



# Saint Petersburg Research Center

---

**INTERNSHIPS**

# Backup GitHub

---

NAME : Anisimova Karina

A solid teal horizontal bar at the bottom of the slide.

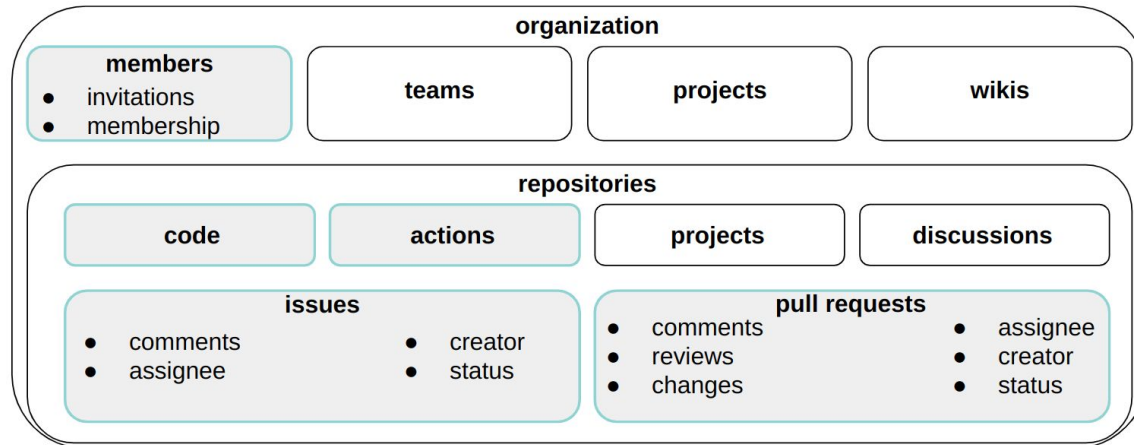
# Project background

---

GitHub organization - a shared account that allows team members to collaborate on multiple projects.

Metadata - information related to a project or organization.

Key metadata: members, issues, pull requests, actions.



# Problem statement

---

Possible risks of using a GitHub organization:

- Modification of organization content without the ability to restore it
  - Force push
  - Repository deletion
  - Organization deletion
- Human factor
  - Accidental or intentional data corruption
- Service failures or errors
- Organization blocking

GitHub does not provide options for backup and restore.

# Competitors Vs proposed solution

	Auto restore	Free	Support	Metadata	Repos content
Backup GitHub	+	+	+	+	all branches
Mirroring	+	—	+	+	all branches
josegonzalez/python-github-backup	—	+	—	± (1)	main branch
Rewind	± (2)	—	+	± (3)	all branches

My project takes advantage of Rewind and solves problems

- (1) with some mistakes
- (2) possible only if the original organization still exists
- (3) not all metadata is restored in its original form

# Different users – different scenarios

---

<b>User type #1</b>	<b>User type #2</b>
Restoring repository content (git data)	Restoring a full organization (metadata)
Force push	Blocked organization
In case of repository data corruption, user can restore previously saved data	In case of blocking the organization, it can be restored on GitHub or other similar service

# What has been done?

---

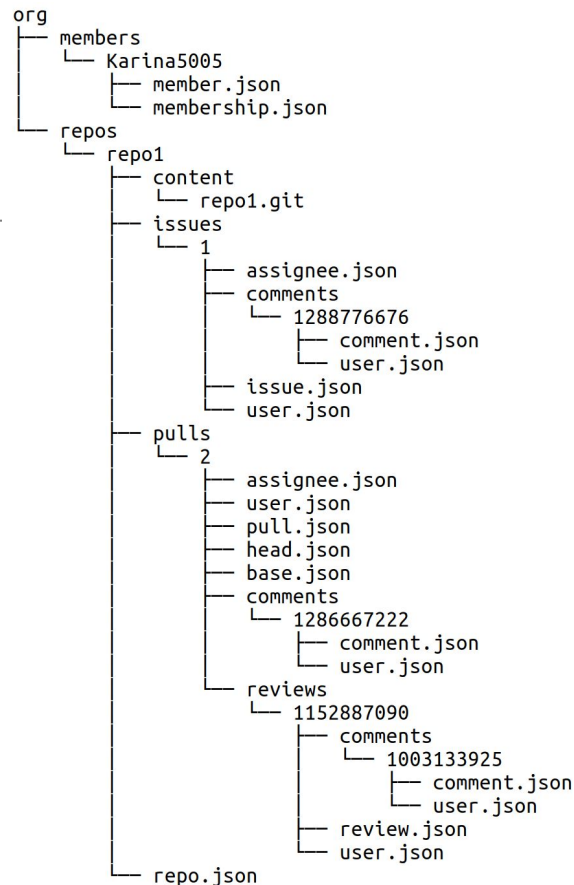
## Backup

### Goals:

- Copying git data
- Copying metadata
- Convenient storage structure
- Minimizing the size of the backup

### Solutions:

- Mirroring repositories
- Importing metadata
- Supporting the structure of GitHub API
- Filtering out redundant information



# What has been done?

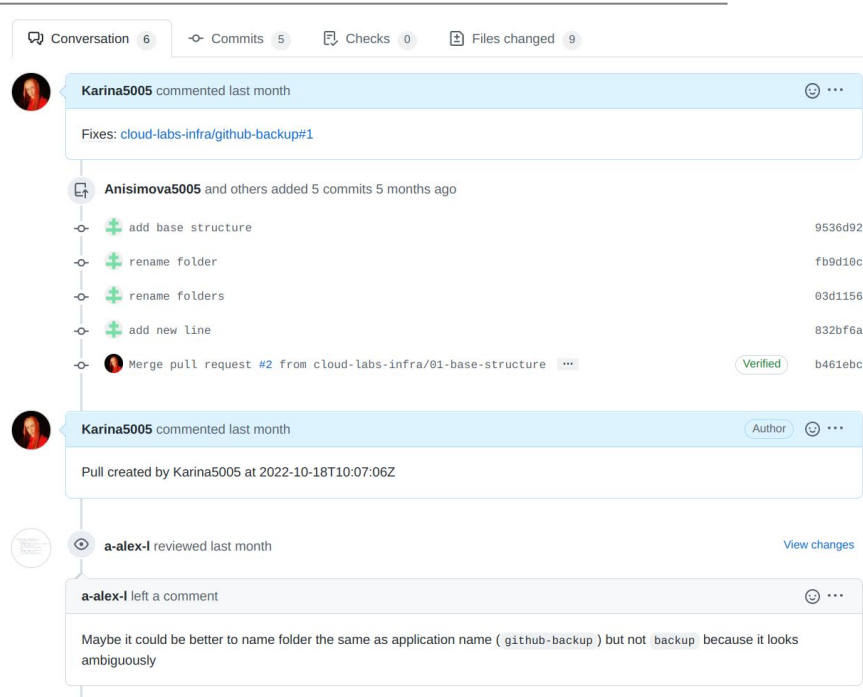
## Restore

### Goals:

- Restoring git data
- Restoring metadata
- The restored organization should be as close as possible to the original

### Solutions:

- Creating repositories from mirrored data
- Reconstructing metadata
- Fault tolerance



The screenshot displays a GitHub pull request interface. At the top, there are navigation tabs for 'Conversation' (6), 'Commits' (5), 'Checks' (0), and 'Files changed' (9). The main content area shows a commit history and a comment section.

**Commit History:**

- Anisimova5005 and others** added 5 commits 5 months ago
  - add base structure (9536d92)
  - rename folder (fb9d10c)
  - rename folders (03d1156)
  - add new line (832bf6a)
  - Merge pull request #2 from cloud-labs-infra/01-base-structure (Verified, b461ebc)

**Comments:**

- Karina5005** commented last month: Fixes: cloud-labs-infra/github-backup#1
- Karina5005** commented last month: Pull created by Karina5005 at 2022-10-18T10:07:06Z
- a-alex-l** reviewed last month (View changes)
- a-alex-l** left a comment: Maybe it could be better to name folder the same as application name ( github-backup ) but not backup because it looks ambiguously



# What has been done?

---

Testing on the organization **Cloud Labs Shared Infrastructure**:

	code	issues	pulls	reviews	members	size	time
backup	+	+	+	+	+	6,7 MB	6m
restore	+	+	±	±	+	—	178m

## Results:

- All necessary data for organization restoration is successfully copied
- Some pull requests are restored as issues instead
- Restoration process takes longer due to limitations of the GitHub API

# What has been done?

---

Testing on the organization **GitTstOrg**:

	code	issues	pulls	reviews	members	size	time
Backup GitHub	+	+	+	+	+	191,2 kB	23s
python-github-backup	+	+	+	±(1)	—	194,5 kB	8s
Rewind	+	+	±(2)	±(3)	—	45,1 kB	12s

- (1) python-github-backup copies reviews as comments
- (2) Rewind copies pulls as issues
- (3) Rewind copies reviews as comments

# What has been done?

---

Testing on the organization **GitTstOrg**:

	code	issues	pulls	reviews	members	time
Backup GitHub	+	+	+	+	+	2m13s
Rewind	+	+	±	—	—	31s

## Results:

- Rewind restores all pull requests as issues
- Better quality of restoration has been achieved
- Restoration process takes longer due to limitations of GitHub API

# Next step

---

- Solution for the restoration speed issue
- Support for other similar services
- Organization synchronization between services

# Anisimova Karina

Contacts for questions

E-mail: [anisimova5005@gmail.com](mailto:anisimova5005@gmail.com)

Telegram: [@ansimova5005](https://t.me/@ansimova5005)

## Project link:

Backup: <https://github.com/cloud-labs-infra/github-backup>

Restore: <https://github.com/cloud-labs-infra/github-restore>

Backup:



Restore:

