

# Transparency: An Architecture Principle for Socio-Technical Ecosystems

Felix Bachmann and Linda Northrop, SEI  
SATURN 2014

# XSEDE

Extreme Science and Engineering  
Discovery Environment



# SEI/Carnegie Mellon Team

- Felix Bachmann
- Chris Bogart
- Jim Herbsleb
- Mark Klein
- Mike Konrad
- Linda Northrop
- Kurt Wallnau





The eXtreme Science and Engineering Discovery Environment (XSEDE) funded by the National Science Foundation (NSF) enhances the productivity of scientists and engineers.



XSEDE is the framework for a national cyber-infrastructure ecosystem, serving as a platform for multi-scale cyber-infrastructure integration for scientific collaboration.

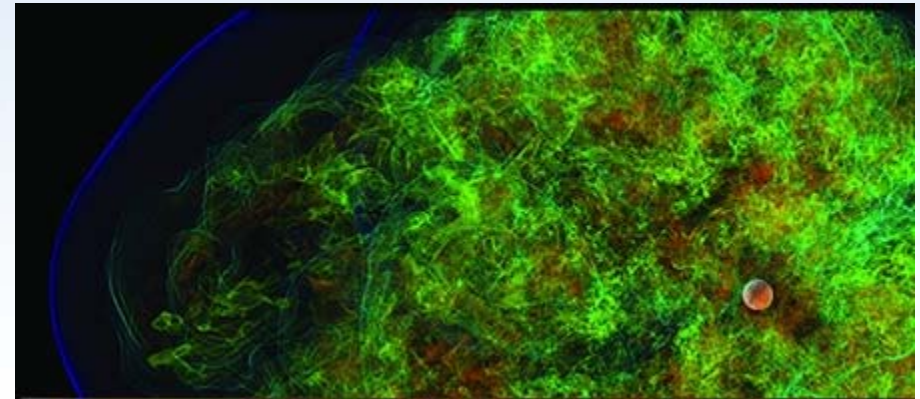


XSEDE





XSEDE's innovative, open standards-based architecture facilitates an unparalleled level of integration.



Enabling this architecture are XSEDE's professional systems engineering approach and technology insertion efforts, which ensure robustness and security while continuously incorporating new technologies.



# XSEDE: A Socio-Technical Ecosystem

## XSEDE

- Provides a **computational infrastructure** that helps scientists access many different supercomputing sites that have different policies and software infrastructures
- Has **users** who come from **different scientific and university communities** with different priorities
- Has **developers and architects** with **global priorities** to continuously evolve the XSEDE cyber-infrastructure, but who come from different institutions and sub-communities with **local priorities**

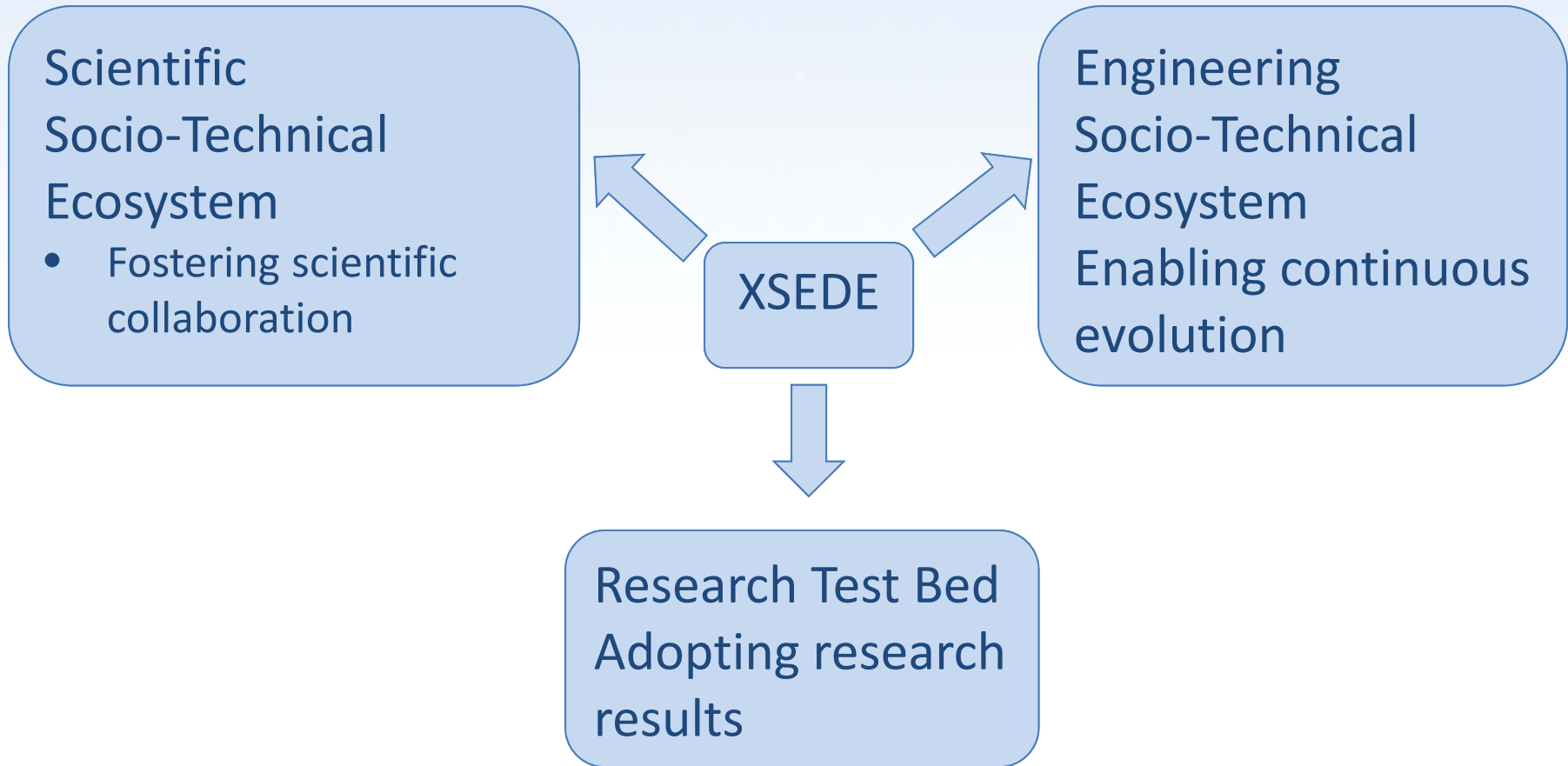


# The Nature of XSEDE

- The strength of the XSEDE community is its ability to engage in collaborative, community-held problem-solving.
- Additional "language and protocol" with more of an engineering mindset are needed to continuously and effectively support collaborative transformational science.



# Views of XSEDE



# SEI's Involvement

The SEI developed an engineering approach for XSEDE that addressed the twin challenges of

- establishing sound architecture-centric engineering practices that enable systematic, measured improvement in products and services; and
- introducing novel engineering practices that address novel challenges posed by XSEDE's status as a socio-technical ecosystem.

Initially, the XSEDE community was unfamiliar with an architecture-centric approach.

- We focused on establishing a core set of practices





# XSEDE Challenges

In a loosely coupled ecosystem like XSEDE, engineering practices that require common, agreed upon understanding and central control do not work.

- For example, it was not feasible to elicit requirements from all groups within XSEDE.
- Requirement gathering was replaced by collecting high level use cases.





# Practices to Support Distributed, Decentralized Environment

- Use case development annotated with quality attributes
- Architecture engineering, including documentation and Active Design Reviews
- Collaborative design and implementation, supported by Google Docs and Jira
- Communication through artifact documentation, wiki, discussion threads, email, meeting minutes, conversation
- Governance via WBS hierarchy, forums, advisory committees
- Progress tracked locally and reported up



# Community Engineering – First Step





## Current State

The XSEDE community is executing architecture-centric practices well and there seems to be a commitment to continuous improvement.

XSEDE staff strive for deliberate transparency in their engineering process efforts.

Productivity is not yet acceptable.

- The issues are often not visible to the XSEDE staff and therefore it is not clear what to improve and how.
- More transparency and accessibility would be helpful.
- Quality assurance is important, but it is difficult to engage people in reviews. Reviews are time consuming.

The XSEDE logo is rendered in a large, bold, white sans-serif font against a dark blue background with a grid pattern. It is located in the bottom right corner of the slide.

XSEDE

# Small-Scale Development vs. Large-Scale

## Small Scale

Short daily meetings allow teams to share all Information.

Can ask a team member at any time for missing information.

Quality assurance done by working in pairs or through short reviews (context is already known).

All team members are informed about changes.

## Large Scale – Socio-Technical Ecosystems (STE) Challenges

Requires creation of extensive documentation. Information is in many places.

Difficult to find the right artifacts. Use of “social network” to find the right person.

Quality assurance requires Input from multiple teams. Difficult to coordinate; difficult to surface local issues and understand global impact.

Difficult to determine who needs change information and therefore usually broadcasted. Leads to information overload.

## XSEDE Challenges

Need to participate in many meetings. Presents logistical problems and is costly.

Many documents and many places where information is stored. Many meetings to understand who knows what.

Reviews take a long time and decisions might be deferred until review is complete.

Trying to avoid change by lowering development risk.



# Lack of Transparency

- Local groups optimize and innovate as appropriate. The information they create is inherently distributed and not necessarily visible to others.
- People desire to keep unfinished changes private until they reach a “publishable state”
- Sheer volume of concurrent activities and artifacts makes information opaque
- Textual documents are in many places
  - Produced using many different tools
  - Used by different participants
  - Hard to get an overview



# Results of Transparency Deficiencies

- Not everyone has the information they need in an appropriate time to efficiently develop products and suggest improvements.
  - More time is needed for reviews and decisions
  - Lack of insight into status and root cause of local issues
- Stakeholder feedback is not efficiently incorporated into the process.
- Engineering process is difficult for broader XSEDE community to see and understand.





# Transparency Goals

- Move from point to point communication to information sharing and collaborative workspaces
- Harness strength of the XSEDE community through a more transparent environment
- Relieve bottlenecks
- Provide more openness and navigability for collections of documents, message, wikis, etc.
- Close the loop to the stakeholders, especially scientist end users
- Work to harmonize current engineering practice with research culture



XSEDE

# Our Efforts to Increase Transparency

We are working to

- discover valuable information
- make that information visible
- make that information accessible
- facilitate greater and more productive collaboration

We are using

- Information flow analysis
- Online deliberation tools
- Machine learning tools
- Visualization tools

We are trying to provide the **right information** to the **right people** at the **right time**.



# Collecting Information for Flow Analysis

- Interviews
- Access to project data

The image displays three overlapping screenshots from the XSEDE project environment:

- System Dashboard:** Shows a JIRA interface with sections for 'Introduction', 'Assigned to Me' (no issues), and 'Activity Stream' (XSEDE).
- XSEDE Project Wiki:** A redirected page from FrontPage, welcoming staff and providing a link to the XSEDE Wiki FAQ.
- Risks Summary:** A detailed risk management dashboard.
 

Risk Summary Stats	
Total Number of Risks:	179
Number of non-retired Risks:	130
High Impact Risks:	35
High Probability Risks:	17
High Risk-Level:	10
Key Risks:	4
Current Risks:	126

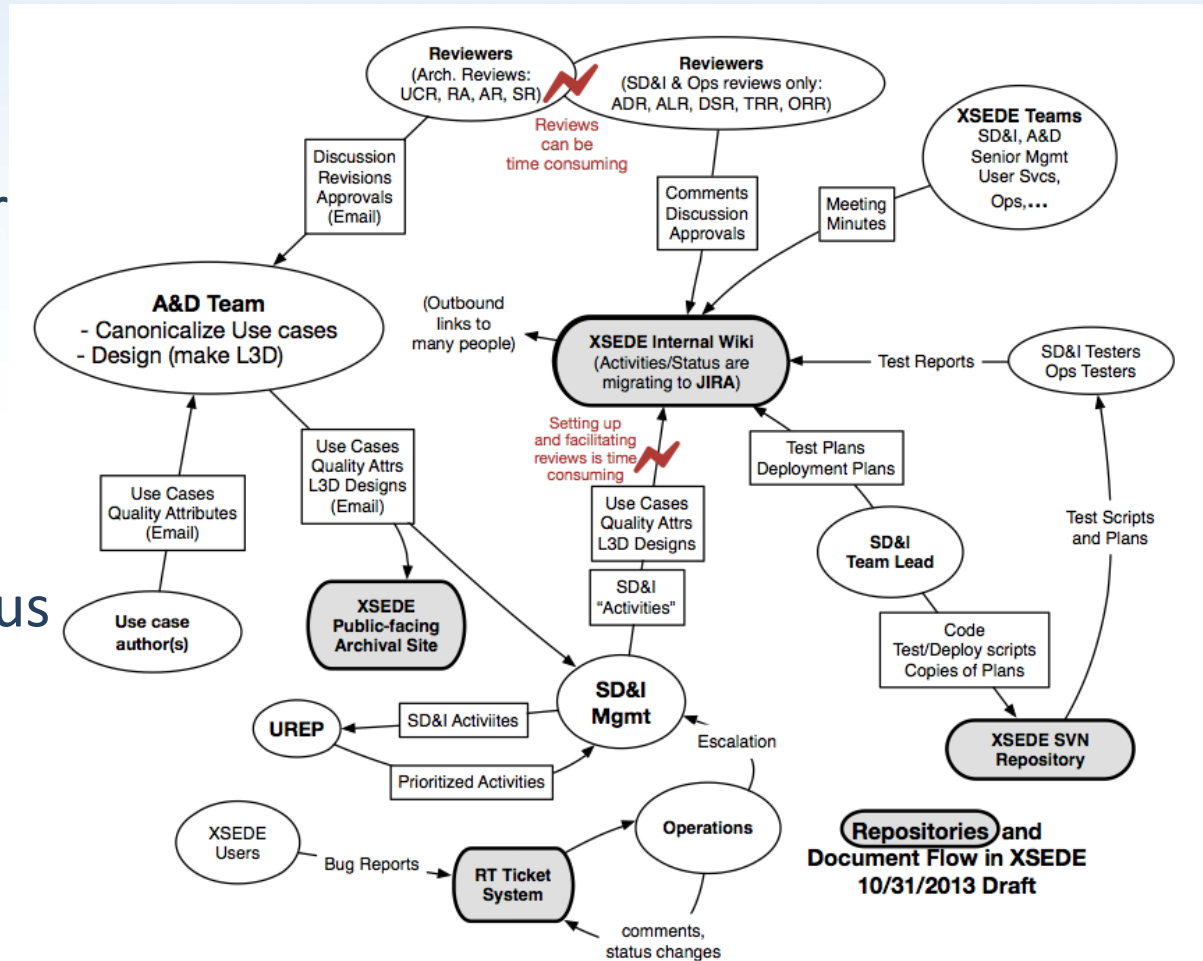
		All Risks			Non-retired Risks		
Probability	Impact	1	8	12	1	6	10
		17	35	23	16	30	9
		31	28	24	21	21	16

Non-retired Risks						
130 risks						
Risk Id	Risk	Risk Level	Probability	Impact	Status	Subproj
398	Extensions to Standard-Based Services	High	High	High	Monitor	1.5.2 Ext Collabora Science Support

# Analyzing the Flow of XSEDE Process - 1

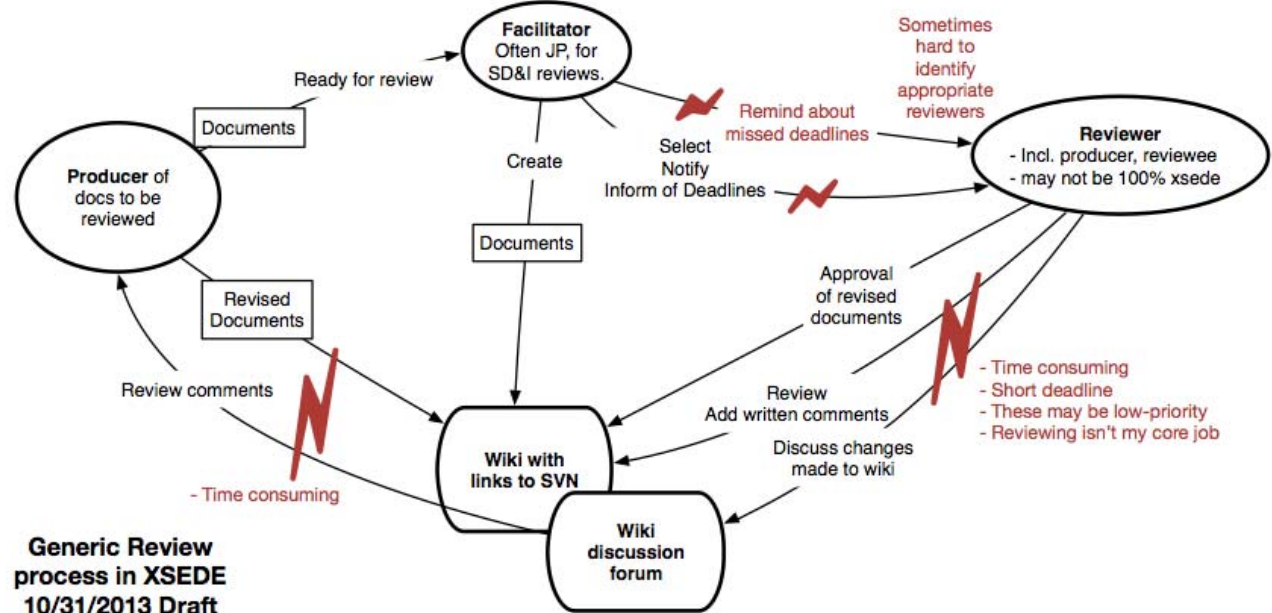
- Teams use different repositories as appropriate for their work: wiki, SVN, google docs, ticket system, email lists, etc.
- Flow diagram helps us understand who is seeing what information





# Analyzing the Flow of XSEDE Process - 2

- Also helps capture where people perceive friction points to be.
- Example: Generalized diagram of reviews shows where people have suggested needs for improvement in the process





# Experiment with Online Deliberation

*Ideascale*: Idea solicitation tool is available for use by any interested XSEDE group. Experiments are being run to determine if communication improves.

Got feedback/ideas for Pugh & You?

I suggest you...

Submit New Idea

## Campaigns

- All Ideas
- Crime
- City Services
- Deficit Reduction
- Education
- Job Development
- Police Response
- Returning Citizens
- Transportation
- Taxes

## Browse Popular Ideas

Recent (28) Popular (28) Hot (0) In Review (0) In Progress (0) Complete (0)

I agree

9  
votes

I disagree

JOB DEVELOPMENT »

### HAVE AN IDEA HOW TO GET RID OF ALL ABANDONED PROPERTY

We need to put all city owned buildings & property that are not in use or don't have current plans in an online auction that runs year round and raise money for the demolition of the buildings that don't sell and for the maintained of the web site and it doesn't have to be crazy expensive like the one they used this year like bid4assets.com on the county side that auction was such a failure if it was run on the existing ... more »



# Machine Learning

- Investigated the use of probabilistic topic models
- Experimented with tensor based mixed membership stochastic block models
- Developed an initial set of ideas for how these results can be applied to XSEDE to structure the information space to enable targeted notification and effective navigation





# Probabilistic Topic Models

*Models help to determine similar artifacts.*

When someone tries to specify an use case, a notification can be provided informing that person that someone else already has a use case that is similar.

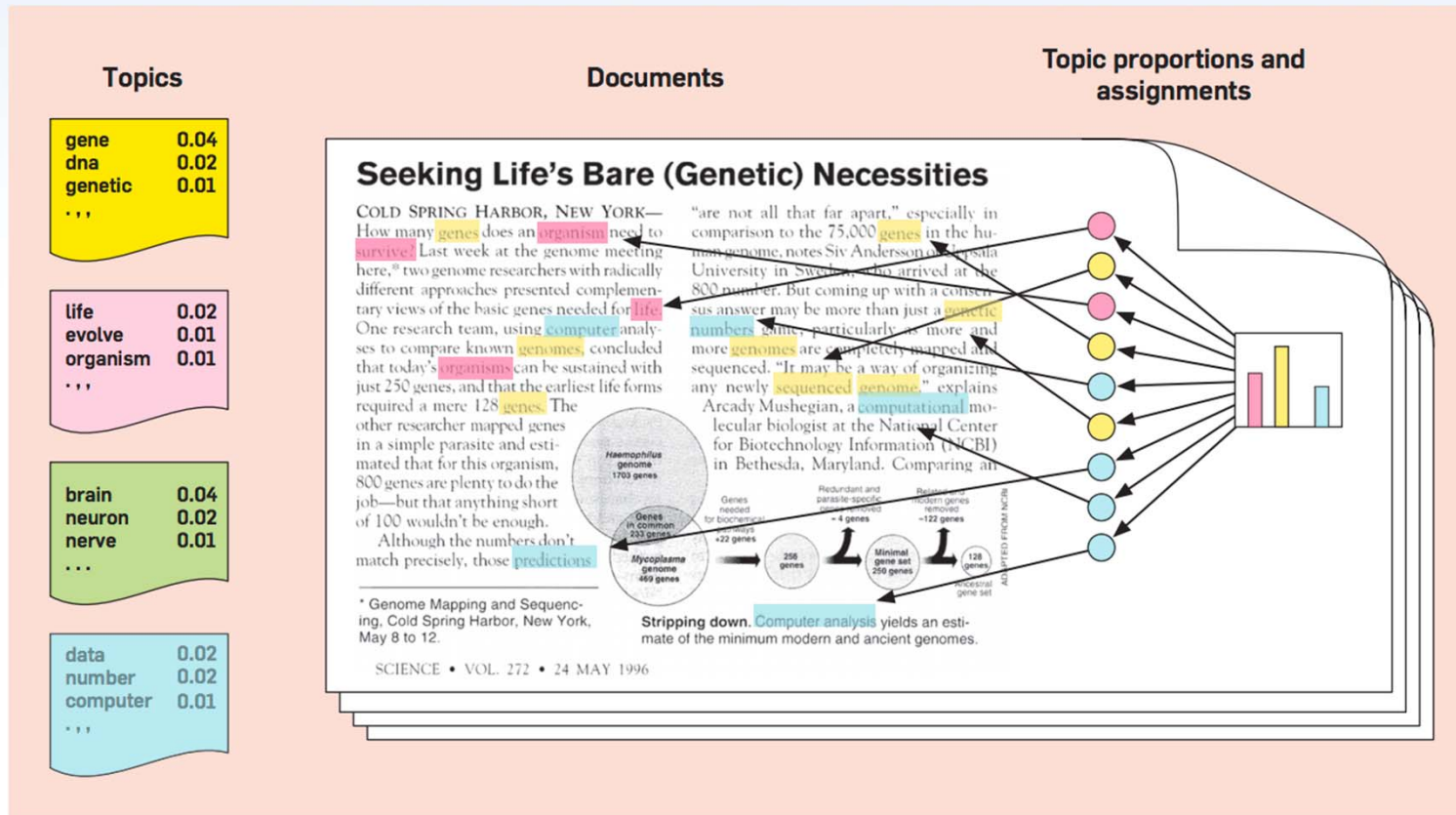
*Also provides “related information.”*

Not only the use case is provided but also the artifact that describes how that use case can be implemented.

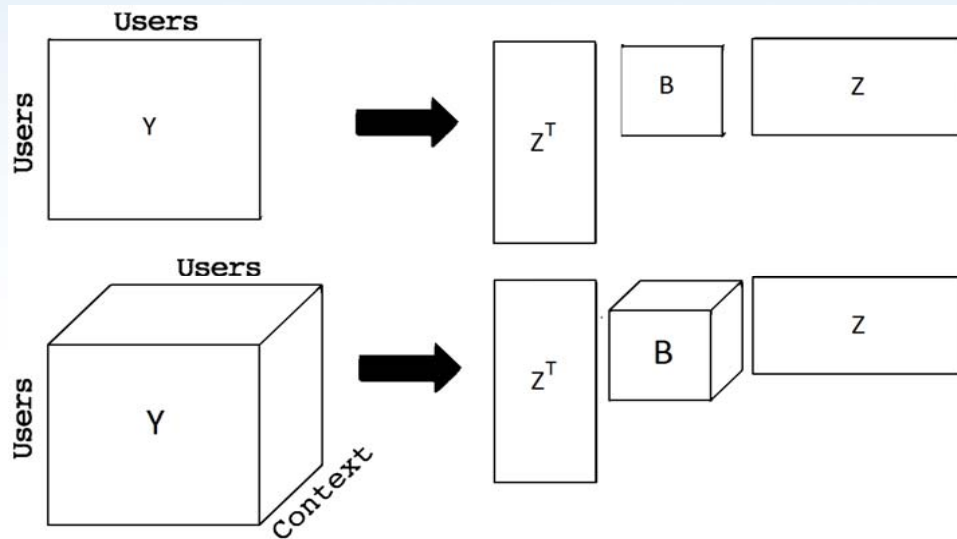


# Probabilistic Topic Model

Are there already existing similar solutions?



# Tensor Based Mixed Membership Stochastic Block Models



Unsupervised approach to identifying sub-community structure in heterogeneous networks.

Enables researchers to understand the effect of sub-community participation on communication.

# Tensor Based Mixed Membership Stochastic Block Models

*Models help to determine groups of people doing a specific kind of work.*

When someone struggles specifying a use case, a notification can be provided informing that person who might be able to help.

More actively, a notification can be provided to an expert that someone might need her help.





# Scraping and Visualization

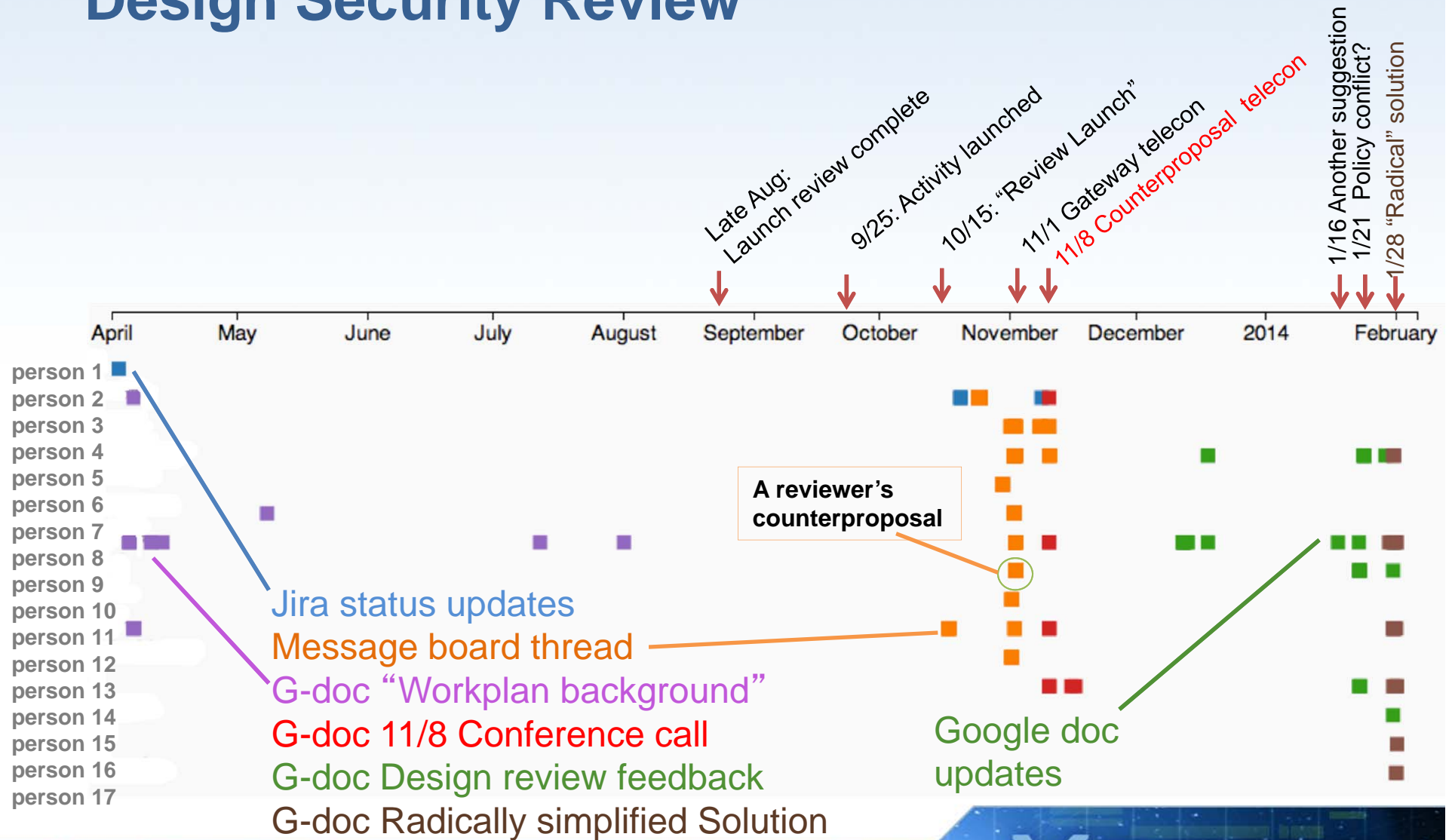
- We were told that “reviews are a problem.”
- We obtained four examples of design reviews that were “complex and protracted.”
- We plotted out activity over time and read through documents and discussions and visualized the results.



XSEDE



# SDIACT-63 Gateway User Count Design Security Review



# Documents referenced in Design Security Review

“SDIACT-63: Science gateway user count”

Key:

Public web site

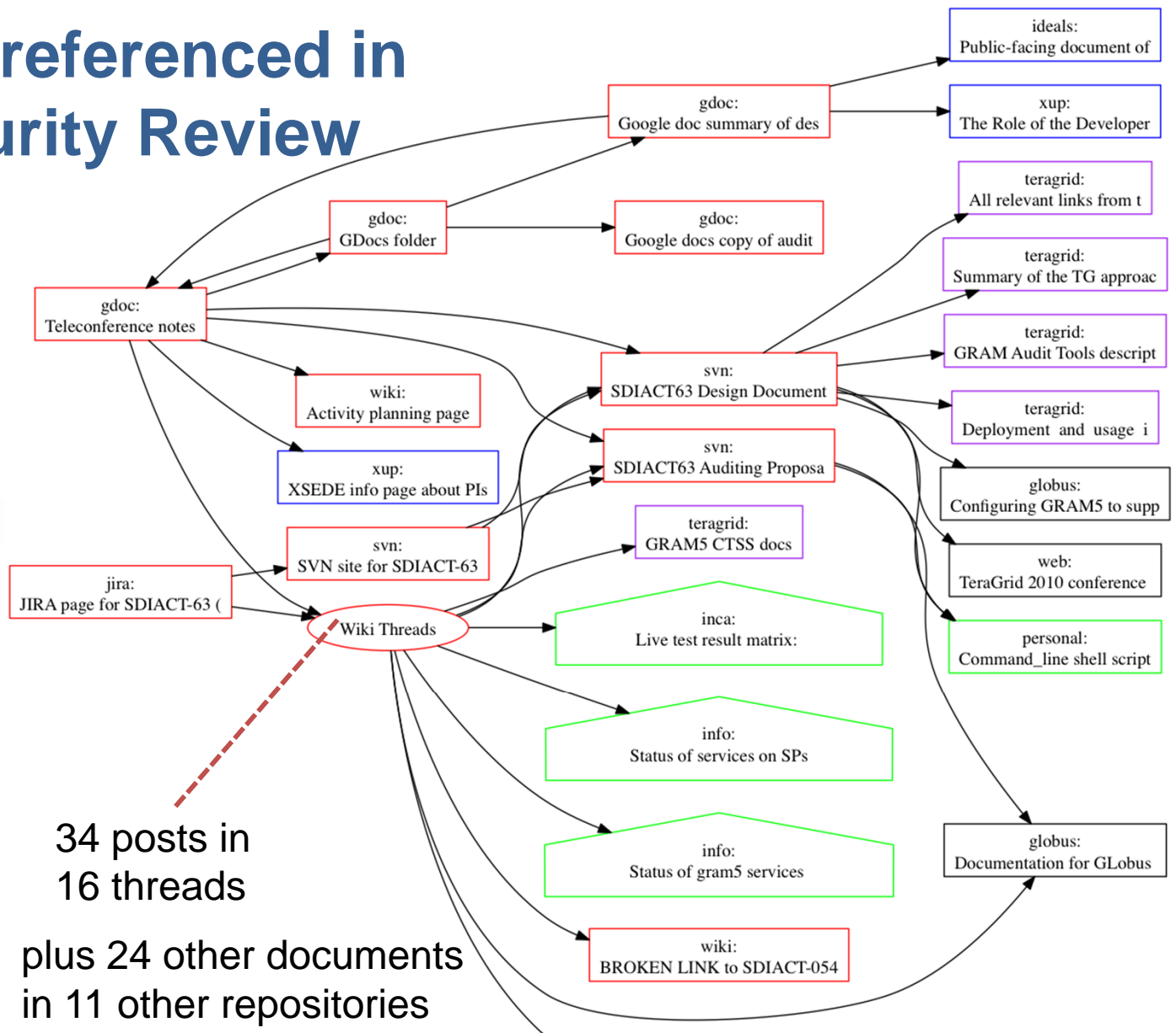
SD&I resource

OPs resource

Teragrid resource

Discussion thread

Real-time data



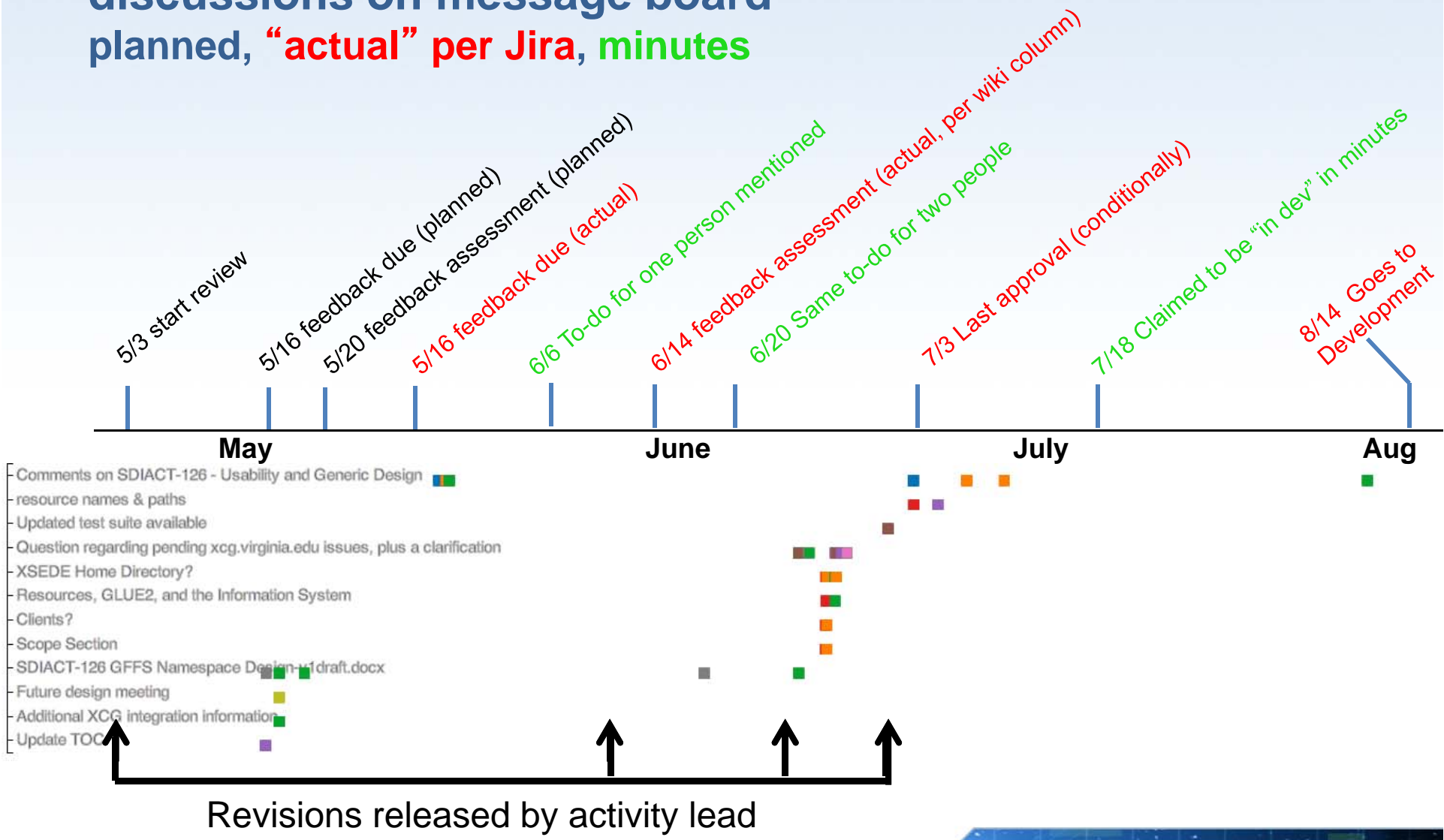
34 posts in  
16 threads

plus 24 other documents  
in 11 other repositories





# SDIACT-126 GFFS Namespace discussions on message board planned, **“actual” per Jira, minutes**



# Resultant SEI Recommendations

1. Involve relevant stakeholders earlier in the design process and make sure that the resultant discussion is reflected in a discussion thread (for transparency and later reference, especially during reviews).
2. Let reviews fail (if need be) with actions to be taken and potentially a smaller set of re-reviewers required for a subsequent review. Do not let reviews go on indefinitely while redesign is actually occurring.
3. Ensure that review dates and outcomes are tracked so that metrics about the review process and status can be easily collected.
4. SEI: provide visualization options for the review process.

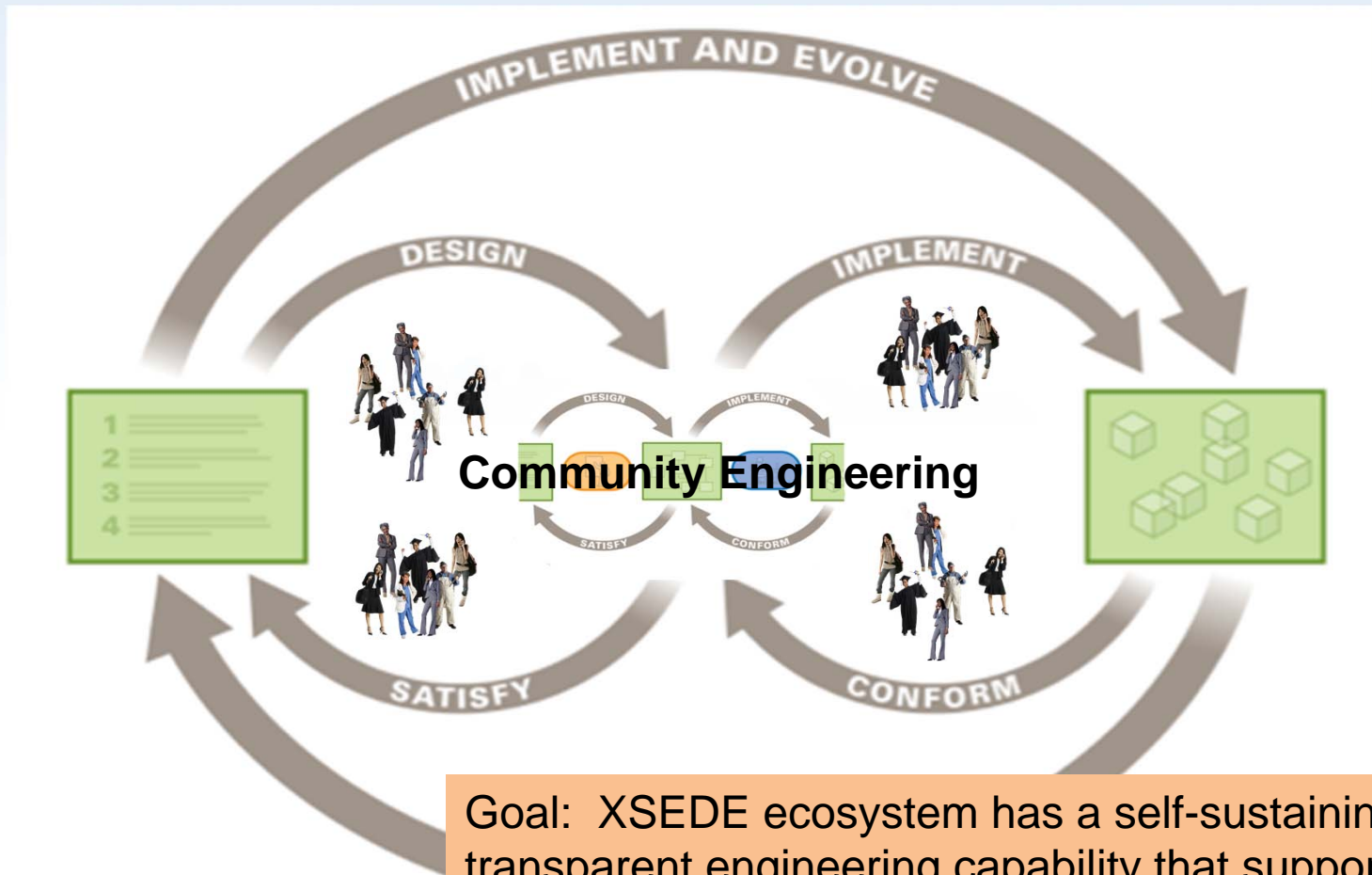


# Future Plans

- **Incrementally improve XSEDE STE**
  - Accessible visualizations
  - More effective search capabilities
  - Data to support XSEDE KPIs and metrics
- **Explore STE enabling technology**
  - Automated support for data-based improvements and transparency
  - Apply existing and new machine learning techniques (e.g., topic modeling and block modeling) to XSEDE repositories to aid in finding new relationships among internal web pages and documents for the purpose of further enhancing XSEDE search capabilities.



# Community Engineering – End State



Goal: XSEDE ecosystem has a self-sustaining , transparent engineering capability that supports ongoing transformational science.



# Question and Comments





Our reach will forever  
exceed our grasp, but,  
in stretching our horizon,  
we forever improve our world.

**XSEDE**

Extreme Science and Engineering  
Discovery Environment

# References

Blei, David M. "Probabilistic Topic Models." *Communications of the ACM* 55.4 (2012): 77-84.

Clements, Paul, Rick Kazman, and Mark Klein. *Evaluating Software Architectures*. Addison-Wesley, 2001.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2013). Leveraging Transparency. *IEEE Software*, 30(1), 37-43.

Howison, J., & Herbsleb, J. (2013). Incentives and Integration in Scientific Software Production, *ACM Conference on Computer-Supported Cooperative Work* (pp. 459-470). San Antonio, TX.

Northrop, Linda, et al. "Ultra-Large-Scale Systems-The Software Challenge of the Future." (2006).

