ECP Community BOF Days

SAND2021-3866C

March 30 – April 1, 2021

Hosts: Ashley Barker (ORNL) and Osni Marques (LBNL)

Enabling Developer Productivity with Software Process Improvement

Elaine Raybourn (SNL)
Reed Milewicz (SNL)
Greg Watson (ORNL)
Elsa Gonsiorowski (LLNL)
David Rogers (ORNL)
Ben Sims (LANL)
Jim Willenbring (SNL)

11:00 AM - 12:30 PM, March 31st, 2021

Plan for Today

1. Look at what software process improvement is and how it matters.
2. Describe the work of IDEAS-ECP in helping ECP teams improve their software projects, and the tools and technologies we are developing to support this effort.
3. Open the floor to general discussion on how teams make progress in sustainable ways.

Note: We'll be using **slido** to gather feedback From the audience!

SLIDO SESSION NUMBER

While We Introduce Our Panelists, A Question for the Audience

Bottom Line Up Front

- Developed by IDEAS-ECP, **Productivity and Sustainability Improvement Planning** (PSIP) is a lightweight workflow for software process improvement we employ with ECP teams to help improve their projects.
- Today…
  - You'll learn about **software process improvement** and how your colleagues approach it.
  - You'll learn about PSIP, who we are, and what we're doing as a team in terms of **automating methodologies** and using a **data-driven approach** to software process improvement.

Enabling Software Quality

A Brief History of Software Process Improvement

First Wave

- Classic Approaches : CMM(I), SPICE, non-software methods translated to software such as ISO9000 or Six Sigma
- Comprehensive standard and certification driven assessments
- In the start, companies were started, and they were failing and certifications were important

Second Wave

- Incorporate lean and agile thinking in their methods
- Iterative improvement and continuous learning and adaptation
- Less focus on certification; more on how to improve methods

Software Process
Is
Software Too!

Characterizing the State of Scientific Software

Studies have found that that researchers are **spending more time** developing software than in the past.

Scientific software projects are increasingly **collaborative**, **multi-disciplinary**, and **multi-institutional**.

The **demand** for and **consequences** of scientific software are **greater than ever**.

There is growing interest in **improving software processes** to manage this scope, scale, and complexity.

Software Process Improvement Can Be Daunting

Adopting a new tool, technique, or process can be **disruptive** and can carry upfront **costs**.

Software process improvement is seen as too **formal**, **heavy-weight**, and **not aligned** with the needs of scientific software projects.

Any productivity or sustainability improvements must be incremental, integrated into the primary feature development process, lightweight, and iterative.

Question for the Audience

Productivity and Sustainability Improvement Planning (PSIP)

- PSIP provides tools and resources to set, measure, and realize improvement goals.
- Key to the success of PSIP has been to develop automated, data-driven approaches to understanding how teams work and developing tools that can meet their needs with minimal overhead.

Enabling Software Quality

IDEAS productivity

Who is using PSIP?

Improvements to documentation to create reference manual, setting code style standards, transition to GitHub

Completed PSIP tutorial, investigating how it can be used in academic context

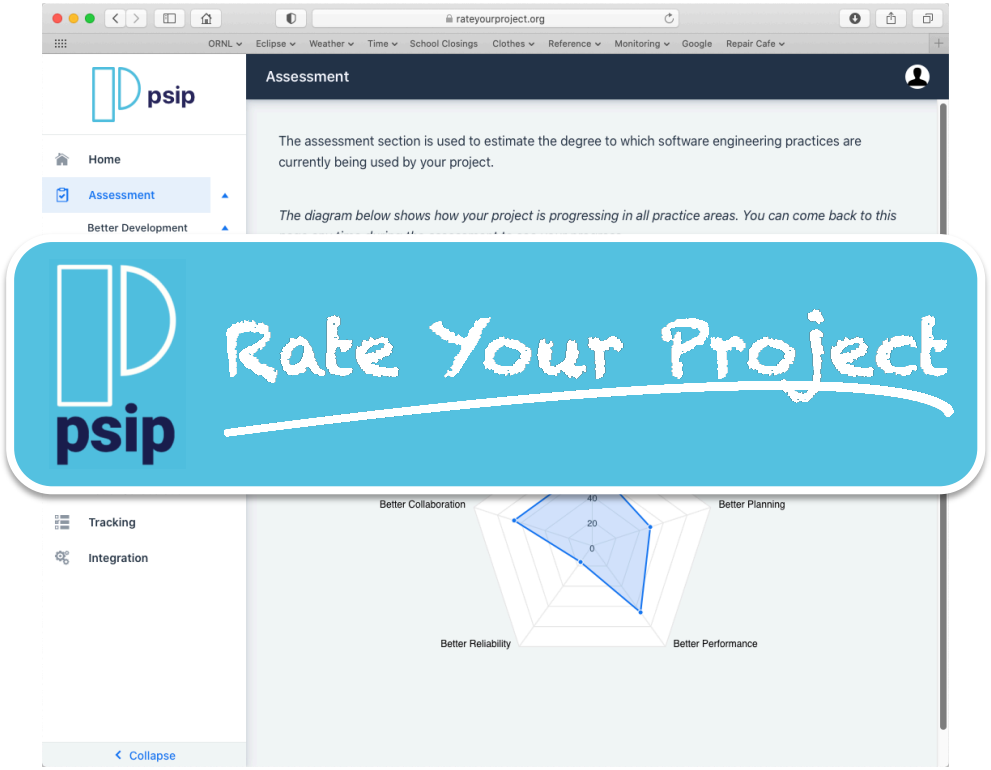
ALPINE/ZFP

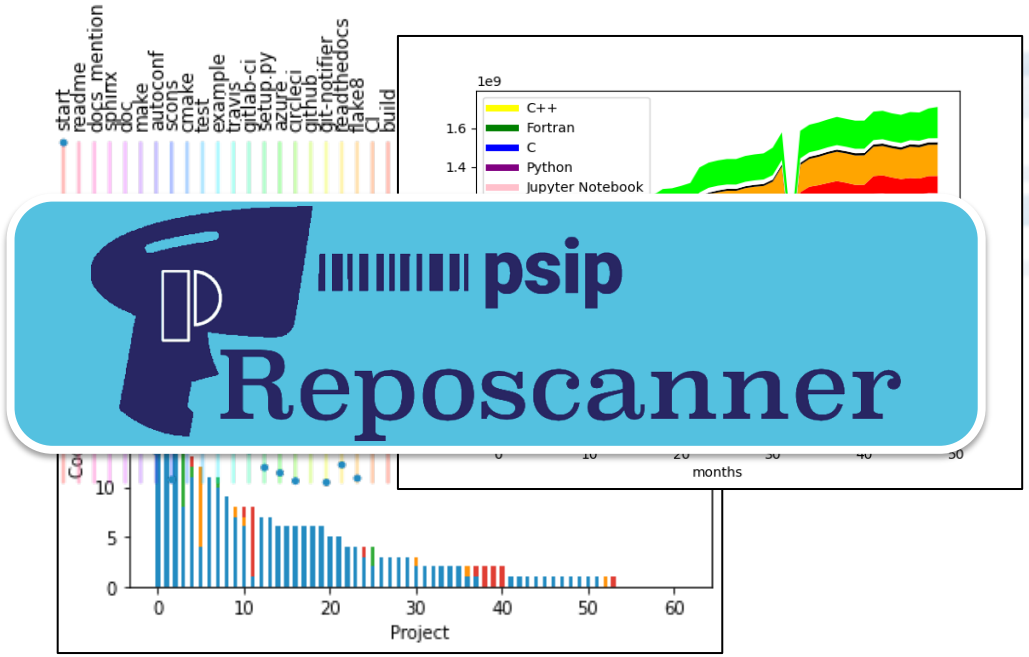
Create a VTK-m filter for APLINE in situ algorithm users

Using a more detailed version for internal project assessment

Using internally for updating version control systems, updating documentation to support better onboarding

How The PSIP Team is Using Automation and Data-Driven Analysis to Deliver Value to ECP Teams



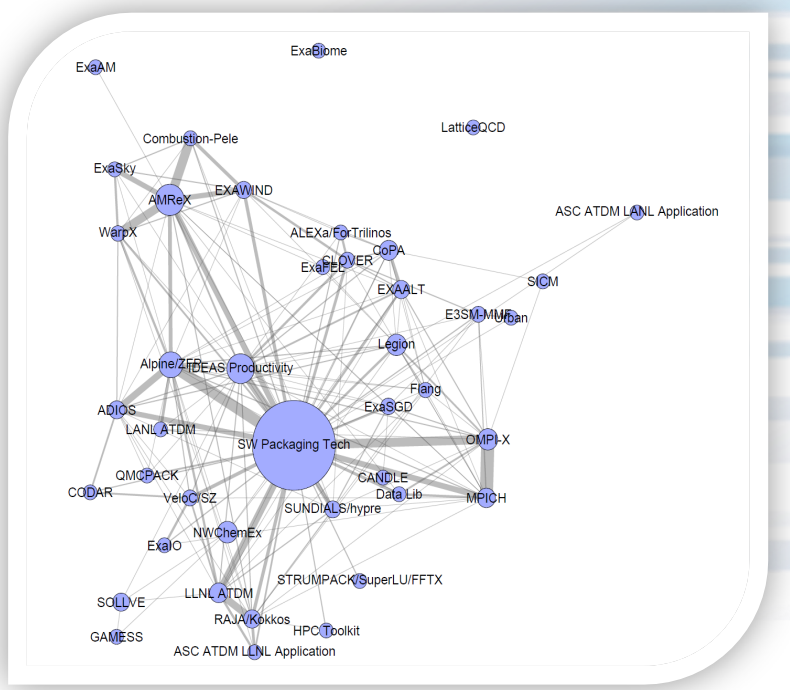


- Our team is actively working on several tools and technologies for software process improvement that your team can start using now to
  - Realize process improvements **without disrupting current development.**
  - Mitigate **technical risk** so that you can develop software with *confidence.*


01 Summarize Current Project Practices
02 Set Goals
03 Construct Progress Tracking Cards (PTCs)
04 Record Current PTC Values
05 Create Plan For Increasing PTC values
06 Execute Plan
07 Assess Progress
Productivity and Sustainability Improvement Planning (PSIP) Workflow

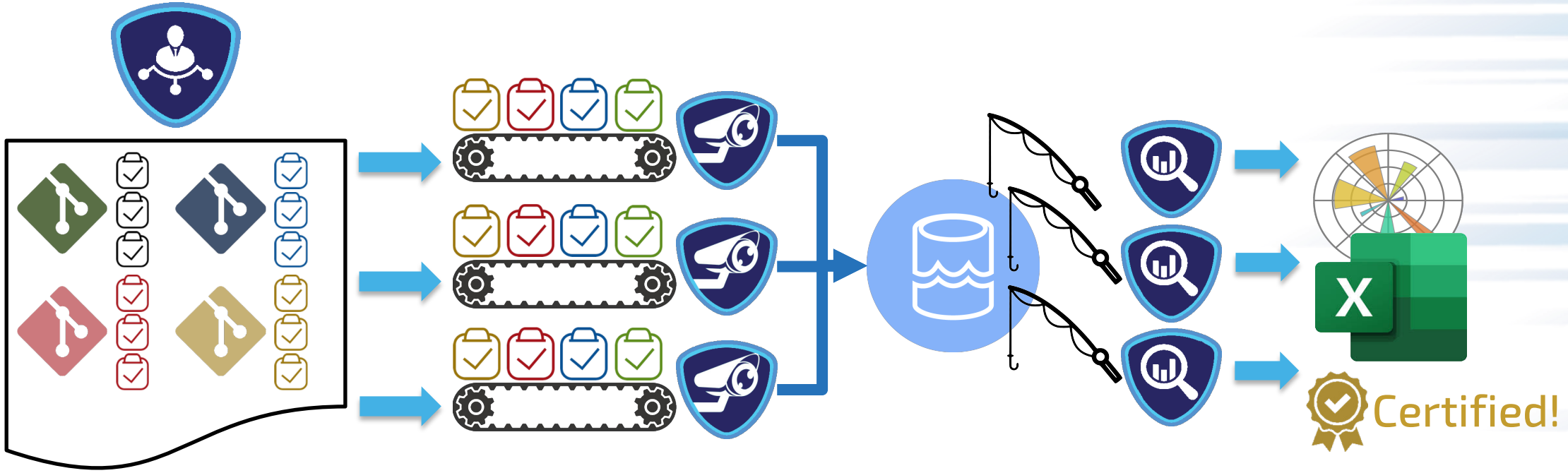
Summarizing Project Practices with Reposcanner

- To match the scale of the ECP and keep pace with its evolution, the PSIP team must leverage solutions for **automated, scalable data collection and analysis**.
- By studying ECP development teams and their practices at scale, the PSIP team can draw general conclusions about the **observable behaviors and characteristics** that contribute to a team's ability to quickly reach the forefront of HPC.
- Ideally, this should be packaged as a tool **that could assist ECP teams** in summarizing own their state of practice.
- These were the driving factors behind the creation of **Reposcanner**, a novel software repository mining toolkit developed by the PSIP team to aid in characterizing collaboration and best practices across ECP.



psip Reposcanner

Reposcanner: Modular, Extensible Architecture



(1) For each repository and each routine that could be applied to it, the orchestrator generates a **task**. Likewise, the orchestrator creates a single task for each analysis downstream.

(2) Routines pick up tasks and carry them out, collecting data from repositories. They output files to disk, and store references to these files as **data entities** held in a communal data store.

(3) Analyses fish out data they need to compute their results, outputting new data files and human-readable graphs and reports. The orchestrator also outputs a provenance log to aid curation.

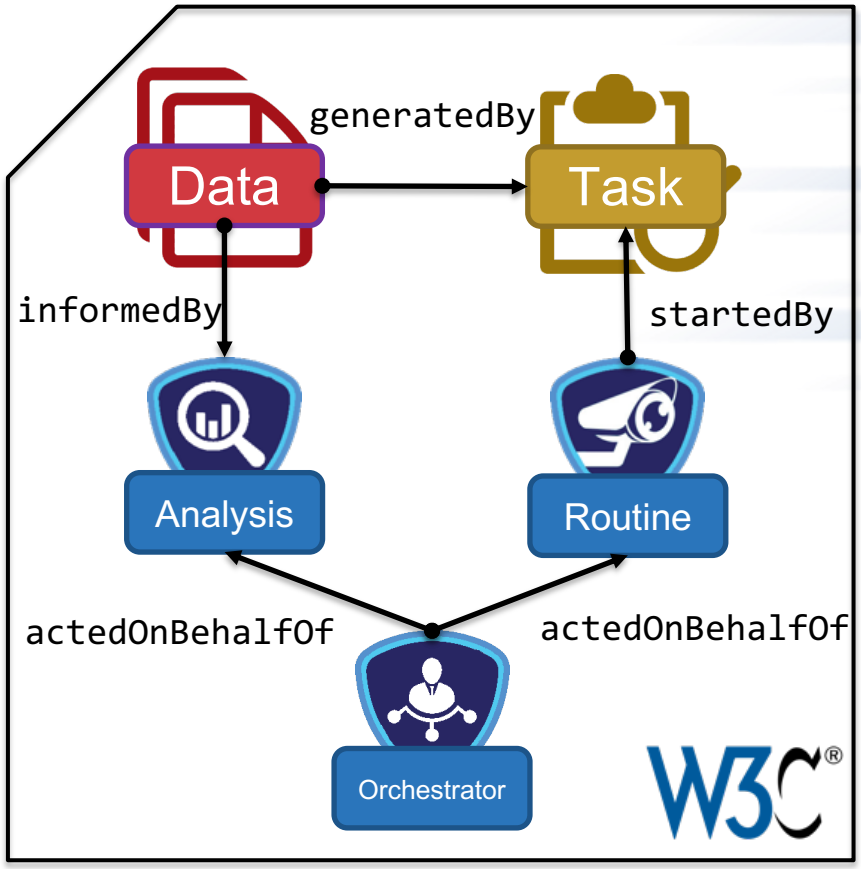
Reposcanner: A Feature-Rich Platform for Analytics

GitHub

GitLab

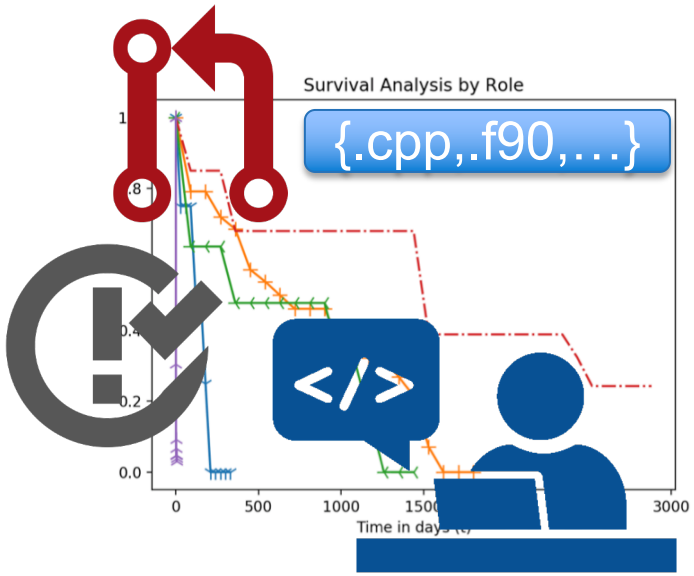
Bitbucket

Content-aware parsing of URLs, intelligent **credential management**, and **seamless support** for different platform APIs enable Reposcanner to meet breadth and depth requirements for analytics.

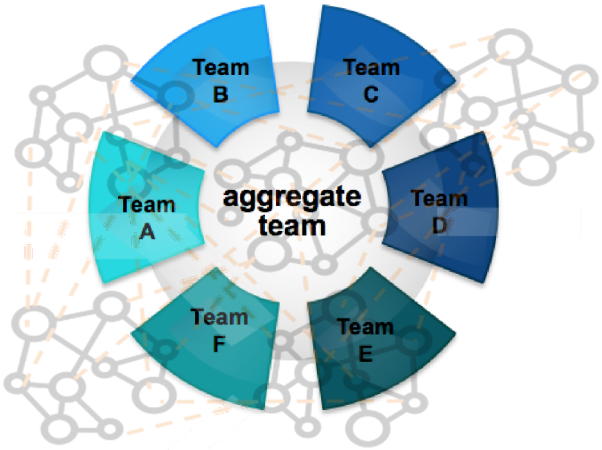


W3C-compliant provenance model helps **preserve the full chain of evidence** from data collection to interpretation and decision-making.

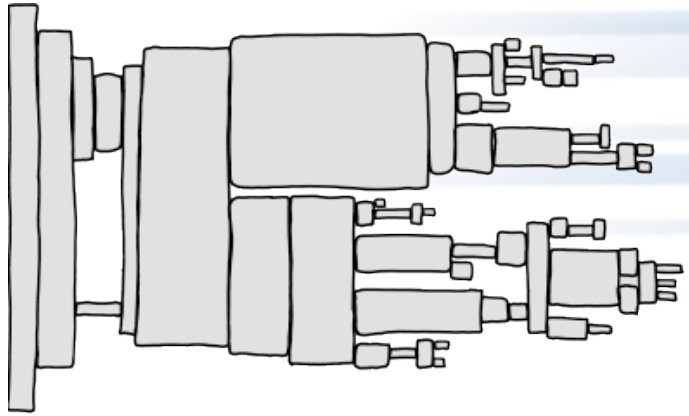
Reposcanner: Supports Team Decision-Making and Research Objectives



Quantitative and qualitative metrics of individual developer activity



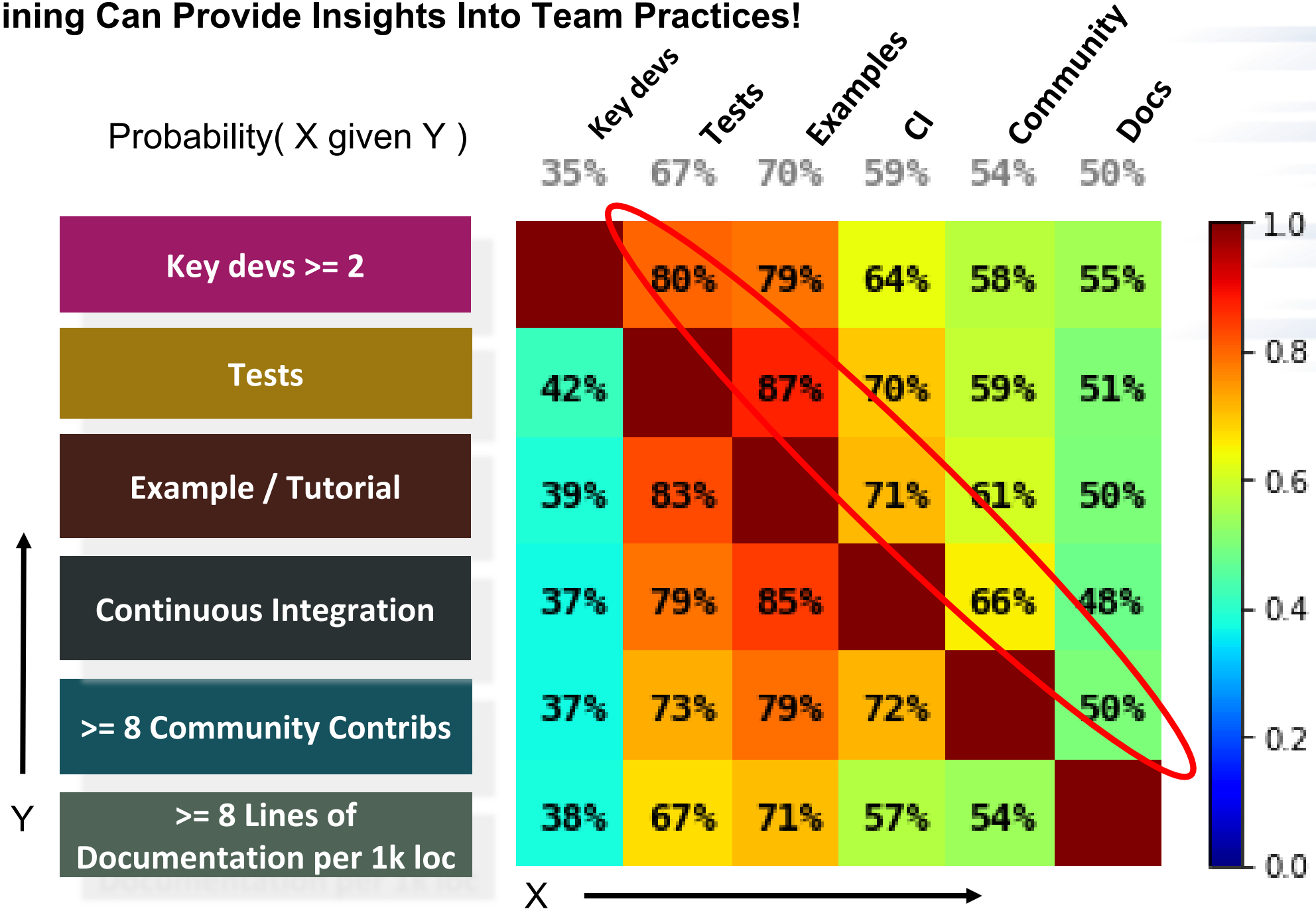
Developer networks and patterns of collaboration

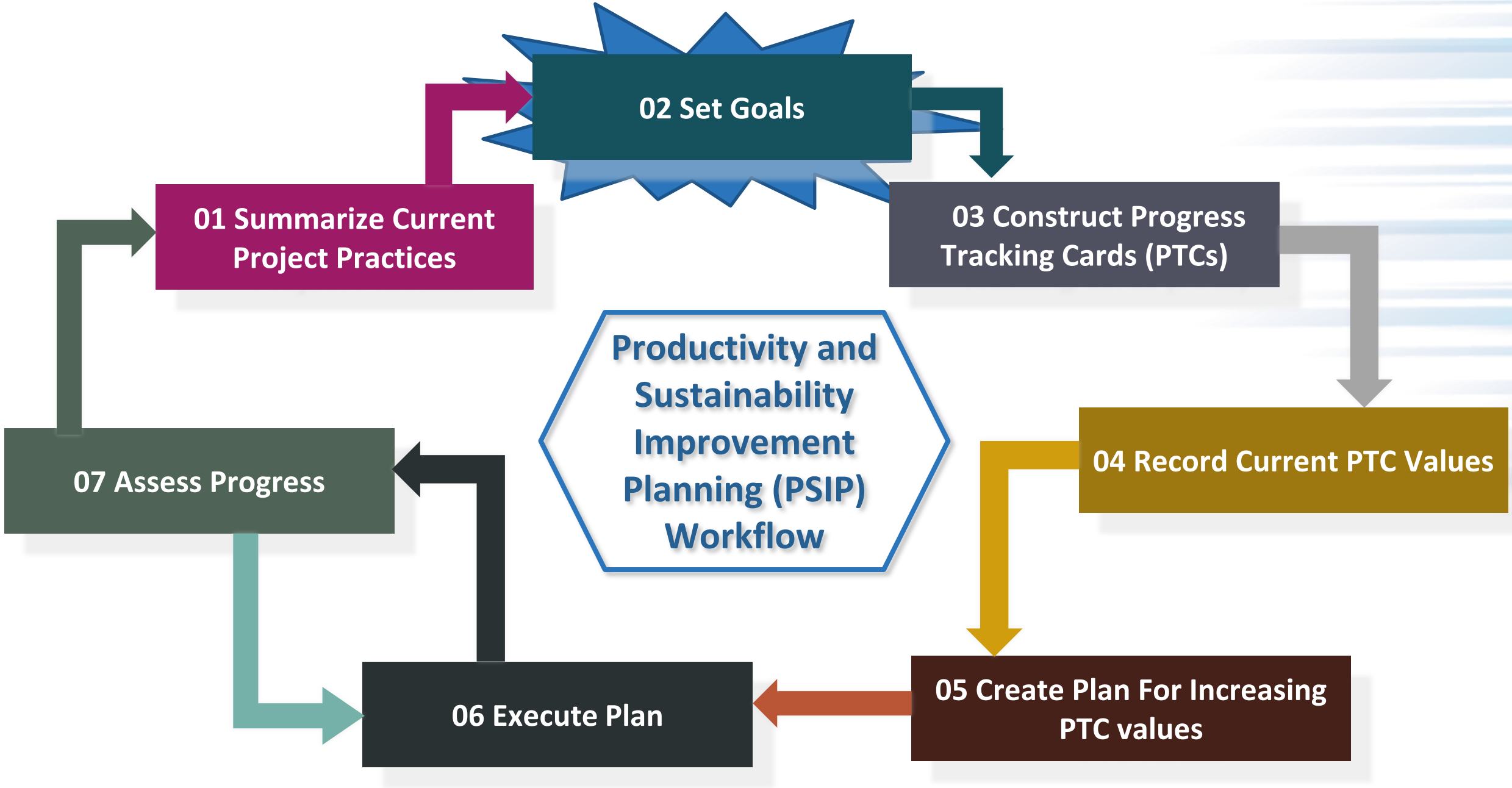


Software co-evolution across dependency "supply chains"

Scope

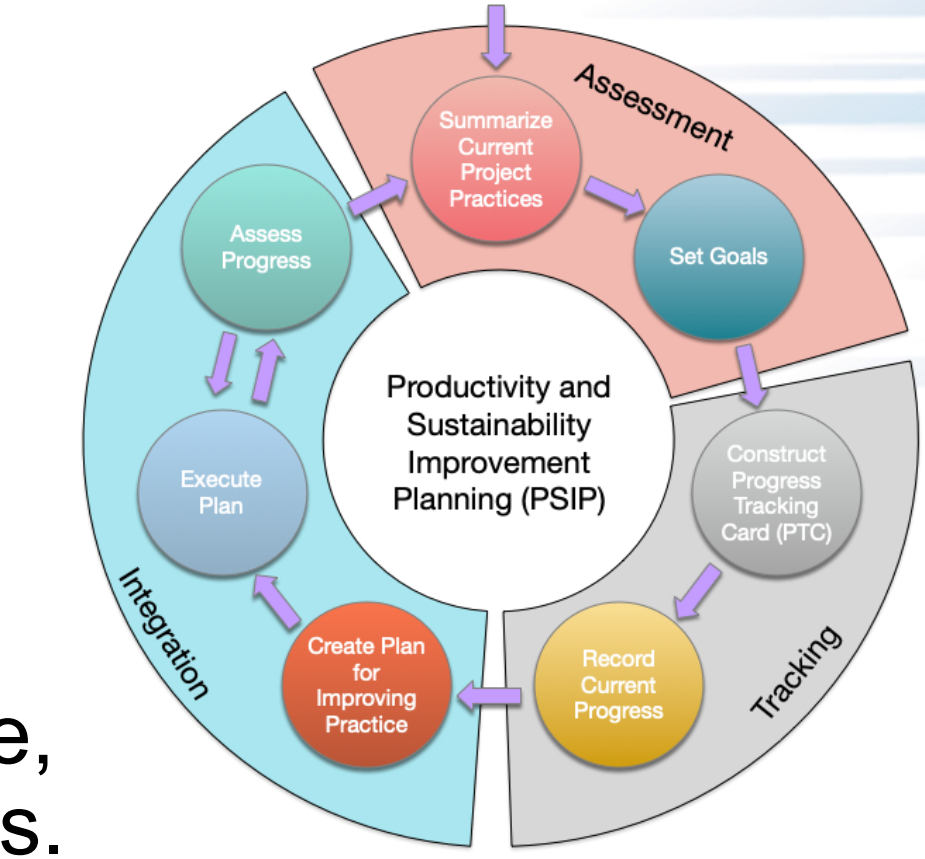
Data Mining Can Provide Insights Into Team Practices!




02 Set Goals
01 Summarize Current Project Practices
03 Construct Progress Tracking Cards (PTCs)
Productivity and Sustainability Improvement Planning (PSIP) Workflow
07 Assess Progress
04 Record Current PTC Values
06 Execute Plan
05 Create Plan For Increasing PTC values

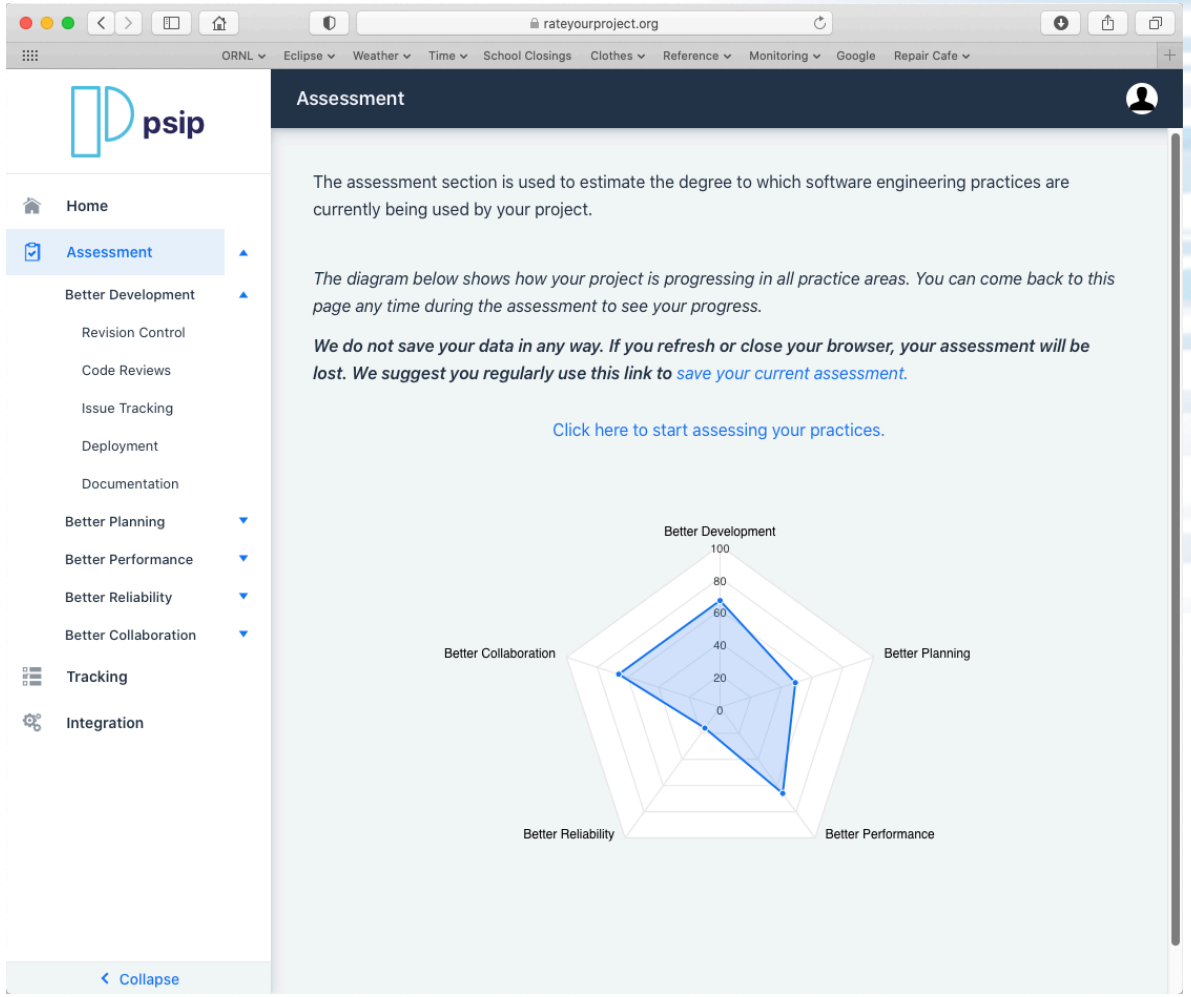
What is RateYourProject?

- In the past, assessment of team practices to identify improvement goals was a labor intensive practice for both the teams and their facilitators.
- What we needed: a guided self-assessment that enables the examination of software development, planning, performance, reliability, and collaboration practices.

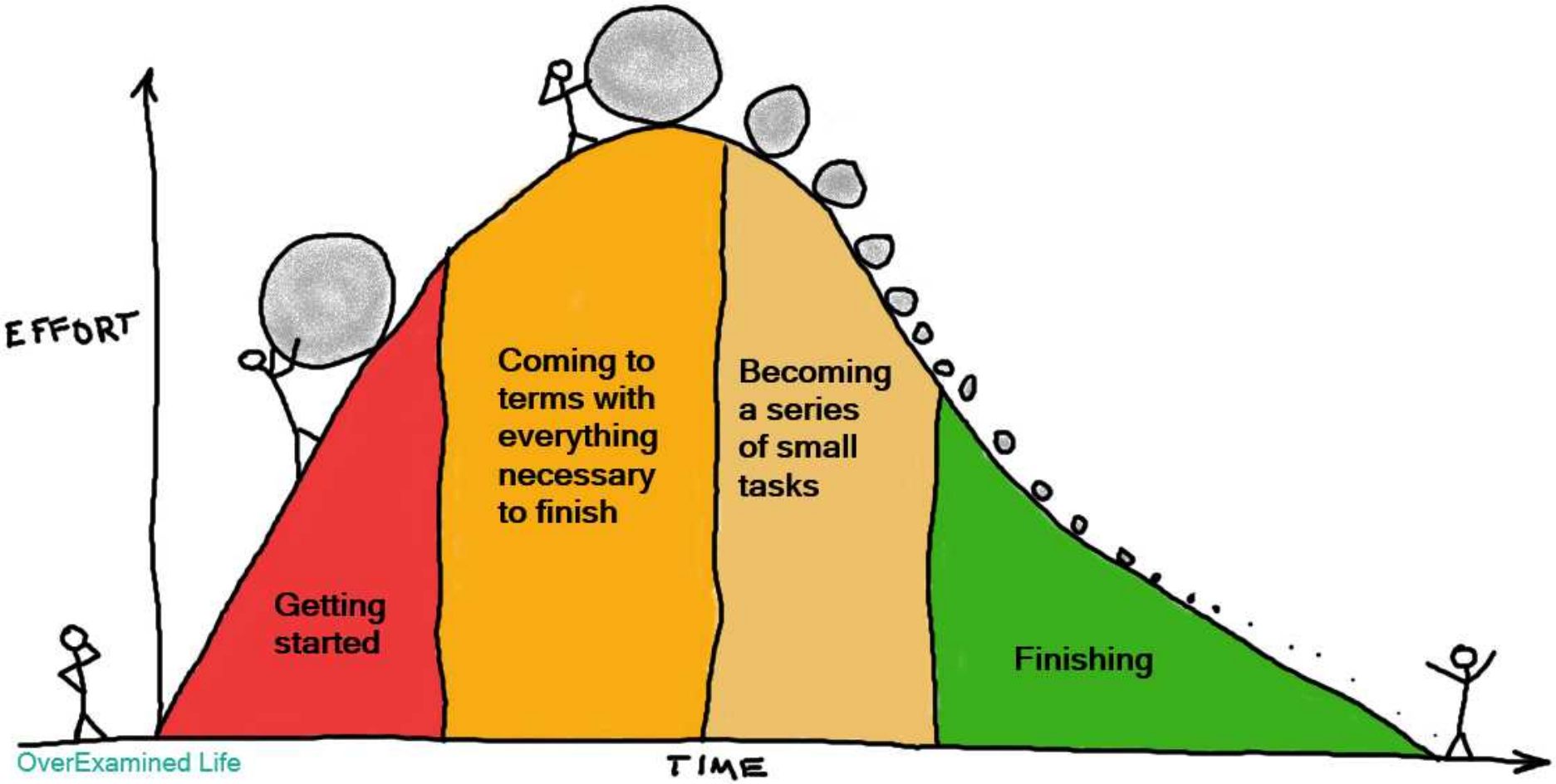


RateYourProject Provides Resources for Setting Improvement Goals

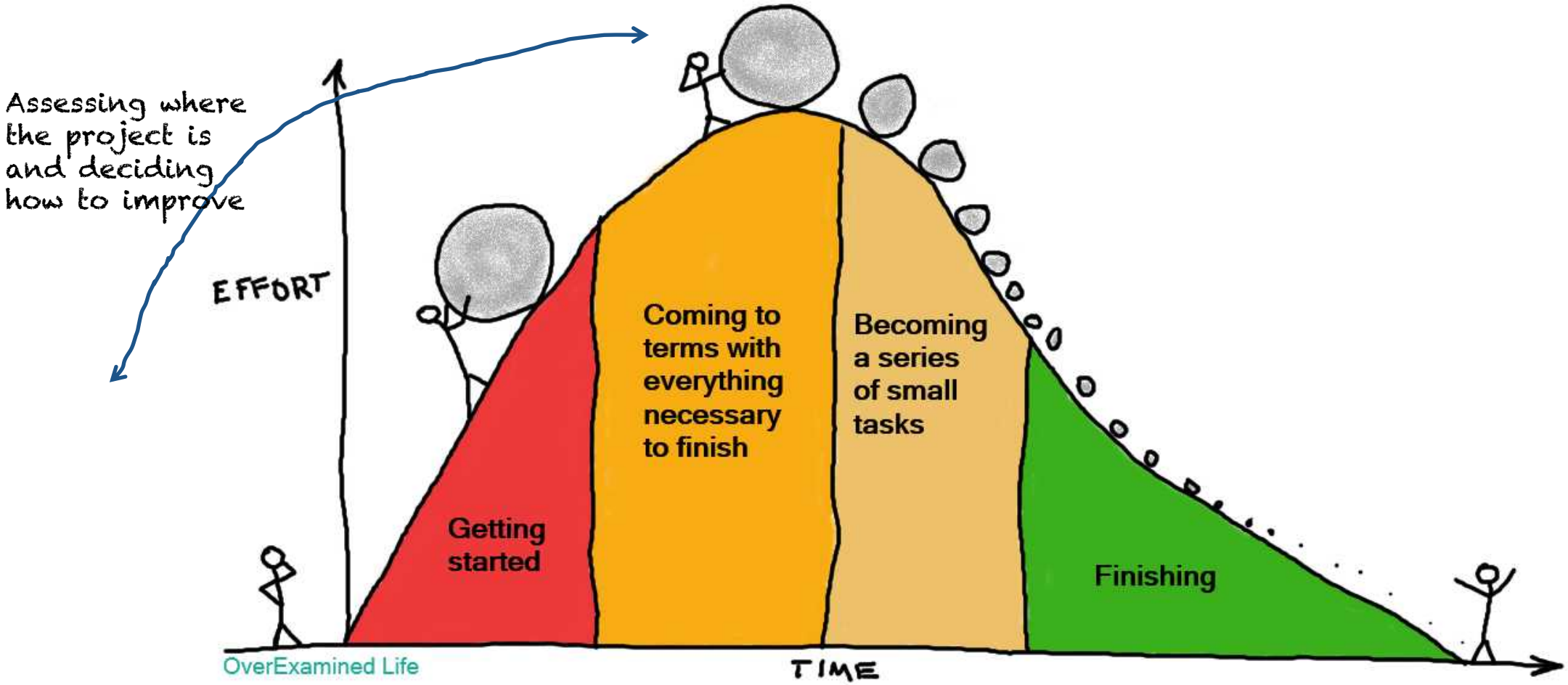
- RateYourProject (https://rateyourproject.org) aims to automate phases of the PSIP process, from self-assessment to PTC creation to integration of PTCs into a project.
- Practices are rated using a modified four-point Likert scale, which forces one of four responses (none, basic, intermediate, advanced) and no neutral response. Scores are aggregated in each practice area, and then used to generate a visual indication of overall progress using a spider chart.



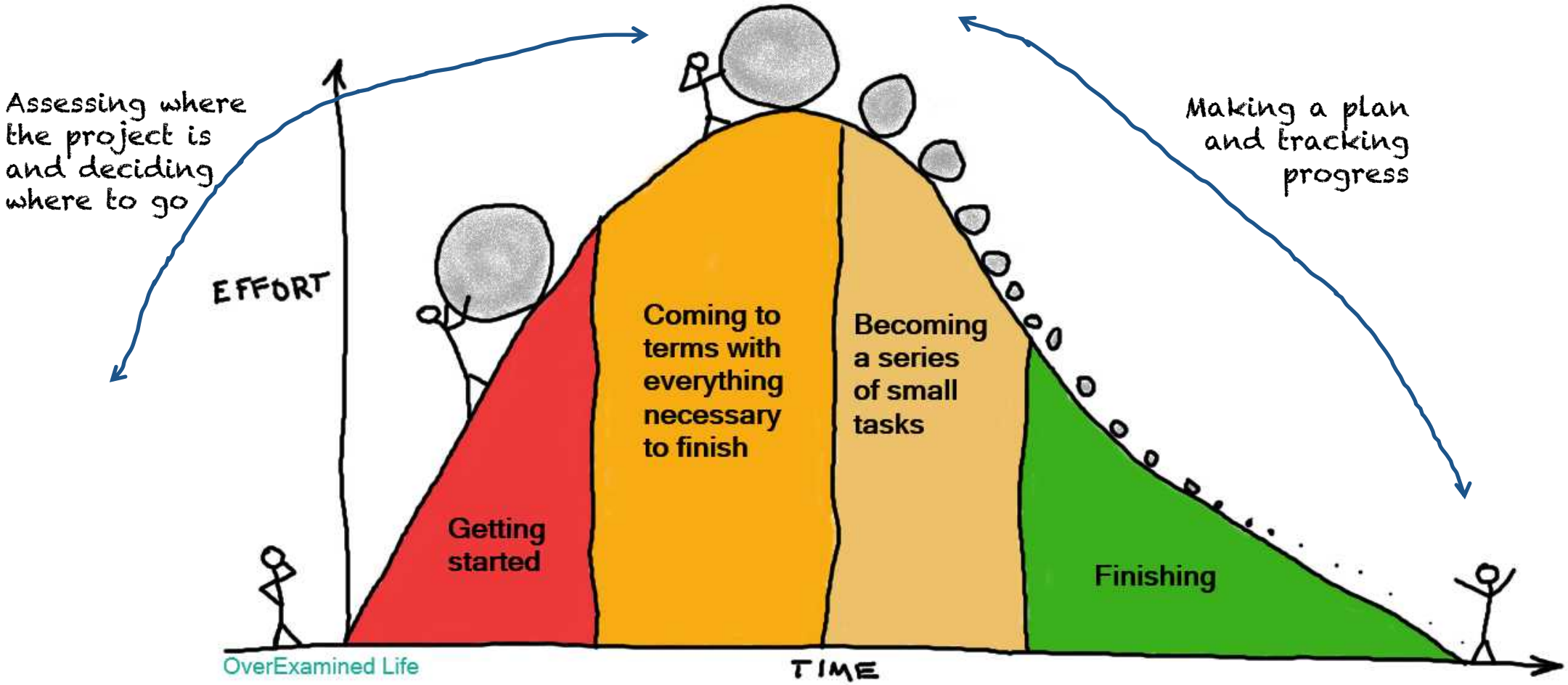
How to overcome the barrier of getting started?



How to overcome the barrier of getting started?



How to overcome the barrier of getting started?




02 Set Goals
01 Summarize Current Project Practices
03 Construct Progress Tracking Cards (PTCs)
Productivity and Sustainability Improvement Planning (PSIP) Workflow
04 Record Current PTC Values
07 Assess Progress
06 Execute Plan
05 Create Plan For Increasing PTC values

Teams Implement PSIP using Progress Tracking Cards (PTCs)

Title	The topic of the card
Target	Practice is changed
User Story	As a ___, I want to ___, so that ___.

Card

Score	Description
0	Initial State
1	Intermediate state of practice (+)
2	Intermediate state of practice (++)
3	Intermediate state of practice (+++)
4	Desired state of practice

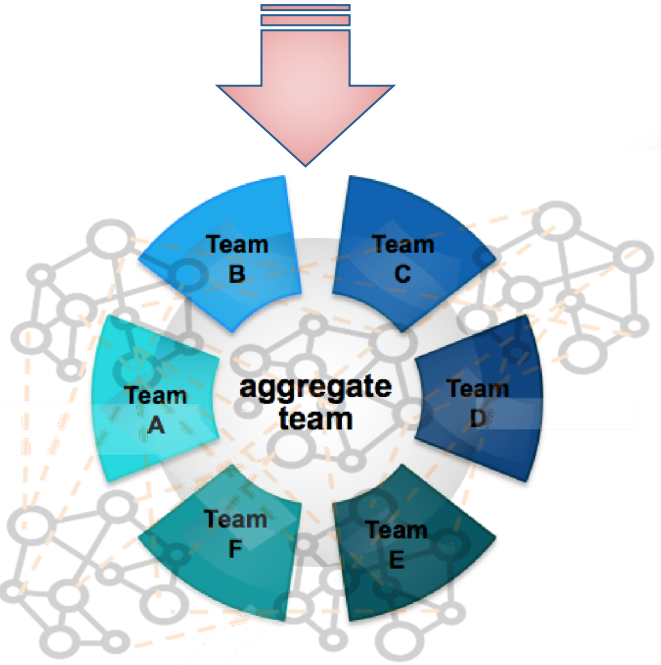
Comments: Relevant links or details

Check out https://bssw.io/psip for links to our PTC catalog and other helpful resources!

What Makes for a Good Process Improvement Goal?

1. **Collaborative** and **team-oriented** (think project policies).
2. **Measurable** and **specific** (think artifacts).
3. **Realistic** and able to be realized in **increments** (over weeks, months)

PSIP helps software teams to **IDENTIFY** opportunities to iteratively and incrementally **IMPROVE** software team practices and processes.



Why PSIP?

- Lightweight
- Easy to Adopt
- Encourages Reflection and Communication
- Can Reveal Differences in Assumptions and Priorities
- Can Help Team Converge on a Consensus

Summary: What is PSIP?

- Software process improvement can carry **upfront costs** and introduce **risk** and **uncertainty** into your project, and should be approached carefully and intentionally.
- The **Productivity and Sustainability Improvement Planning** (PSIP) is a lightweight workflow for software process improvement.
- PSIP is implemented via **Progress Tracking Cards** (PTCs).

Enabling Software Quality


psip
Breakout Sessions


psip
Regroup

Recap: What Have We Covered Today?

- Developed by IDEAS-ECP, **Productivity and Sustainability Improvement Planning** (PSIP) is a lightweight workflow for software process improvement we employ with ECP teams to help improve their projects.
- Today…
  - We talked about about **software process improvement** and how your colleagues approach it.
  - You learned PSIP, and what the PSIP team is doing as a team in terms of **automating methodologies** and using a **data-driven approach** to software process improvement.

Enabling Software Quality

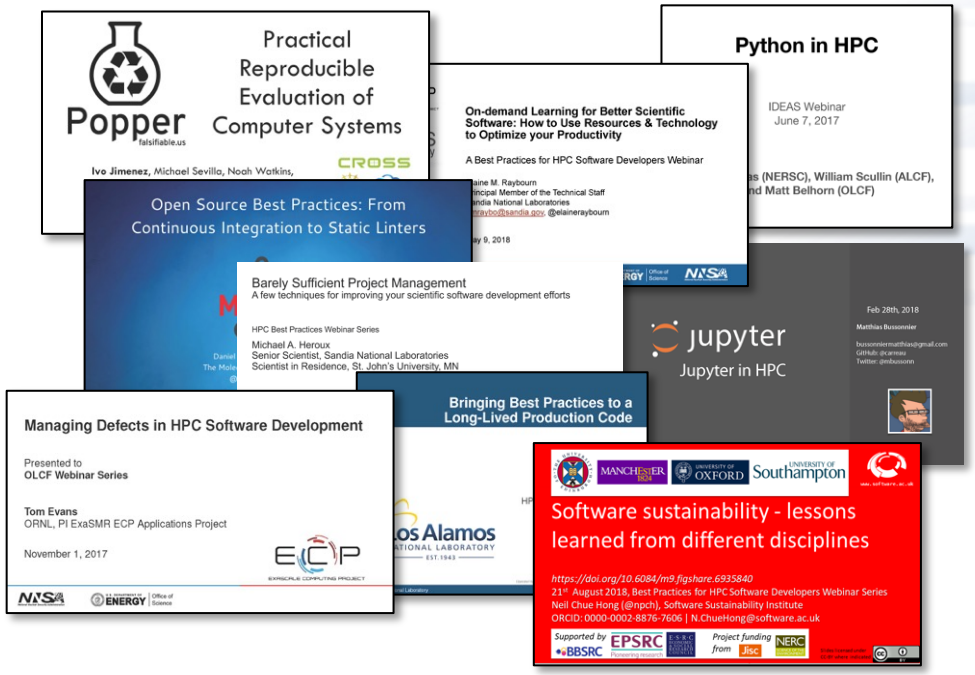
ADDITIONAL RESOURCES…

Additional Resources: Better Scientific Software Tutorials

- Full-day and half-day variants
  - Hands-on in full-day
- Recent venues
  - ATPESC (2016-2020)
  - ECP Annual Meeting (2017-2021)
  - ISC (2017-2019)
  - SIAM CSE17
  - Supercomputing (2016-2021)
- **Slides (and some recordings) of past tutorials**
  - https://ideas-productivity.org/events/
- Current tutorial modules
  - Motivation and Overview of Best Practices in HPC Software Development, Agile Methodologies, Git Workflows,
    Software Design, Software Testing, Refactoring,
    Continuous Integration, Reproducibility, Software Licensing

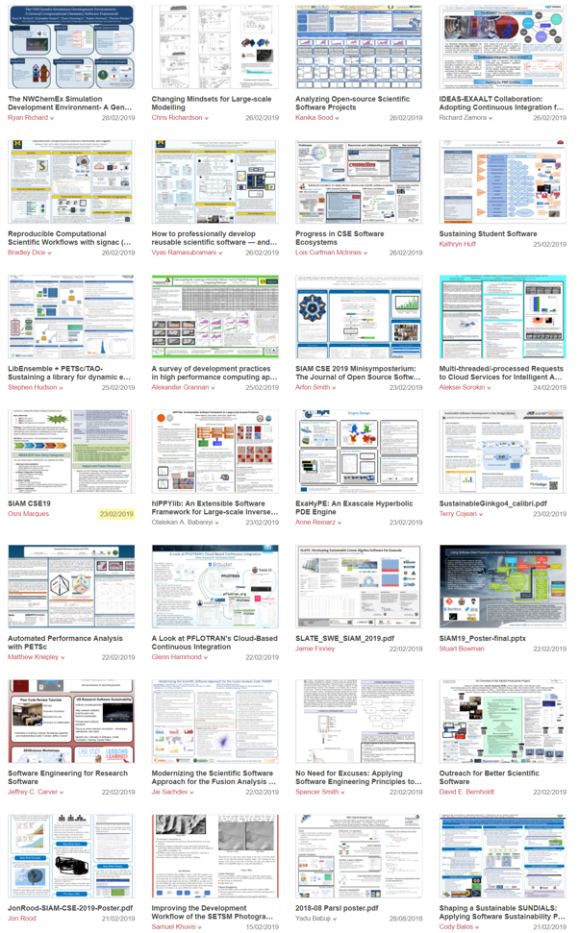
Additional Resources: Best Practices for HPC Software Developers Webinar Series (HPC-BP)

- Monthly series, since May 2016
  - Traditionally 1-2pm ET on a Wednesday
  - Which Wednesday varies
- Offered live and archived
- Presented by the community to the community
  - Not just IDEAS
- 45 webinars to date
  - 77 per webinar on average
  - 3214 attendees total, to date
- **Series info, archives, and mailing list for announcements**
  - https://ideas-productivity.org/events/hpc-best-practices-webinars/



Additional Resources: Technical Meetings and Birds of a Feather Sessions

- We help create opportunities to talk about software development, productivity, and sustainability in more traditional “academic” settings
  - https://ideas-productivity.org/events/
- Minisymposia
  - SIAM Computational Science and Engineering (2015, 2017, 2019, 2021)
  - PASC (2018, 2019)
- Thematic poster sessions
  - SIAM CSE (2017, 2019, 2021)
- Birds of a Feather sessions
  - Software Engineering and Reuse for Computational Science and Engineering
    - SC15, SC16, SC17, SC18, ISC19, SC19, SC20 see http://bit.ly/swe-cse-bof



SIAM CSE19 Miniposterium posters, archived on FigShare

License

- This presentation is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Acknowledgements

- This work was supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR), and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.
- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. Images used by permission.
- This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC.