Loom Workflow Engine: Collaboration through portable, shareable data analysis

Nathan A. Hammond¹, Isaac Liao¹, Sowmi Utiramerur², Somalee Datta¹

¹Stanford Center for Genomics and Personalized Medicine
²Stanford Health Care

ACADEMIC RESEARCH must be REPRODUCIBLE

Think back to a computational result from 3 years ago

• How did you generate this result?
• What was the input data? Can you verify it?
• What were the software versions? Are you sure?
• Can you rerun the analysis and verify the result?
• Can someone else run the analysis without your help?

Why use a workflow framework?

• A workflow framework can ensure reproducibility and implement other common functions in a consistent way across all steps.
• This is easier, cleaner, and safer than depending on the pipeline developer to implement these functions for every step in every workflow.

Reproducibility & Traceability

To reproduce an analysis, you need to reassemble and verify:

• Input data
• Runtime environment
• Commands executed

Loom automatically keeps track of these for you.

Files are always identified by hash, and verified by default.

Runtime environment is saved using Docker and tracked by an immutable image id.

Portability architecture

With Loom, the same workflow you run on your laptop can be run in the cloud without modification.

• Loom abstracts platform-level services such as file storage, compute, and database operations. With simple adaptors, it can run on many different platforms.
• Loom’s client-server architecture allows it to scale from a single user running it on one desktop to many users sharing a remote Loom server.

With Loom, the same workflow you run on your laptop can be run in the cloud without modification.

<table>
<thead>
<tr>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford Center for Genomics and Personalized Medicine</td>
</tr>
<tr>
<td><a href="http://scgpm.stanford.edu/">http://scgpm.stanford.edu/</a></td>
</tr>
<tr>
<td><a href="https://github.com/StanfordBioinformatics/loom">https://github.com/StanfordBioinformatics/loom</a></td>
</tr>
<tr>
<td><a href="https://pypi.python.org/pypi/loomengine">https://pypi.python.org/pypi/loomengine</a></td>
</tr>
</tbody>
</table>

Special thanks to our partners

• Palo Alto VA Hospital
  • Phil Tsao
  • Cuiping Pan
• Stanford Clinical Genomics Service
  • Stanford Health Care
  • Lucile Packard Children’s Hospital

Contact

Nathan Hammond - nhammond@stanford.edu
Isaac Liao – iliao@stanford.edu
info@loomengine.org

Getting started

Loom is still in pre-release but we invite you to check it out and let us know what you think!

You can find us on github: https://github.com/StanfordBioinformatics/loom