Reproducibility as Side-Effect

Shu Wang*, Zhuo Zhen*, Jason Anderson* , Kate Keahey*§

* University of Chicago, § Argonne National Laboratory

The ability to keep records and reproduce experiments is a critical element of the scientific method. However, recording and publishing of research artifacts that allow to reproduce and directly compare against existing research continue to be a challenge. We propose the experiment précis framework that automatically records experiment management actions providing a reliable record of the experiment. Using this framework, we implement a prototype tool *ReGen* which generates experiment scripts that can be used to re-enact the experiment or shared with others.

The Scientist's Dilemma

Current Status for Computational Experiment
[Victoria Stodden et al. Enabling the Verification of Computational Results: An Empirical Evaluation of Computational Reproducibility. P-RECS, 2018]

No discussion & no artifacts made available	58.8%
Some discussion of artifacts, none made available	35.6%
Some artifacts made available	5.6%

- **None** are easy to reproduce
- **50.9%** is impossible to reproduce even with artifacts

Importance of Repeatability
"Non-reproducible single occurrences are of no significance to science."
—Karl Popper

- Expose hidden assumptions
- Define a baseline for future studies
- Keep records for ourselves
- Share research artifacts with others
- Evaluate the contribution thoroughly
- Avoid dissemination of flawed results

Why is it so hard?
- **Cultural factors**: Paper is primary consideration for acceptance
- **Incentives**: Strong emphasis on novel and only positive results
- **Mechanisms**: Difficult to keep track of, communicate and repeat

Dilemma

New Research   **Repeatability**
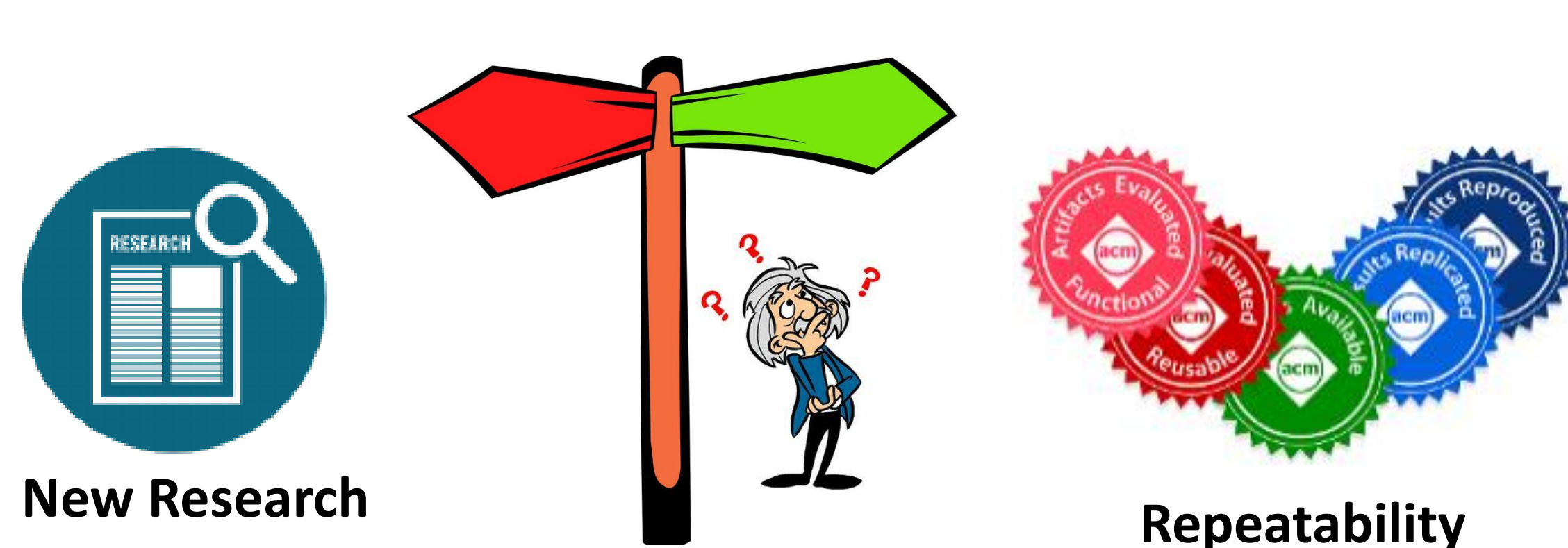
Opportunity in Testbeds

- **Repeatability fundamentals**
  - Testbed hardware is versioned
  - Appliances/Image is versioned
- **Experiment configuration is recorded by testbed**
  - Allocation of specific resources
  - Deployment of specific appliances
- **Monitoring information is recorded**

Contributions

- **Experiment précis:**
  - A summary of experiments information as by-produce of an experiment
  - Can be used in multiple ways: to repeat an experiment, generate a description, etc.
- **Prototype Tool:**
  - Collects crucial user actions from Chameleon Events
  - Generates a command-line script for experiment repeatability
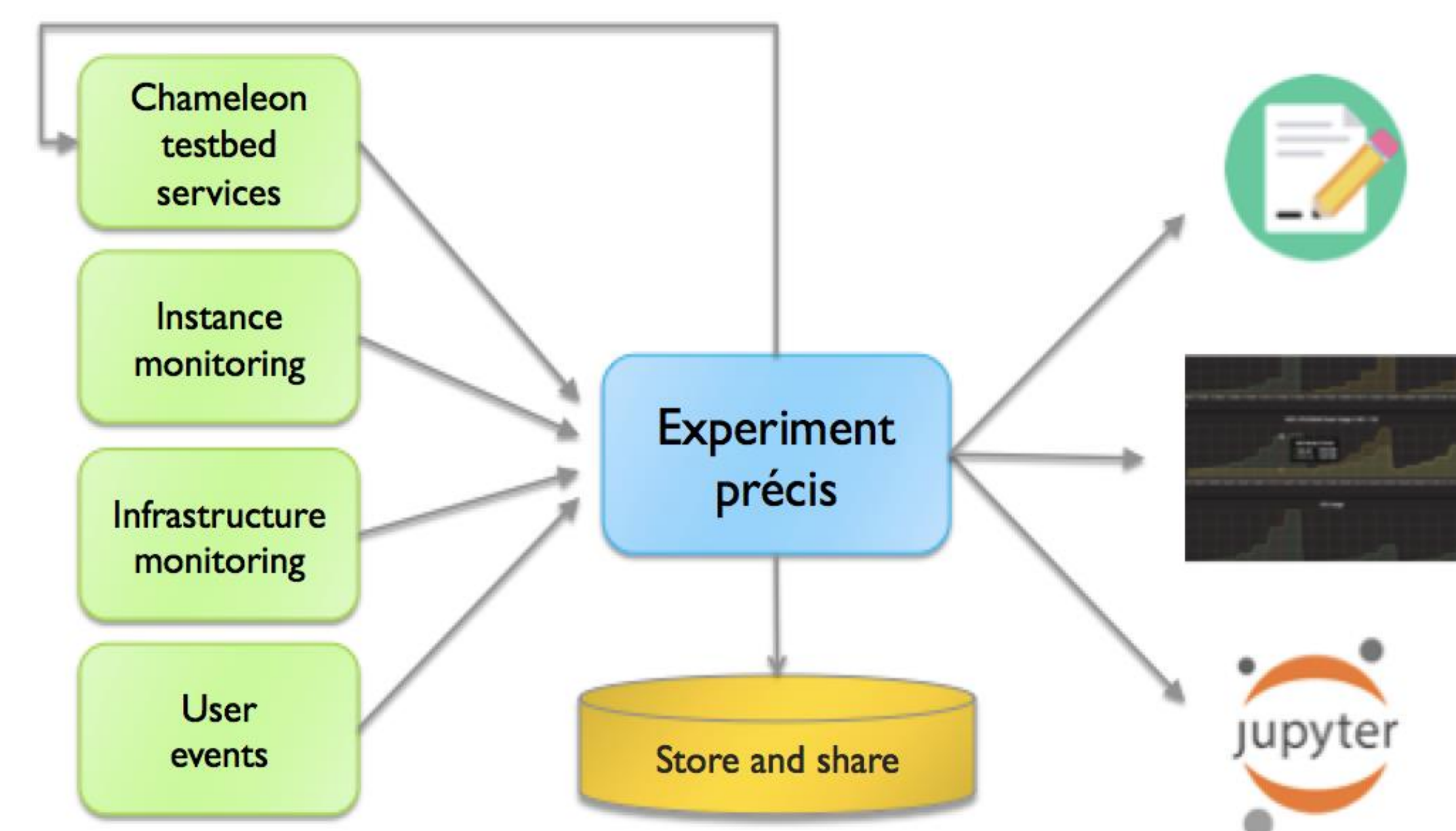 - Generates a detailed experiment description

Experiment Précis

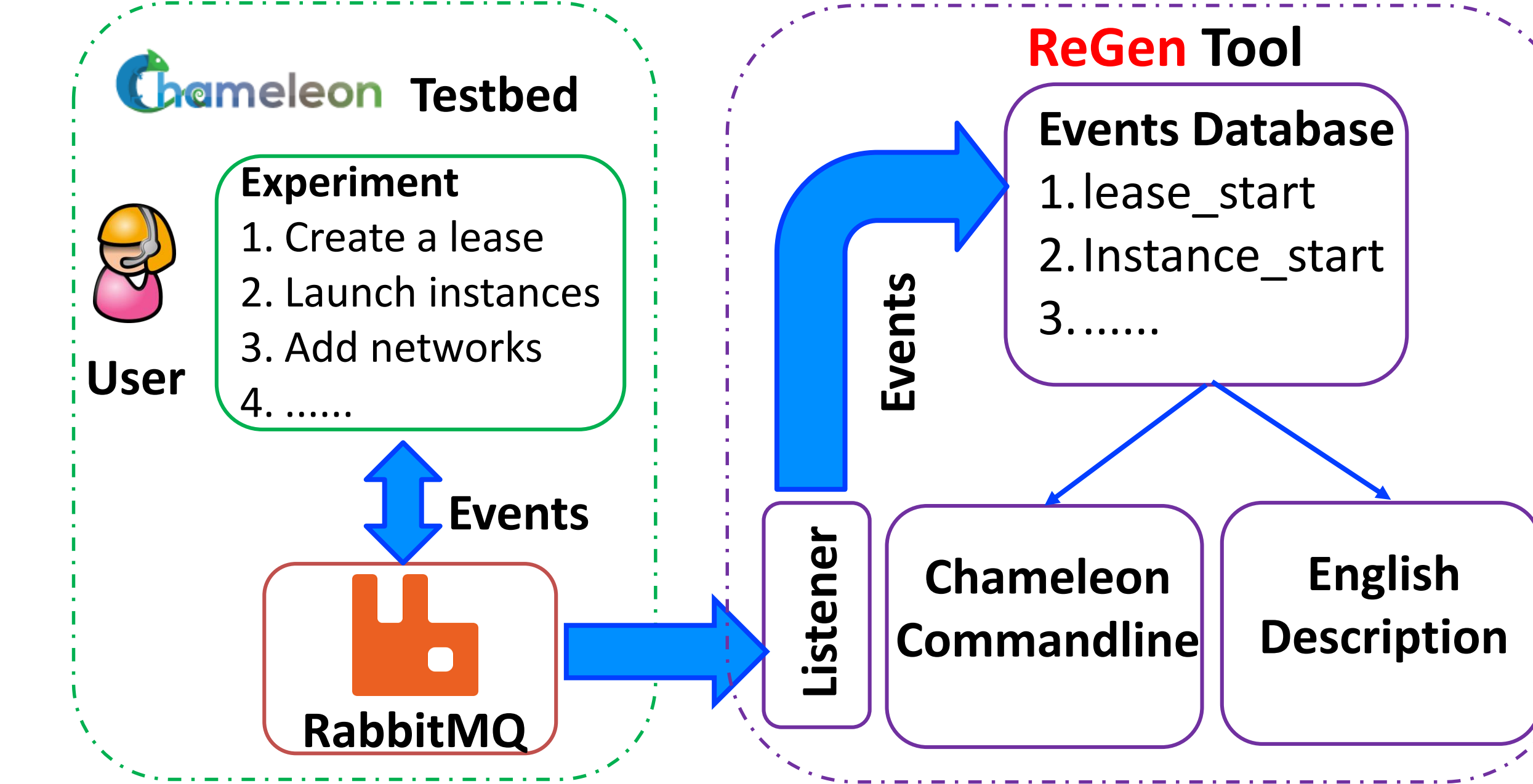Figure 1. Experiment Précis Framework

Figure 2. ReGen Tool

Experiment Précis ≈ Linux "history"

Evaluation

Goal
Evaluate how ReGen helps to reproduce Hbase3813 benchmark from Wang et al.
[Understanding & Auto-Adjusting Performance-Sensitive Configurations. ASPLOS, 2018]

Experiment
The paper above proposed a framework that automatically tunes the configuration with respect to a specific constraint, e.g. by adjusting RPC queue size given memory limitations. In the evaluation, we first run the experiment under default conditions, generating an experiment precis that describes the configuration of our experimental setup. We then modify the experiment precis to represent a controlled modification of an experiment; we use the framework from the paper above to recommend a better configuration. We then repeat the experiment with the new configuration and compare results across the two runs.

Results
- **ReGen** generates a detailed experiment description
- **ReGen** generates a Experiment Précis as a side effect
- **Experiment Précis** can be modified to represent a new experiment with discrete changes to configuration.
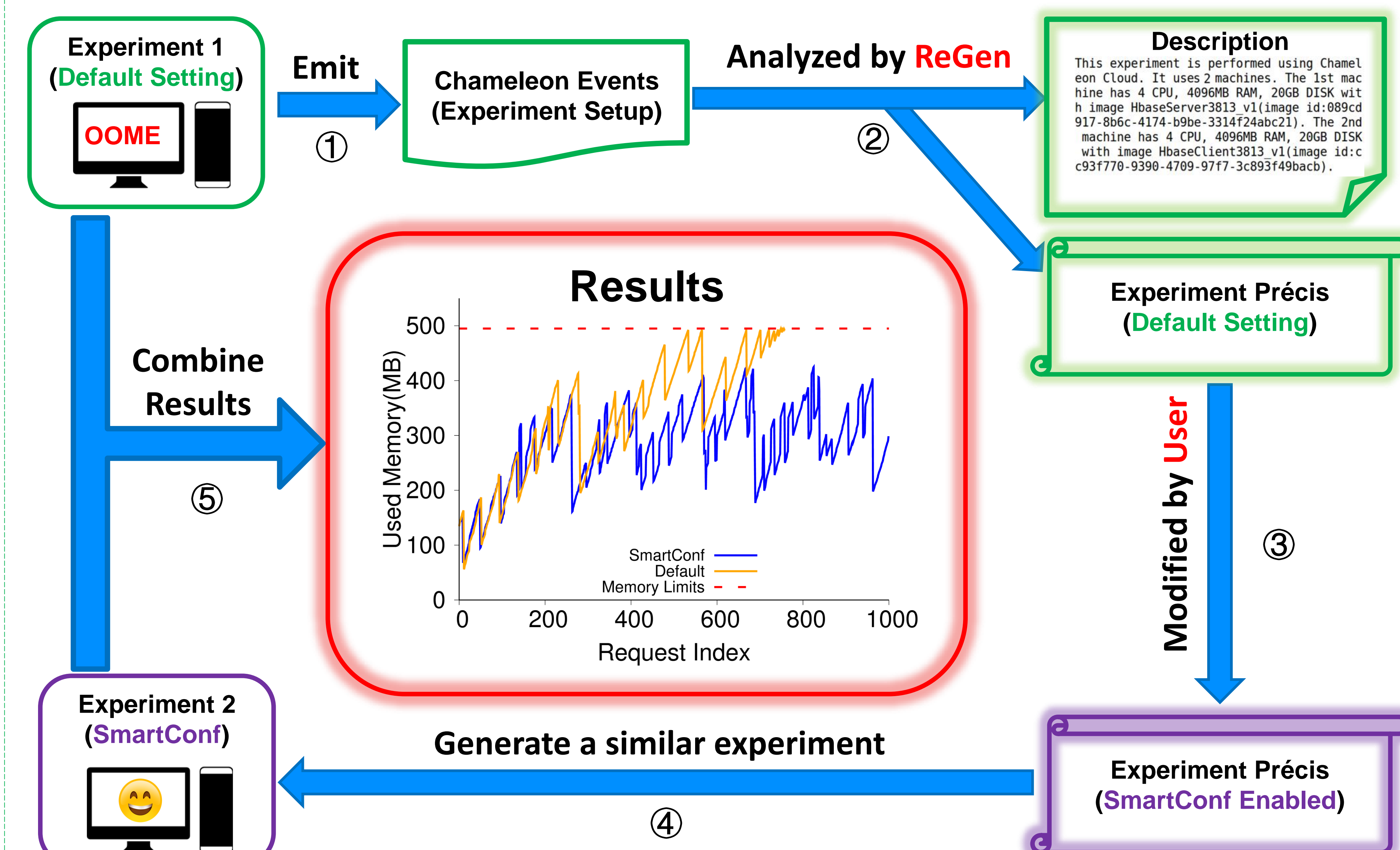

Figure 3. Evaluation Procedure

Conclusions

- We demonstrated that it is possible to capture a major part of experiment information automatically and faithfully
- We used the captured information to repeat the experiment with controlled modifications
- This allow us to share this information with others who may want to make further modifications

Acknowledgement