



# Facilitating Researchers with HTC Workloads

Lauren Michael, Research Computing Facilitator  
Center for High Throughput Computing  
22 Mar 2016



# Overview

- **Research Computing at UW-Madison**
- **The right people for support**
- **The right model of support**
- **Reaching researchers with HTC needs**
- **Facilitating researchers with HTC needs**

# CHTC Services

Center for **H**igh **T**hroughput **C**omputing, est. 2006

## > Large-scale, campus-shared computing systems

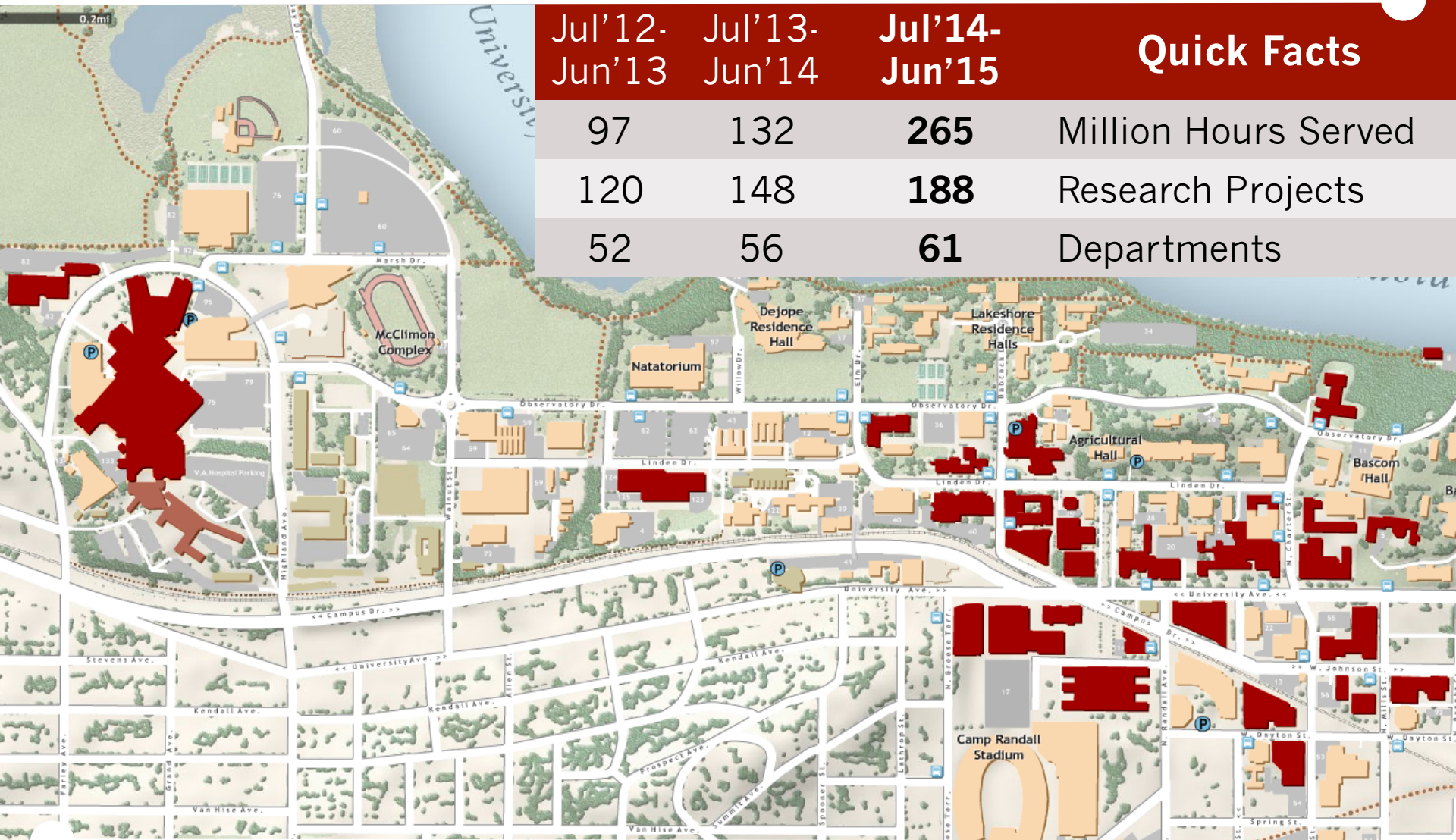
- high-throughput computing (HTC) and high-performance computing (HPC) systems

- **all standard services provided free-of-charge**

- hardware buy-in options
- support and training for using our systems
- proposal assistance

- **[chtc.cs.wisc.edu](http://chtc.cs.wisc.edu)**





Researchers who use the CHTC are located all over campus (red buildings)

# Accessible HTC Computing:



# CHTC-Accessible Computing:



**“UW Grid”**

**CHTC**

# CHTC-Accessible Computing:



## Open Science Grid

“UW Grid”

CHTC





Jul'12- Jun'13	Jul'13- Jun'14	Jul'14- Jun'15	Quick Facts
97	132	<b>265</b>	Million Hours Served
120	148	<b>188</b>	Research Projects
52	56	<b>61</b>	Departments

Individual researchers:  
**30 years of computing**  
per day

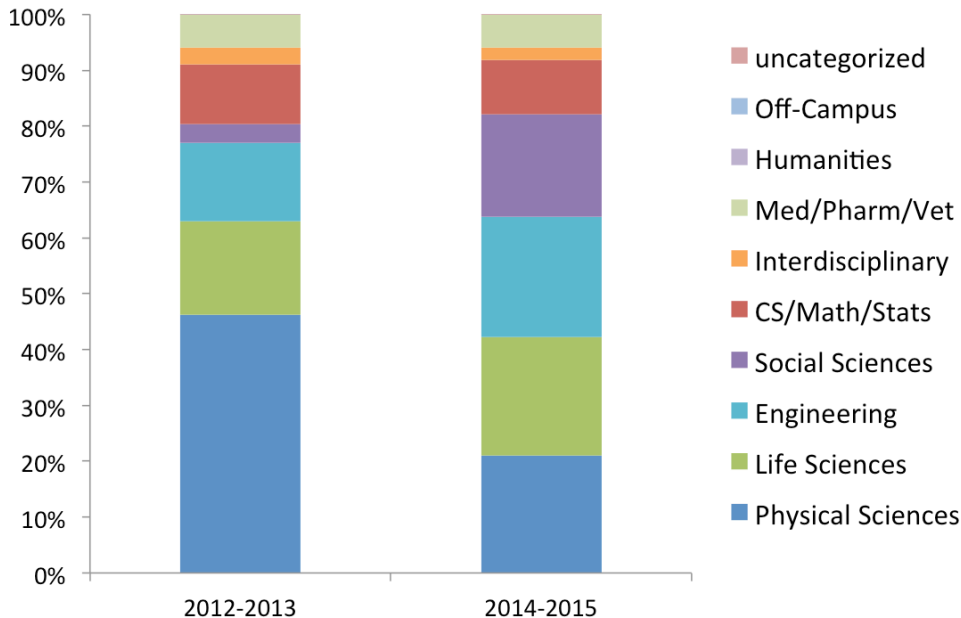
Researchers who use the CHTC are located all over campus (red buildings)



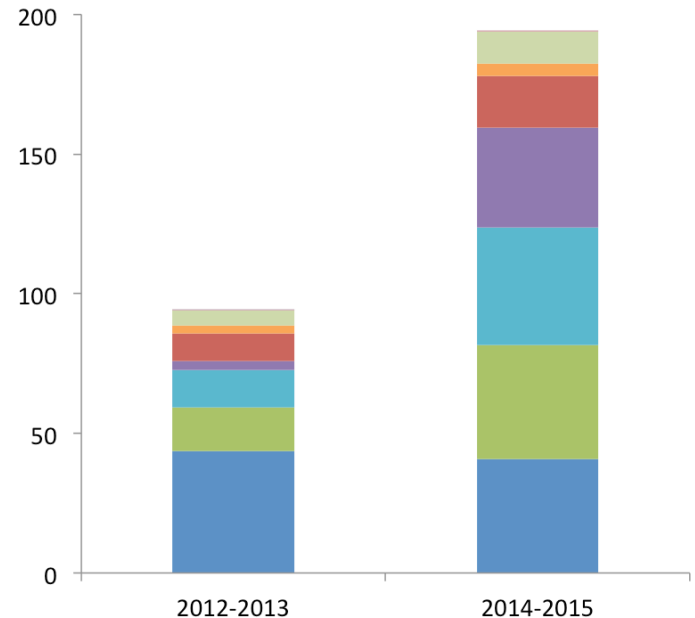


Jul'12- Jun'13	Jul'13- Jun'14	Jul'14- Jun'15	Quick Facts
97	132	<b>265</b>	Million Hours Served
120	148	<b>188</b>	Research Projects
52	56	<b>61</b>	Departments

**% of Compute Hours**



**Absolute Compute Hours**



Researchers who use the CHTC are located all over campus (red buildings)



~~Users~~ are people.

Researchers are people.



Make it easy for researchers  
to find **the right people.**

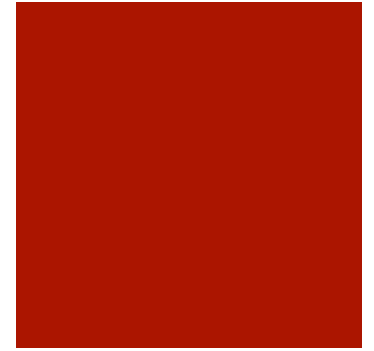
## “Facilitators”

- consultants/liaisons for  
research computing
- identify with the *user*  
perspective



# Research Computing

## Facilitation



- proactive engagement
- personalized guidance
- teach-to-fish
- relationship building
- advocating for researchers
- professional networking

[http://aci-ref.github.io/facilitation\\_best\\_practices/](http://aci-ref.github.io/facilitation_best_practices/)

Michael, L. and Maas, B. Research Computing Facilitators: The Missing Human Link in Needs-Based Research Cyberinfrastructure. ECAR, May 16, 2016.

# Research Computing

## Facilitation



- proactive engagement
- personalized guidance
- teach-to-fish
- relationship building
- advocating for researchers
- professional networking

***beyond-the-helpdesk***

***deliberate  
designated  
dedicated***

[http://aci-ref.github.io/facilitation\\_best\\_practices/](http://aci-ref.github.io/facilitation_best_practices/)

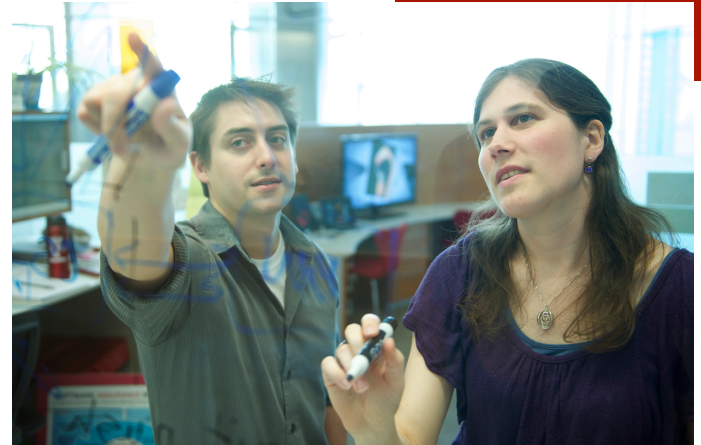
Michael, L. and Maas, B. Research Computing Facilitators: The Missing Human Link in Needs-Based Research Cyberinfrastructure. ECAR, May 16, 2016.

# Research Computing *Facilitators*

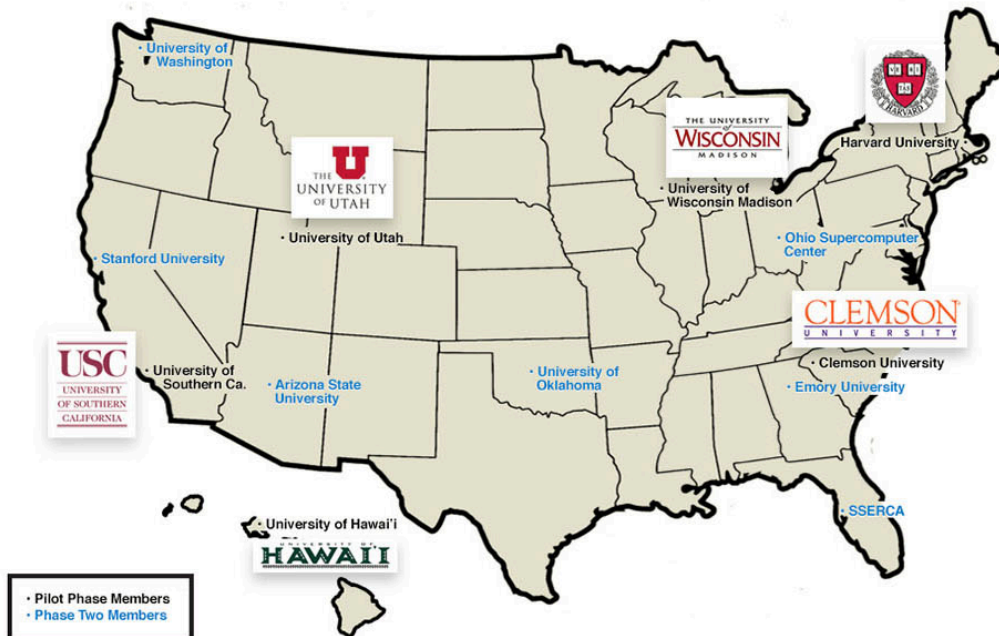
**Communication,  
Teaching,  
Leadership**

**Scholarship  
Experience**

**Technical  
Skills**

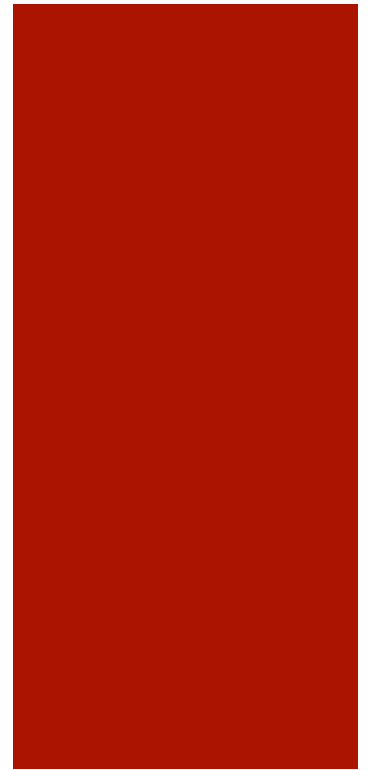


# Share support practices with others!



**Advanced Computing Infrastructure  
Research and Education Facilitators (ACI-REF)**

<https://aciref.org>



# Reaching Researchers with HTC Needs





# **How to attract researchers with HTC needs.**



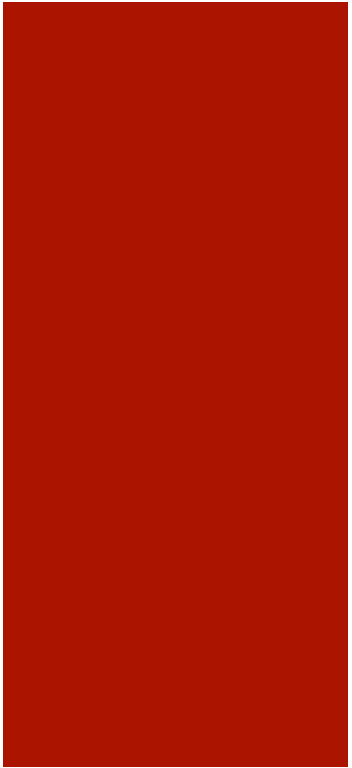
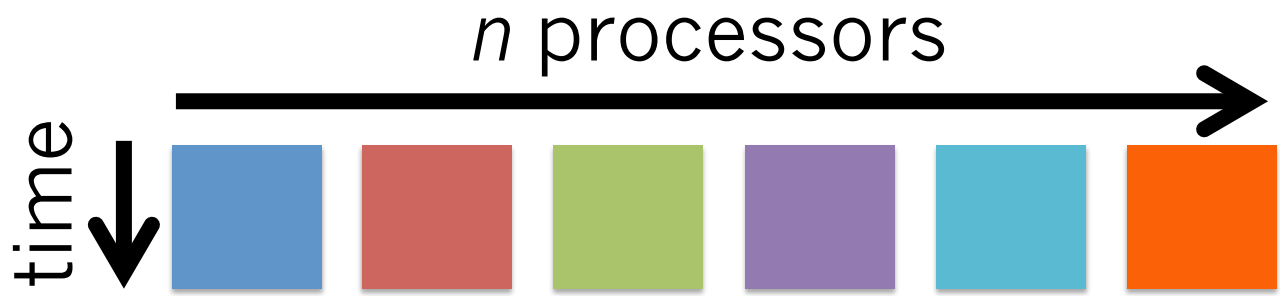
~~How to attract researchers  
with HTC needs.~~

**Just help researchers with**

**COMPUTING**

(not “HTC” or “HPC”, etc.)

**Promote AWARENESS**



# Identifying HTC-able Research Problems

# Observables of HTC-able workloads



**long for-loops**

**numerous similar files per job**

file system performance issues?

**numerous similar jobs in the queue**

**single-server, multi-core software**

Matlab's "parfor", OpenMP, etc.

# Researcher indicators of potential HTC



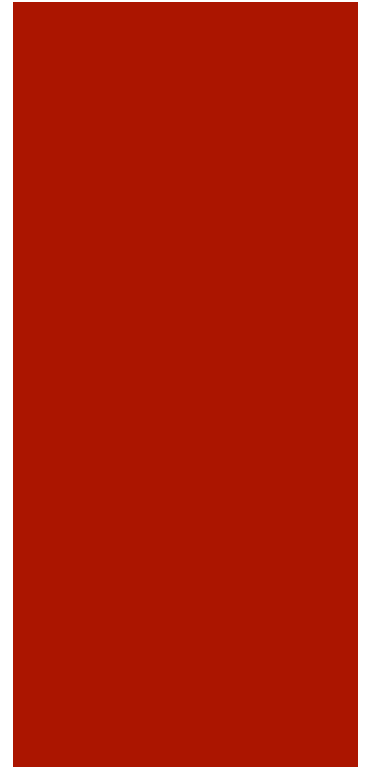
**“model optimization”**

**“parameter sweep”**

**“text analysis”**

**“image analysis”**

**“library” “database for output”**



# Facilitating Researchers with HTC Needs

# Teach Key Components of Executing HTC Work



**Breaking Up Workloads**

**Data Management**

**Software Portability**

**Automation**

**Long-Term Pay-Offs**

# Breaking Up Workloads

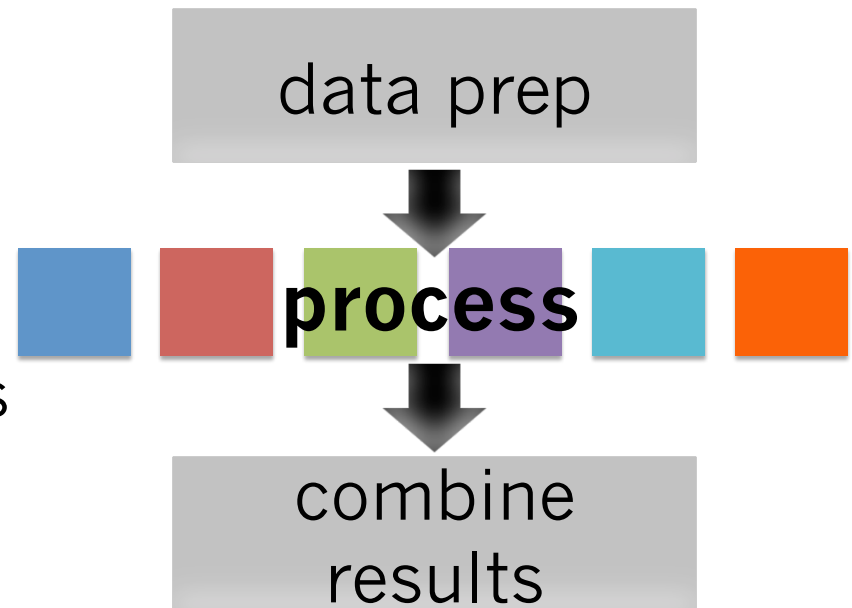
Discuss *conceptually*  
for programs *AND* files

Plan programming and/or  
execution modifications

Batch multiple *very* short tasks  
together in each “job”

Test small, understand resource  
requirements, optimize, scale up

Checkpointing!





# Data Management



(beyond typical management of metadata, paths, etc.)

**scalable file-naming and organization**

**organize “shared” data and job-specific data separately, avoiding unnecessary duplicates**

**compression for file transfers**

**stage large files off of the submit server**


**organize software files, too**

Teach Researchers

**How Software Works (!!!)**



# Teach Researchers How Software Works (!!!)



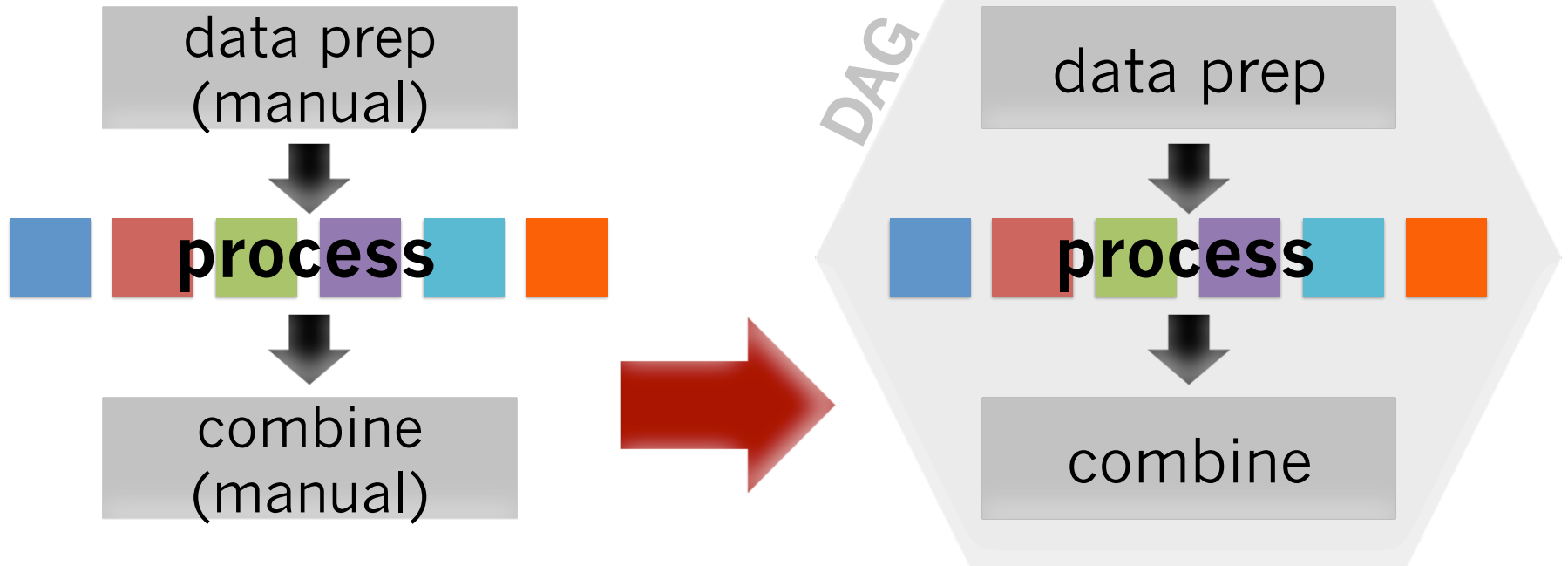
**Portability is a must for HTC scalability**

**Empower full control over versions**  
reproducibility! experimentation!

**Reduce researcher dependency on staff and specific, local configurations**  
extensibility for future work and compute systems

**Acknowledge licensing limitations to inform software selection**

# Encourage Automation



# Help Researchers to See **Big Picture Payoffs**

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?  
(ACROSS FIVE YEARS)

HOW MUCH TIME YOU SHAVE OFF

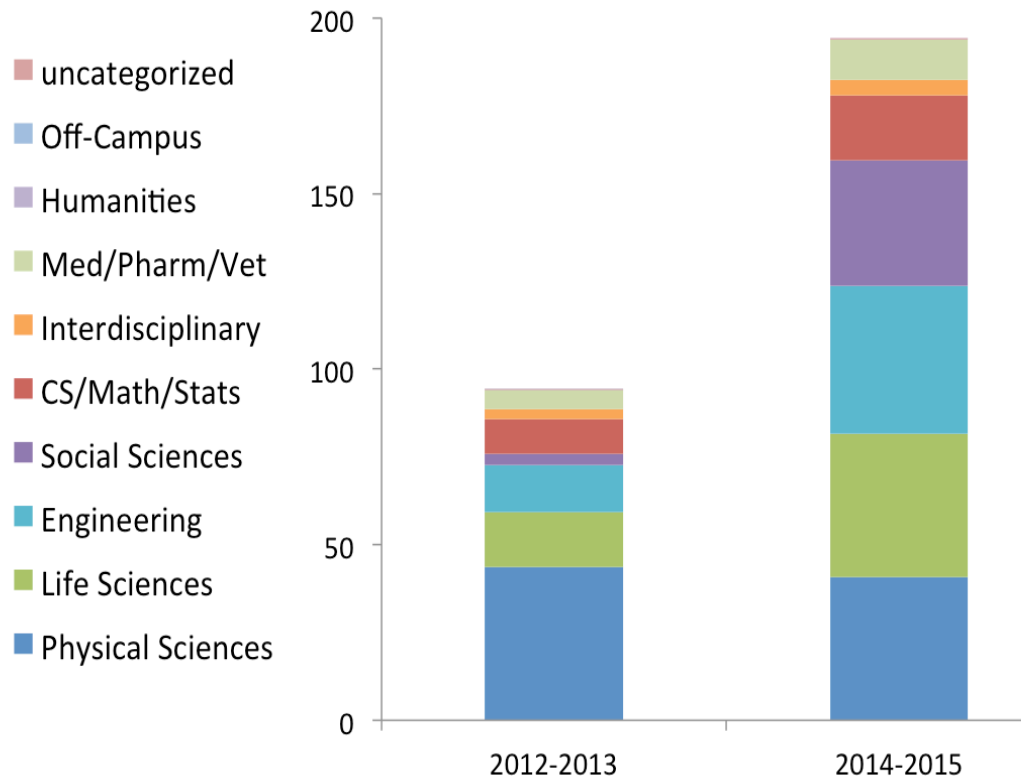
	HOW OFTEN YOU DO THE TASK					
	50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
6 HOURS				2 MONTHS	2 WEEKS	1 DAY
1 DAY					8 WEEKS	5 DAYS

# Infrastructure

= Technology + People



## Compute Hours Delivered at UW-Madison

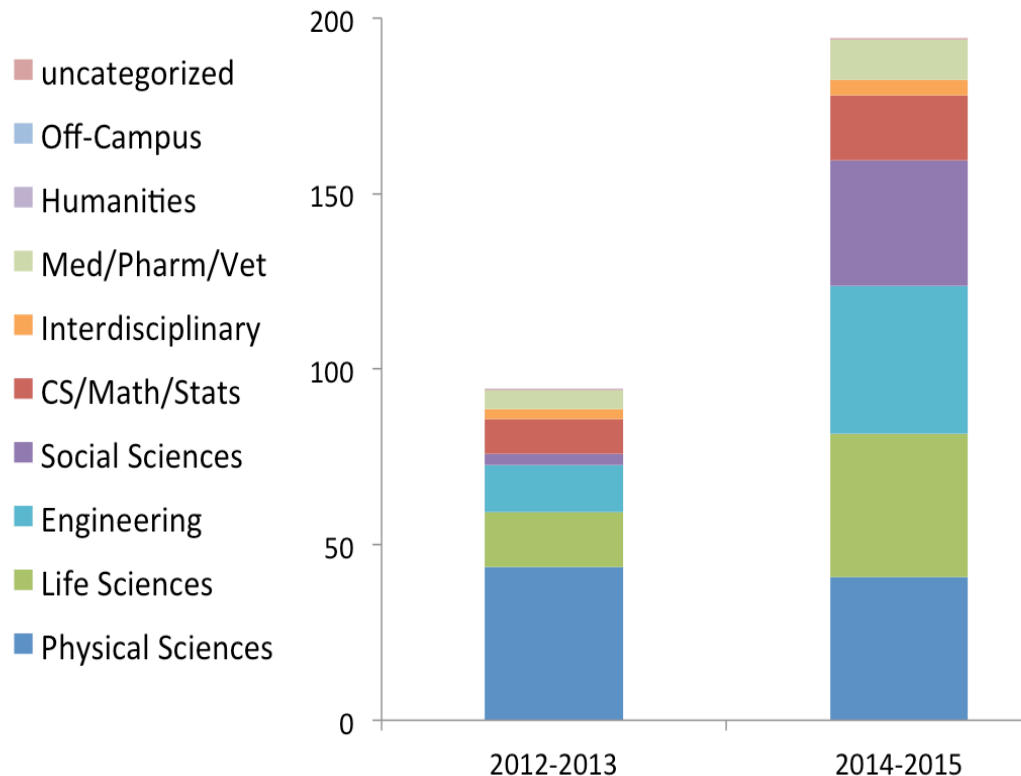


# Infrastructure

## = Technology + People



### Compute Hours Delivered at UW-Madison



“Lauren Michael’s value to the overall enterprise is hard to overestimate. Putting someone so capable of explaining procedures in simple but powerful terms greatly enhances the effective power and utility of the millions invested in the hardware.”

-Tom Givnish, Botany

“With extensive help from [RCFs] we adapted our workflow to run seamlessly on the HTC platform and have now utilized these resources to construct the largest computed diffusion database in the world.”

-Dane Morgan,  
Materials Science and Engineering



# Facilitating Researchers with HTC Workloads

Lauren Michael, Research Computing Facilitator  
Center for High Throughput Computing  
22 Mar 2016