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’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

Scientific Socio-Technical Ecosystem
- Fostering scientific collaboration

Engineering Socio-Technical Ecosystem
Enabling continuous evolution

Research Test Bed
Adopting research results
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.
XSEDE Architecture Centric Engineering (ACE)

... in progress
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

Use Cases with Quality Attributes

Design Alternatives

Community Goals

Evaluation Results

Architectural Design

Architectural Documentation Artifacts
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.
## Small-Scale Development vs. Large-Scale

<table>
<thead>
<tr>
<th>Small Scale</th>
<th>Large Scale – Socio-Technical Ecosystems (STE) Challenges</th>
<th>XSEDE Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short daily meetings allow teams to share all information.</td>
<td>Requires creation of extensive documentation. Information is in many places.</td>
<td>Need to participate in many meetings. Presents logistical problems and is costly.</td>
</tr>
<tr>
<td>Can ask a team member at any time for missing information.</td>
<td>Difficult to find the right artifacts. Use of “social network” to find the right person.</td>
<td>Many documents and many places where information is stored. Many meetings to understand who knows what.</td>
</tr>
<tr>
<td>Quality assurance done by working in pairs or through short reviews (context is already known).</td>
<td>Quality assurance requires Input from multiple teams. Difficult to coordinate; difficult to surface local issues and understand global impact.</td>
<td>Reviews take a long time and decisions might be deferred until review is complete.</td>
</tr>
<tr>
<td>All team members are informed about changes.</td>
<td>Difficult to determine who needs change information and therefore usually broadcasted. Leads to information overload.</td>
<td>Trying to avoid change by lowering development risk.</td>
</tr>
</tbody>
</table>
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
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
• 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.

Generic Review process in XSEDE 10/31/2013 Draft
An XSEDE Visualization

- Static diagrams of flow are useful, dynamic diagrams can show bottlenecks.
- Interactive widget that shows use cases being turned into designs.
- Shows, for example, that highest priority use cases were indeed prioritized in practice.
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.
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.
Calendar of four reviews
Jira, Message board, Google doc

The squares are document-events: comments made in threads, document updates, Jira status changes.
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*

<table>
<thead>
<tr>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/3 start review</td>
<td>5/16 feedback due (planned)</td>
<td>6/6 To-do for one person mentioned</td>
<td>7/18 Claimed to be “in dev” in minutes</td>
</tr>
<tr>
<td>5/16 feedback due (planned)</td>
<td>6/14 feedback assessment (actual)</td>
<td>6/20 Same to-do for two people</td>
<td>8/14 Goes to Development</td>
</tr>
<tr>
<td>5/20 feedback assessment (planned)</td>
<td>7/3 Last approval (conditionally)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Revisions released by activity lead
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.
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.
References


