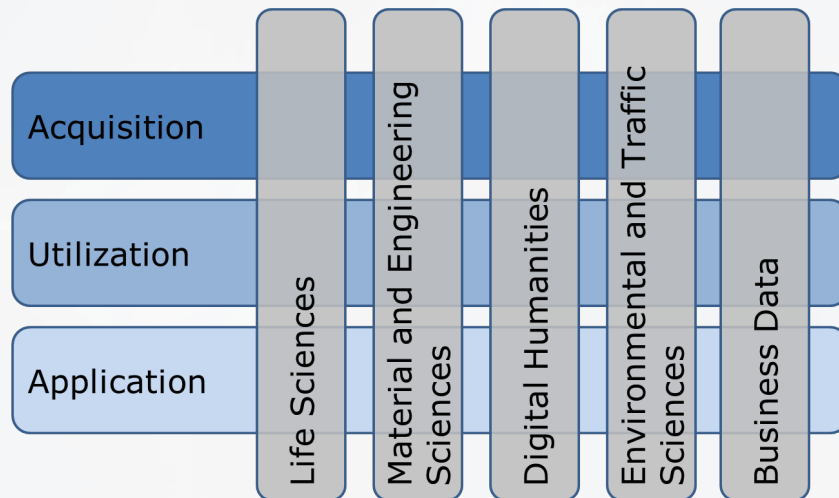


Execution Environments for Big Data: User Centric Scenarios

- Number of users which need HPC on ExaScale level gets smaller
- Tremendous increase in the availability of data for sophisticated scientific analysis
- Scientific data deluge: not only large consortia, but also broad scientific spectrum (long tail of science)

Competence Center for Scalable Data
Services and Solutions ScaDS
Dresden/Leipzig



User Centric Scenarios: Challenges are Manifold

Requirements from the Users perspective

- **What to do with my data? Variable and different data sources:**
 - Large streams of raw data (e.g. microscopes, sensor arrays, ...)
 - Integrate heterogeneous data sets into common analysis (open data, collaborative aspects etc.)
- **Keep control of data:**
 - Cover all aspects of data life cycle
 - Ensure validity and quality of data
 - Is there more knowledge in the data?
- **Deal with heterogeneous environments:**
 - Different data and meta data formats
 - Data not self-explanatory (missing documentation or no meta data at all)

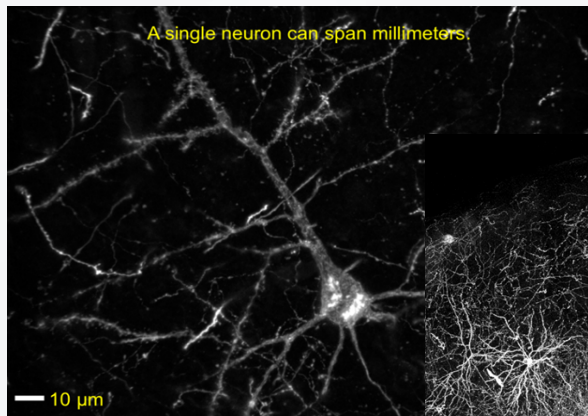
User Need is often not “FLOPS” but “Insight”

Not just static HPC applications – transformation towards more user interaction

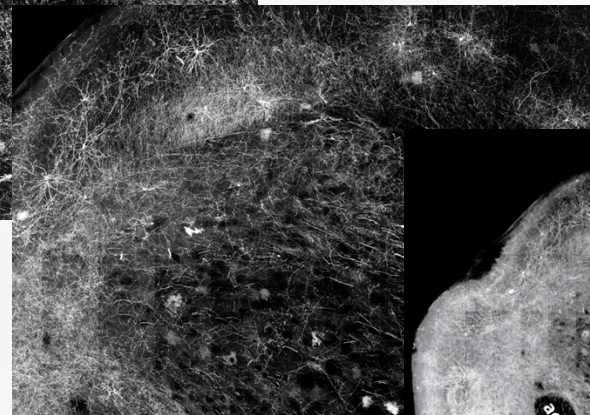
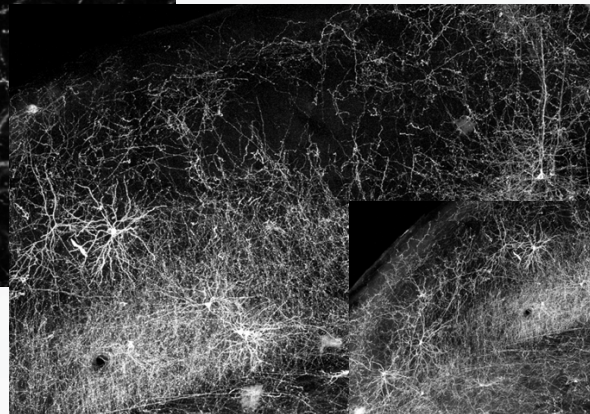
- More user interaction in analysis chain

- Explore large data sets
- Visualization + Analysis

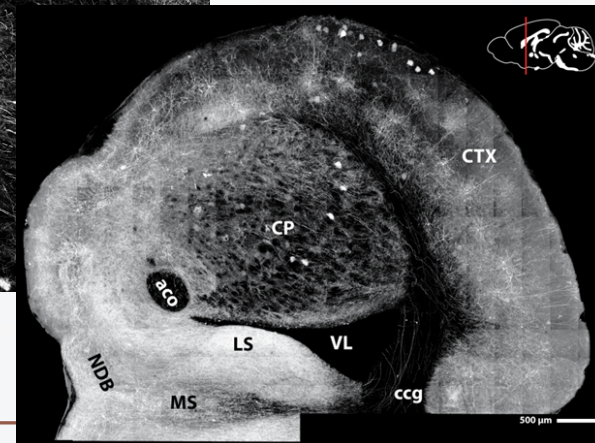
reconstruction,
visualization and
interactive exploration



10 µm



300 µm



Make “ExaScale Research” Productive

- Support the users in their daily work software environments
- Design “HPC scalability” as part of the whole workflow chain
- Requirements to make flexible Big Data analysis possible
 - Easy-to-use tools in the data analysis chain
 - Standards for the interoperability of data processing steps over the full data life cycle (from acquisition to long term storage)
 - Generalized methods for data and meta-data access
 - Integration into easy-to-use workflows supporting the reusability for other purposes or in different scenarios
 - Joint developments and collaborative work towards real Big Data services

- Big Data and Exascale at the push of a button
- It might not be 10^{18} FLOPS every day, but ...
- It easily can be 100.000 times faster than today

