

# Improving and Expanding our Internet Scan Ecosystem

## Motivation

Our team conducts large-scale Internet scans that generate substantial amounts of data in periodic cycles. There is also a growing collection of various datasets from the Internet. We are working on improving the infrastructure and have two students develop monitoring and metadata tools for us. We seek interested students who want to expand on the previous work.

You can work either on an IDP in this area, expanding the functionality of these tools, for example, with a visualization of the RO-Crates, which we use for the metadata collection, or implement a better way to handle the ever-growing number of downloads.

Alternatively, we also offer HiWi positions ( $\approx 10$  hours per week, details can be discussed), where you can work on the topics mentioned above and/or help us integrate the tools into our ecosystem by enriching the datasets with meaningful metadata. We are also working on transitioning our scanning infrastructure from manually configured servers to an infrastructure as code model using Ansible, which you can also support us with.

Please write us a mail with a short introduction of yourself and in what you are interested and why.

## Your Task

- Familiarize yourself with our ecosystem
- IDP: Work together with us to improve and expand the tools
- HiWi: Work together with us to integrate the tools with our workflows and datasets

## References

- <https://net.in.tum.de/projects/gino/>
- RO-Crate Metadata: previous work here, <https://www.researchobject.org/ro-crate/>
- Live Monitoring: previous work here, <https://grafana.com/>

## Contact

Tim Betzer      [betzer@net.in.tum.de](mailto:betzer@net.in.tum.de)  
Christian Dietze      [diec@net.in.tum.de](mailto:diec@net.in.tum.de)

<https://net.in.tum.de/members/betzer/>  
<https://net.in.tum.de/members/diec/>

