

# Tracker miners: File creation times

## GSOC 2021 Proposal

### Nishit Patel

**Email:** nishitlimbani130@gmail.com

**Github profile:** <https://github.com/nis130>

**GNOME GitLab profile:** <https://gitlab.gnome.org/nis130>

**Blog:** <https://www.nishitpatel.tech/blog>

**IRC nick:** nishit130 (Nishit Patel)

**Mobile:** +91 7045576438

### About Me

I am a sophomore pursuing B.Tech in Computer Engineering.

I love to do coding and creating projects and have a curiosity to learn and contribute to the low level projects. Most of my projects in the past have been in the field of App Development and Backend. I have contributed to few open source organizations in the past and being a GNOME user myself I am very much curious to learn new technologies and contribute to this organization.

### Project Abstract

Tracker-miner is the indexer used for extracting the metadata from different file types. The project aims to provide the support for storing creation time of the files and later provide the feature of searching by creation time in Nautilus.

## Project Description

Tracker currently doesn't store the creation time in the database as it was historically not tracked on UNIX file systems, and it's not part of the POSIX specification. Linux now provides a `statx` system call interface for retrieving birth time (Creation time) of files for the filesystems that supports it since kernel version 4.11. Which can be used to store creation time in tracker-miners.

## Known Issues:

- [improve the API for `crttime` in Glib:](#)

The need for fixing this issue is to properly address the dates before the Unix epoch. As you can see in the image below `gio` sets the `time:modified` well ahead in future as it treats the time as unsigned int while `time_t` can handle negative time but `gio` implicitly converts signed `time_t` to unsigned, hence causing this issue. POSIX requires time to be an integer type, but does not mandate that it be signed or unsigned.

```
nishit@nishit:~/Documents/dev$ gio info before-epoch | grep -i time
unix mount: /dev/sda5 / ext4 rw,relatime,errors=remount-ro
time::modified: 18446744073709551615
time::modified-usec: 0
time::access: 18446744073709551615
time::access-usec: 0
time::changed: 1616992033
time::changed-usec: 482488
time::created: 1616992033
time::created-usec: 482488
```

As tracker-miner directly use the `GFileInfo` api provided by the `GLib` to extract the `ctime`, `mtime` etc. On the tracker-miners side, a [temporary fix](#) has been added to use `time_t` for the time being. We can create a permanent patch by using `g_file_info_get_modifcation_date_time()` instead of `g_file_info_get_attribute_uint64()` which will return a `GDateTime` object which can handle time before epoch. Merge request

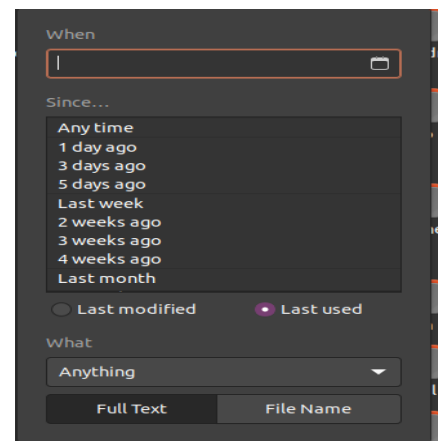
for adding similar api for creation time and access time is already open at [!2017](#).

- **Add file creation time support to Tracker filesystem indexer:**

Tracker right now doesn't provide a feature of querying creation time of the file that needs to be fixed during the summer. It can be fixed by using the new api `g_file_info_get_creation_date_time` which the GLib will provide as stated above. Creation time can be stored in the `nfo:fileCreated` property which is already present in the tracker-miner.

- **Add search by creation time to Nautilus:**

Nautilus currently allows files to be searched using last-modified and last-used times. This issue proposes to add the feature of searching the files by creation time. It will be possible by using the information provided by tracker miners. Design of the search box needs to be brain stormed on #gnome-design IRC.



## Overview:



This flow chart gives an overview of the tasks that are to be done during the summer. It sums up all the above listed issues

# Proposed Schedule

Pre GSOC [April 13, 2021 - May 17, 2021]:

- Getting familiar with Glib, Tracker-miners and Nautilus projects.
- Setting up a local development environment for Nautilus.
- Going through the GLib Documentation regarding time handling.

Community Bonding [May 17, 2021 - June 7, 2021]:

- Communicate with the mentors to formulate a Project Plan based on this timeline, making changes wherever necessary and discuss the design aspect in Nautilus.

Coding Period [June 7, 2021 - August 16, 2021]:

Week 1, 2,3:

- Work on the Tracker-miners to make use of new GLib API for getting file attributes.
- Storing creation time of the files and allowing it to be queried.
- Make use of the [date\\_time helpers](#) for storing time as GDateTime by replacing the current way of storing time as string in Tracker-miners.

Week 4:

- Open the MR for the work done in previous weeks.
- Resolve the threads and testing for the possible bugs.
- Writing documentation.

Week 5:

- Start work on Nautilus for search by creation date.
- Discuss with Nautilus maintainers for design of the nautilus-search-popover.

Week 6,7:

- Make changes as per feedback by the mentors.
- Continue working on the Nautilus part.
- The first round of evaluations would happen around this time.

Week 8,9:

- Complete work on Nautilus.
- Writing necessary documentation.
- Test and resolve threads, if any.

Week 10:

- Finish any pending work or the changes suggested by mentors.
- buffer.

### Tentative Spare Tasks

I have made the above timeline, keeping in mind the maximum time any task could take and the worst-case scenario of having a lot of bugs. Most probably, the work will be done before the end of the GSoC period and hence I would love to work on resolving the issues listed on these projects and use the time efficiently.

## Contributions

### MR's

- **Tracker!346:** [README.md update](#)
- **Tracker!352:** [Strip punctuation](#)
- **Tracker-miners!301:** [tracker-search.c: allow to search the filename](#)
- **Tracker-miners!304:** [tracker search folder fix](#)
- **Tracker-miners!302:** [cli.py: tests for tracker-search](#)
- **Tracker!368:** [tracker\\_offsets\\_function\(\) fix](#)
- **Tracker!381:** [Datetime helpers](#)

Any future MR [here](#)

## Additional Question:

- What is your email address? What is your Matrix ID, or IRC nick?

E-mail: nishitlimbani130@gmail.com

IRC nick: nishit130 (Nishit Patel)

- What is your blog address? While no prior blogging experience is required, having a blog is required for a valid proposal. If you don't have one already, it's not a problem and you can create a free one on, e.g., wordpress.com, before submitting your proposal.

Blog link: <https://www.nishitpatel.tech/blog>

Will write at least 4 blogs during the google summer of code sharing my experiences and journey.

- What city and country will you reside in during the summer?

Location: Mumbai, India

- What is your academic background, including university, major, and expected or achieved graduation date for any degree you have or are pursuing?

I am in 2nd year of Computer Engineering and expected to complete graduation by 2023.

- Who is a possible mentor for the project you are proposing?

mentors: [@sthursfield](#) and/or [@carlosg](#).

- What is the ultimate goal of your proposal?

The goal of the proposal is to store the creation time of the file in tracker-miners and then to use that information to add functionality of search by creation time in nautilus.

- What applications/libraries of GNOME will the proposed work modify or create?

[GLib](#), [Tracker-miners](#), [Files \(Nautilus\)](#)

- What benefits does your proposed work have for GNOME and its community?

The proposed work aims to add the functionality of storing creation time in tracker-miners and it is one of the core library which is used by [many projects](#) and by adding a feature of search by creation time end users of GNOME Desktop will be benefited.

- Why are you the right person to work on this project?

I have been primarily working on tracker project for the past 4-5 months and have understood the code base and working of the project well. I am also planning to continue my contribution for this project irrespective of my selection for GSOC.

- How do you plan to achieve completion of your project?

Detailed schedule and timeline is given above in the **Proposed Schedule** section.

- Please provide a sequence of tasks and subtasks and how long (days) you estimate it will take you to complete each of them. Highlight important milestones/deliverables.

Detailed schedule and timeline is given above in the **Proposed Schedule** section.

- For GUADEC, the GNOME conference will be online. We encourage you to do a short presentation of 3 minutes. What part of your project will be showable?

As GUADEC is planned from July 21st – 25th, 2021 Major part of the project will be completed and tracker-miners successfully storing the information related to creation time can be shown.

- What are your past experiences with the open source world as a user and as a contributor?

I have in the past contributed to few open source projects and I am also part of the tech team in my college coding council ([KJSCE Codecell](#)) which organizes various workshops and talks about open source programmes and maintains [various repositories](#). As a user I have been using open source tools knowingly or unknowingly ever since I started using computers.

- Please include links to your code contributions which have already been merged, or to Gitlab merge requests for the issues you fixed for the project of your proposal or any other GNOME projects. This demonstrates your willingness to learn and familiarity with development workflow.

I have listed my past contributions to the GNOME Organization in the **Contribution** section above.

- If available, please include links to any code you wrote for other open source projects.

Github profile link: <https://github.com/nis130>

- What other relevant projects have you worked on previously and what knowledge you gained from working on them?

I have no prior experience working on desktop search projects apart from contributing to tracker. In general I learned to write well documented and



tested code and following coding style from my past experience contributing to open source.

- What other time commitments, such as school work, exams, research, another job, planned vacation, etc., will you have between May 17 - August 16, 2021? What are the dates for these commitments and how many hours a week do these commitments take?

I will be having my semester exams for which the tentative dates are 20th May to 10th June apart from that I don't have any other commitments or vacations planned.

- What are your Post GOSC Plans?

I'm **not** applying for any other organisation since I am motivated by the journey of the Tracker project and I want to be an integral part of it. I'll also be looking forward to contributing to the project, resolve the issues and hopefully work on the future ideas that Sam Thursfield suggested in his [blog](#) and various other ideas listed in the [roadmap](#).

## Reference

In the research and making of this proposal, multiple sources have been used and they have been hyperlinked at the place of their usage for ease of reading and verification.