OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

**Parallel Computing and I/O**

## Efficient Data Structures for Interactive Data Analytics

Tools and data structures for searching and analyzing large repositories of media objects are currently mainly designed and optimized for domain-specific applications. However, if the underlying data should be re-analyzed with a different objective in mind, there are currently no efficient data structures and flexible and efficient access infrastructures available. NIFTI and DICOM are typical examples in the medical domain.

DICOM is an open standard for storing and managing medical imaging data. DICOM files can include so-called tags that include additional metadata about the images in question. These tags can be used to, for example, find all images belonging to one pati or study. Traditionally, tags are stored within the DICOM files themselves, making it hard to query tags across all available image

As part of this thesis, you will 1) review possible solutions, as well as 2) design and implement a solution that stores image data within an object store, while tags will be stored either in a key-value store or a database. The goal is to improve performance of tag access and enable efficient operations such as clustering tags or finding images with similar properties. This is especially important if interactive visualization tools should be used to explore or analyze the image data. This solution should make use of existing libraries such as DCMTK or pydicom. Moreover, JULEA provides object, key-value and database interfaces that can be leveraged for this. DICOM files should be read using the aforementioned libraries and imported into JULEA for further processin

- https://www.dicomlibrary.com/dicom/dicom-tags/
- https://github.com/DCMTK/dcmtk
- https://pydicom.github.io/
- https://github.com/julea-io/julea

Contact: › Michael Kuhn (https://parcio.ovgu.de/People/Michael+Kuhn.html) and › Andreas Nürnberger (https://www.findke.ovgu.de/nuernberger.html)