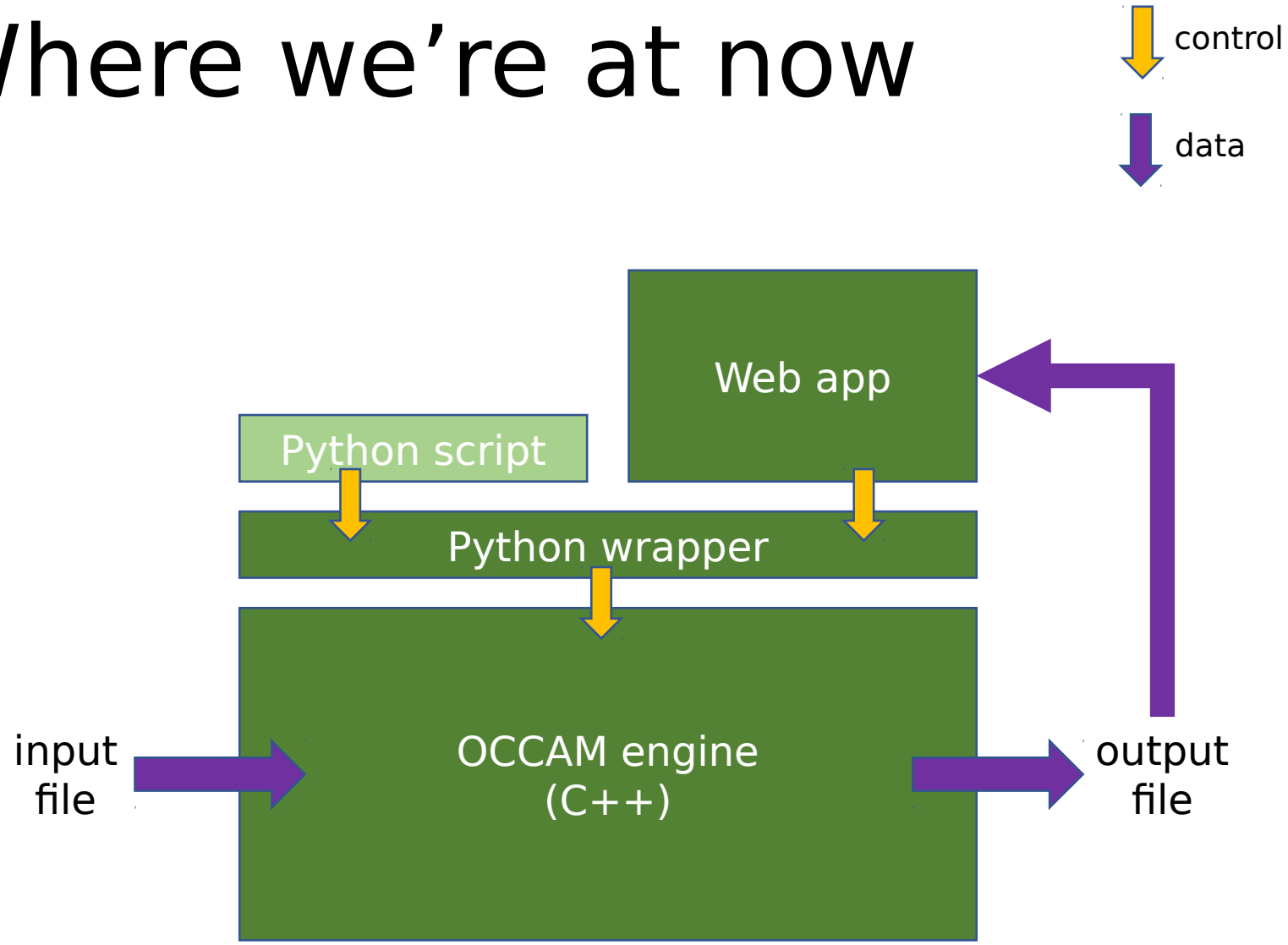
RA has user-friendly search



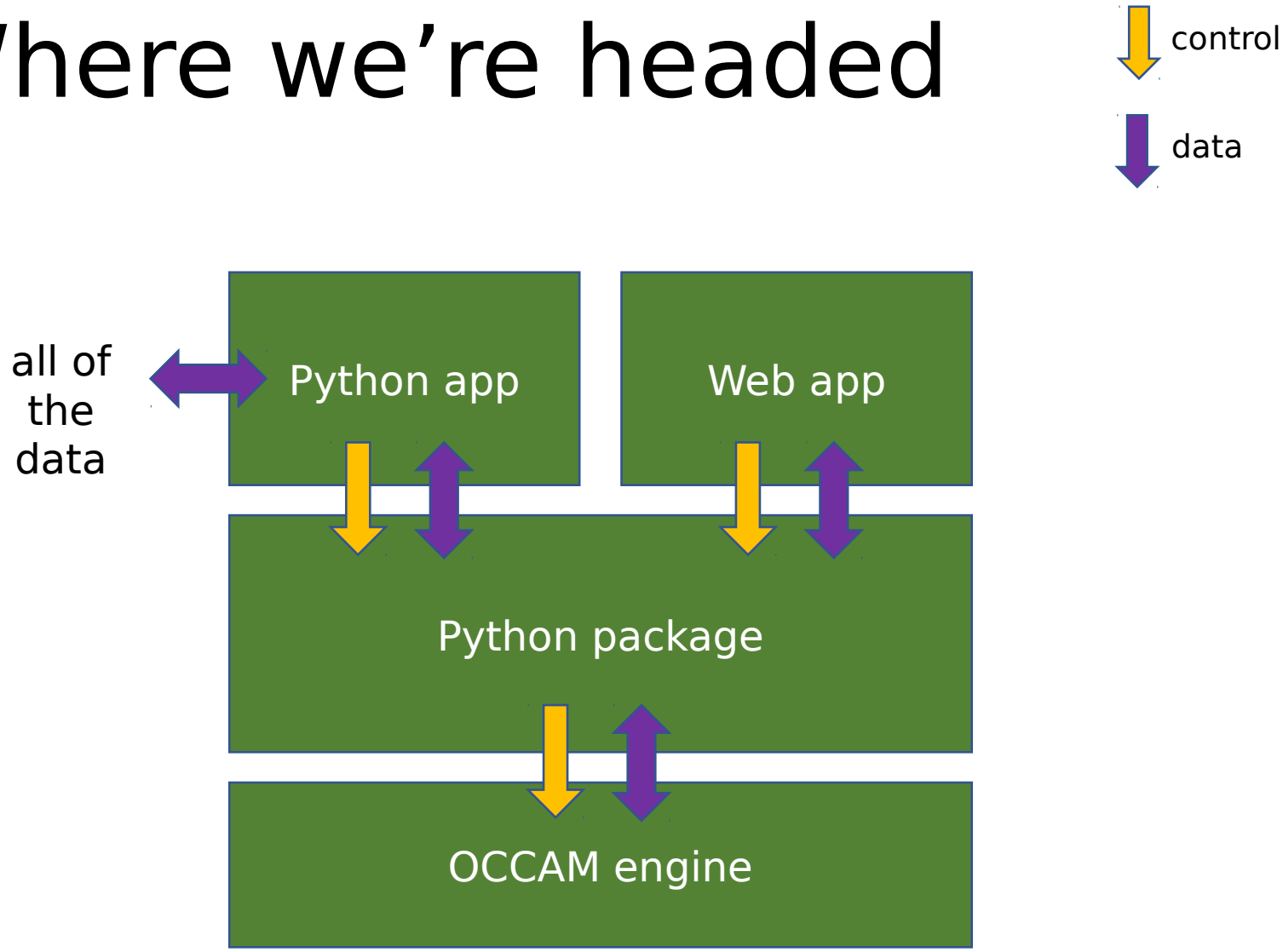
What is OCCAM

- a software implementation of RA
- a C++ application with Python wrapper
- used almost exclusively via web interface through Python cgi scripts
- an established research tool
  - taught in 2 classes, part of 3 graduate certificates
  - part of many papers, dissertations, grants, more
  - developed for decades – one GA at a time
- recently released as open source!

Where we're at now



Where we're headed



What we’re looking for

- a proper Python package
  - suitable for inclusion in a major repository
  - standard structure, interface, I/O
  - plays nicely with other ML & data packages
- a well-organized open-source project
- a design that’s ready for the future
  - moving engine code into Python
  - switching to better libraries for math, stats
  - more power: parallelizing, TensorFlow

Why we’re asking for this

- open-source is the future for OCCAM
- get RA into more people’s hands
- take advantage of better software & hardware
- solve more, bigger, harder problems!

What you get out of this

- work in machine learning & data mining
- design & implement a Python package
- learn about RA
- be part of a new open-source project
- work with an established research team
- contribute to many dissertations, theses, grants, and class projects, for years to come

Resources

the OCCAM project on GitHub

occam-ra.io

jfusion@pdx.edu

zwick@pdx.edu