



Gesundheit – gemeinsam. Kooperationstagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP), Deutschen Gesellschaft für Epidemiologie (DGEpi), Deutschen Gesellschaft für Medizinische Soziologie (DGMS) und der Deutschen Gesellschaft für Public Health (DGPH)

08.09. - 13.09.2024, Dresden

Article

XML version

Send article

Search Medline for

Zaschke P

Hempel P

Bowden J

Bender T

Hanß S

Spicher N

Krefting D

>>

>>

>>

>>

>>

>>

>>

Meeting Abstract

Extending the Biosignal and Imaging Data Managing Platform XNAT by High Performance Computing for Reproducible Processing

- Philip Zaschke** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany
- Philip Hempel** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany
- James Bowden** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany
- Theresa Bender** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany
- Sabine Hanß** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany
- Nicolai Spicher** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany
- Dagmar Krefting** - Institut für medizinische Informatik, Universitätsmedizin Göttingen, Georg August Universität Göttingen, Göttingen, Germany

Gesundheit – gemeinsam. Kooperationstagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP), Deutschen Gesellschaft für Epidemiologie (DGEpi), Deutschen Gesellschaft für Medizinische Soziologie (DGMS) und der Deutschen Gesellschaft für Public Health (DGPH). Dresden, 08.-13.09.2024. Düsseldorf: German Medical Science GMS Publishing House; 2024. DocAbstr. 853

doi: 10.3205/24gmds113 , urn:nbn:de:0183-24gmds1131

DOI: 10.26907/279105113, URL: <https://doi.org/10.26907/279105113>

Published: September 6, 2024

© 2024 Zaschke et al.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/> 

Outline

^ Top

Text

References

Text

Introduction: Computationally expensive methods, such as training deep learning models on large datasets, require increasingly computing power in medical research [1]. Processing on small local machines is oftentimes insufficient w.r.t resource requirements. High Performance Computing (HPC) clusters enable extensive analyses with powerful interconnected hardware.

To ensure reproducibility, data can be stored in a central research data management system. Subsequently, it can be transferred to the cluster through a bridge between the two systems.

In this work, we present a method to connect the research data management platform eXtensible Neuroimaging Archive Toolkit (XNAT) [2] to an HPC system to establish an HPC data analysis pipeline. As a use-case we conducted the annually proposed George B. Moody challenges (GBMC) which address current medical research questions. Typically a prediction task needs to be solved on training data in a certain time period which necessitates efficient workflows.

Methods: The established, open-source XNAT enables storage of medical data and file transfer using a standardized Representational State Transfer (REST) interface. The sytems HPC and XNAT are hosted by our local infrastructure provider. We implemented a so-called HPC Jobscrip and a Python data transfer script as an extension. The HPC Jobscrip executes four steps:

1. starting a XNAT workflow,
2. file exchange via the data transfer script,
3. executing the user analysis script and
4. ending the workflow.

To exchange project files via REST, the data transfer script extends our command line tool XN.

By transferring five recent GBMC datasets (Experiment 1) and self-generated data with predefined projet sizes (Experiment 2) from XNAT to HPC, we evaluated the infrastructure while capturing the transfer speed and determined the impact of three indicators project size, file size, and subject size.

Results: We transferred (4.307 ± 0.622) subjects/min for the GBMC 2023 data to (108.817 ± 2.427) subjects/min for GBMC

proposed George B. Moody challenges (GBMC) which address current medical research questions. Typically a prediction task needs to be solved on training data in a certain time period which necessitates efficient workflows.

Methods: The established, open-source XNAT enables storage of medical data and file transfer using a standardized Representational State Transfer (REST) interface. The systems HPC and XNAT are hosted by our local infrastructure provider. We implemented a so-called HPC Jobscrip and a Python data transfer script as an extension. The HPC Jobscrip executes four steps:

1. starting a XNAT workflow,
2. file exchange via the data transfer script,
3. executing the user analysis script and
4. ending the workflow.

To exchange project files via REST, the data transfer script extends our command line tool XN.

By transferring five recent GBMC datasets (Experiment 1) and self-generated data with predefined project sizes (Experiment 2) from XNAT to HPC, we evaluated the infrastructure while capturing the transfer speed and determined the impact of three indicators project size, file size, and subject size.

Results: We transferred $(4,307 \pm 0.622)$ subjects/min for the GBMC 2023 data to (108.817 ± 2.427) subjects/min for GBMC 2019. In context of transferred files/min, we achieved (108.817 ± 2.427) (GBMC 2019) to (346.784 ± 50.047) (GBMC 2023). Scores from (0.701 ± 0.016) (GBMC 2019) to (180.214 ± 26.008) (GBMC 2023) were observed for the indicator MiB/min.

Experiment 2 showed positive correlations in terms of elapsed times with both total project size and total file amounts (0.09h to 9.73h).

Discussion: The number of subjects stored in XNAT plays the most significant role in transfer behavior, but the combination of the project files and project size may have a greater impact. Nevertheless, all parameters showed an influence on the transfer speed. The combination of overhead of the REST interface and servers, different system load and total amount of Bytes transferred could explain the behavior. The infrastructure will be used in our BMBS funded Project Somnolink. Due to the standardized XNAT REST interface, the bridge is portable to other XNAT systems and SLURM-enabled HPC clusters. We currently extend the infrastructure by a secure RESTful authorization method, that would allow to start and end the workflow directly from XNAT [3]. With our approved secure HPC workflow [4], processing of health data in the care context is possible.

Code availability: Published code: <https://gitlab.gwdg.de/medinfpub/mi-xnat/extensions/clients/XNAT-HPC> 

The authors declare that they have no competing interests.

The authors declare that an ethics committee vote is not required.

by transferring five recent GBMC datasets (Experiment 1) and self-generated data with predefined project sizes (Experiment 2) from XNAT to HPC, we evaluated the infrastructure while capturing the transfer speed and determined the impact of three indicators project size, file size, and subject size.

Results: We transferred (4.307 ± 0.622) subjects/min for the GBMC 2023 data to (108.817 ± 2.427) subjects/min for GBMC 2019. In context of transferred files/min, we achieved (108.817 ± 2.427) (GBMC 2019) to (346.784 ± 50.047) (GBMC 2023). Scores from (0.701 ± 0.016) (GBMC 2019) to (180.214 ± 26.008) (GBMC 2023) were observed for the indicator MiB/min.

Experiment 2 showed positive correlations in terms of elapsed times with both total project size and total file amounts (0.09h to 9.73h).

Discussion: The number of subjects stored in XNAT plays the most significant role in transfer behavior, but the combination of the project files and project size may have a greater impact. Nevertheless, all parameters showed an influence on the transfer speed. The combination of overhead of the REST interface and servers, different system load and total amount of Bytes transferred could explain the behavior. The infrastructure will be used in our BMBS funded Project Somnolink. Due to the standardized XNAT REST Interface, the bridge is portable to other XNAT systems and SLURM-enabled HPC clusters. We currently extend the infrastructure by a secure RESTful authorization method, that would allow to start and end the workflow directly from XNAT [3]. With our approved secure HPC workflow [4], processing of health data in the care context is possible.

Code availability: Published code: <https://gitlab.gwdg.de/medinfpub/mi-xnat/extensions/clients/XNAT-HPC>

The authors declare that they have no competing interests.

The authors declare that an ethics committee vote is not required.

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019 Jan;25(1):44-56. DOI: 10.1038/s41591-018-0300-7
2. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*. 2007;5(1):11-34. DOI: 10.1385/nl:5:1:11
3. Köhler C, Biniarz MH, Bingert S, Nolte H, Kunkel J. Secure Authorization for RESTful HPC Access with FaaS Support. *International Journal on Advances in Security*. 2022 Dec;15(3 and 4):119-31. Available from: <https://resolver.sub.uni-goettingen.de/purl?gro-2/119762>
4. Nolte H, Spicher N, Russel A, Ehlers T, Krey S, Krefling D, et al. Secure HPC: A workflow providing a secure partition on an HPC system. *Future Generation Computer Systems*. 2023 Apr;141:677-91. DOI: 10.1016/j.future.2022.12.019