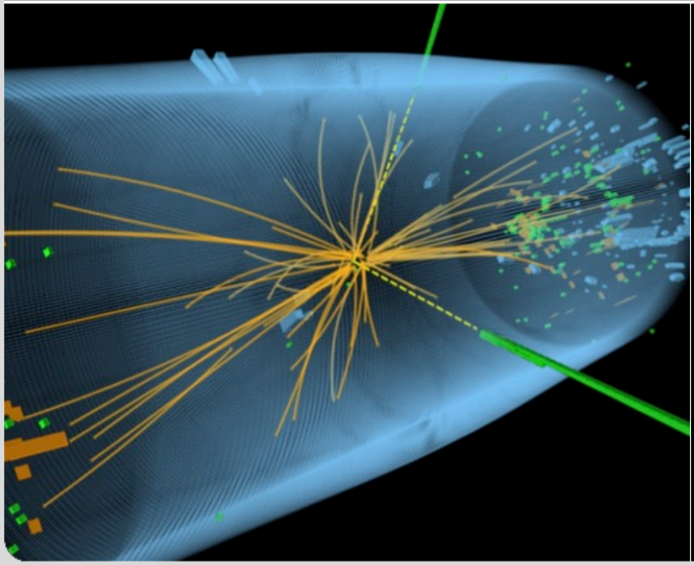
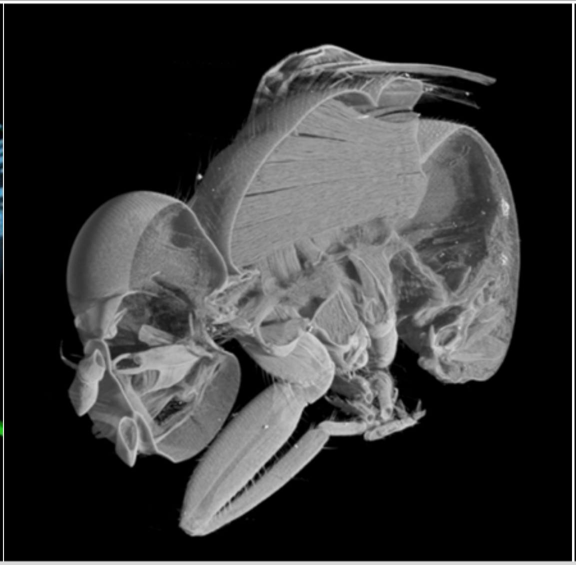
ARD topical workshop on “ps - fs Electron and Photon Beams”

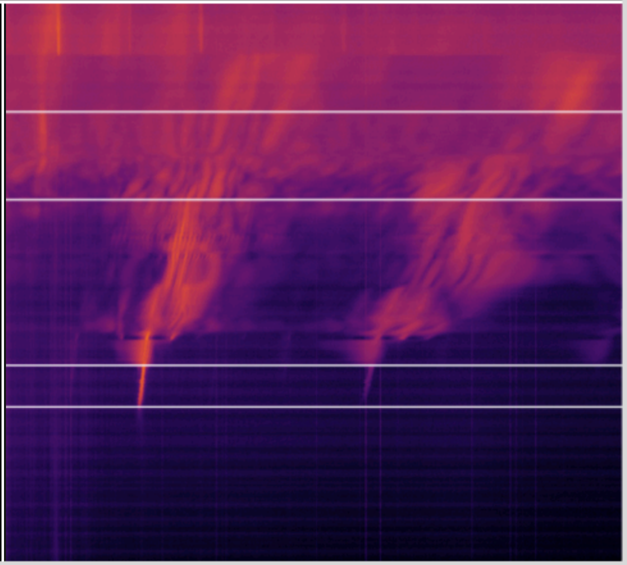
# Online processing with GPUs in data acquisition systems for improved data quality and control


**Andreas Kopmann**

Institute for Data Processing and Electronics





# Scientific computing with GPUs

- Single core age ended -> *parallel programming is required!*
- GPUs are fast, cheap and scalable (up to  4 in a PC)
- TeraFLOP applications: e.g. tomography (2TFLOP/GB)

*Moore's law and CPU*

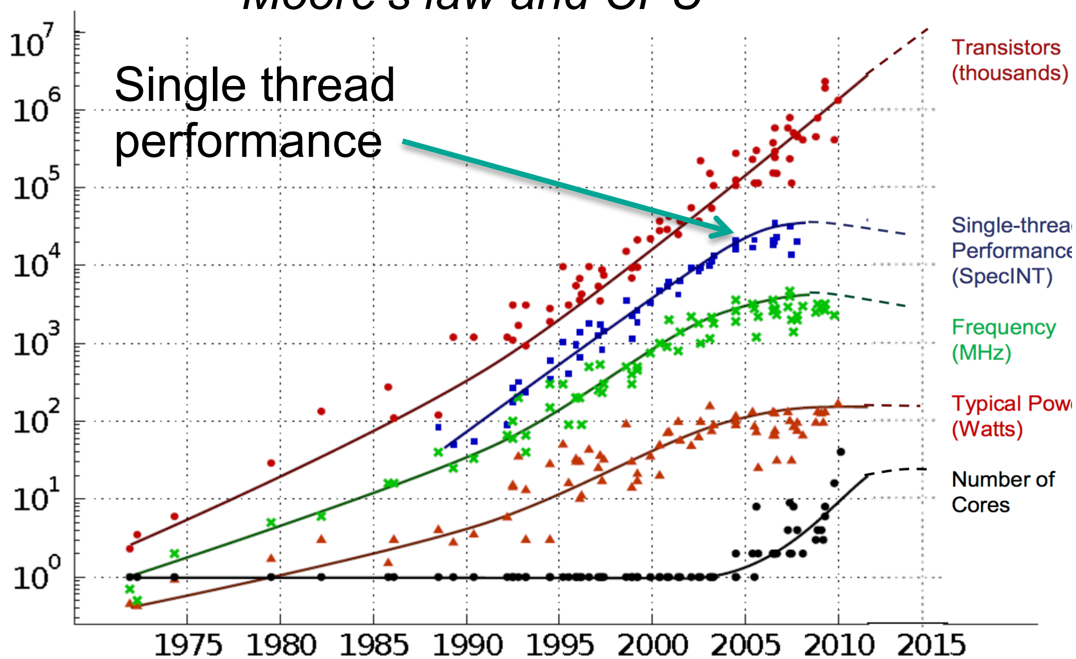
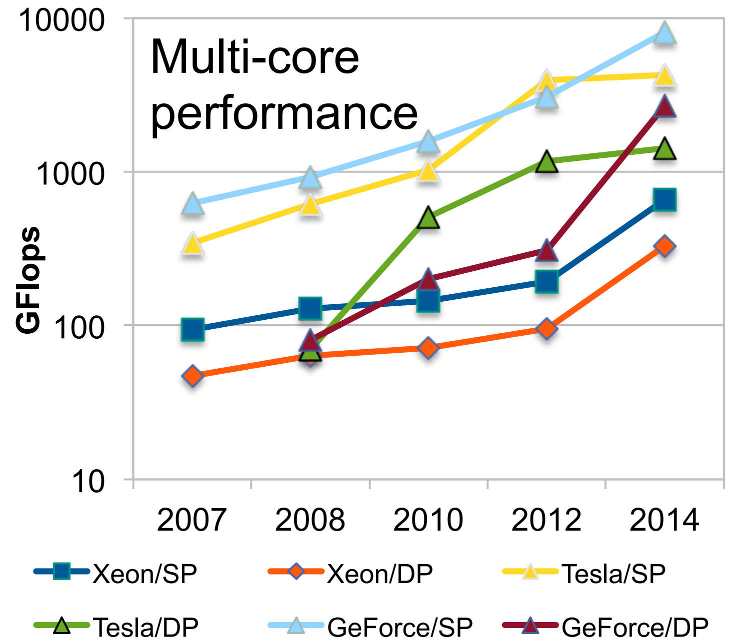



Figure: Chuck Moore, AMD Technology Group CTO

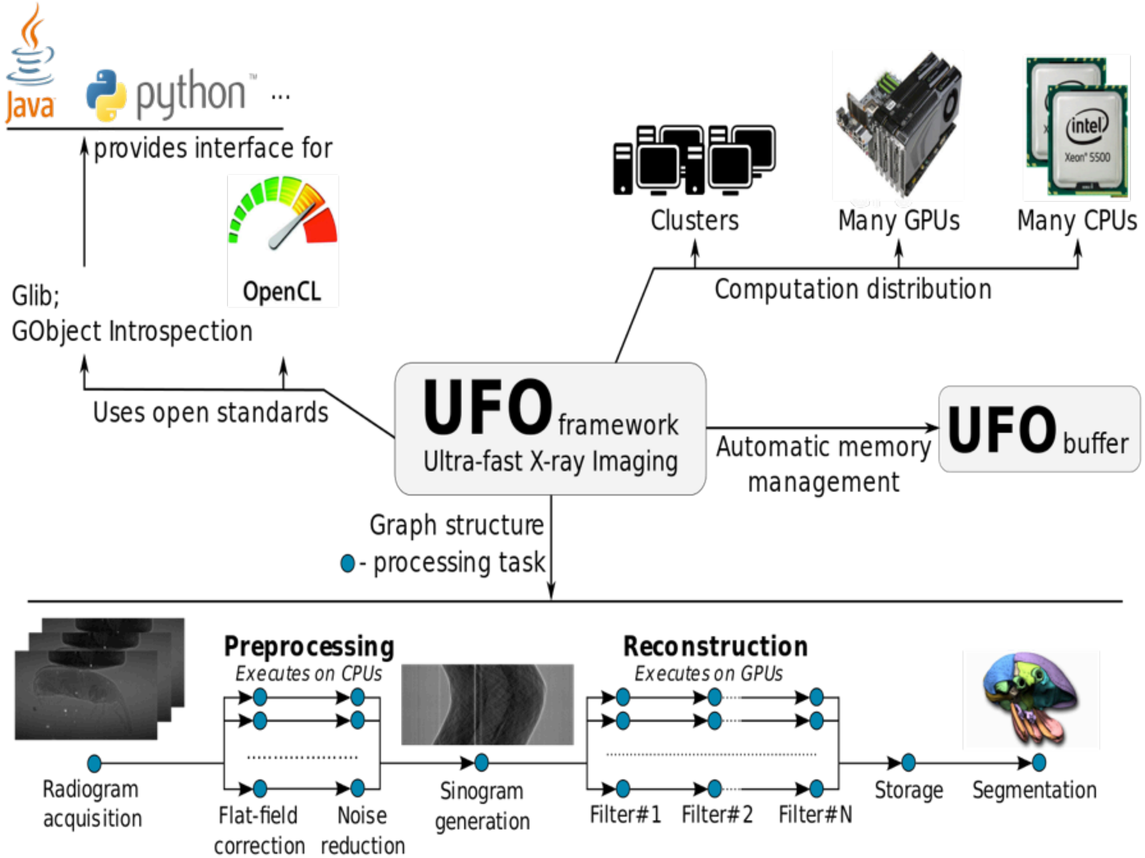
*CPU vs GPU*



# Ultra-fast X-ray imaging – “UFO Framework”

**Goals:**

- Pipelined processing of **data streams**
- (Re-)use of **optimized algorithms**
- Automatic scheduling
  - **CPUs, GPUs**, …
- **Integration** in
  - Control system
  - Analysis tools
- **Easy to use** for
  - Users, Admins, Developers



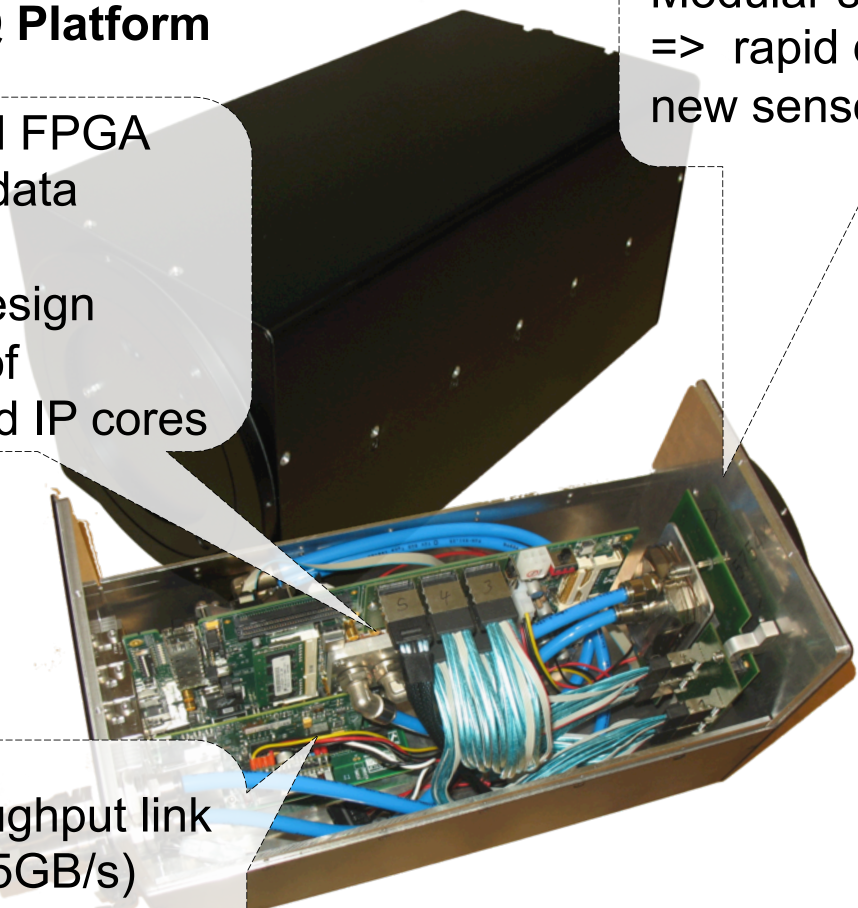

http://ufo.kit.edu

Vogelgesang M et al, Proc HPCC-ICESS (2012) 824-829
Vogelgesang M et al, Proc ICALEPS (2013)

# Custom DAQ Electronics

Karlsruhe Institute of Technology

## UFO DAQ Platform



Embedded FPGA for online data analysis
⇒ open design
⇒ library of standard IP cores

Modular sensor interface
=> rapid development for new sensors

High-throughput link (PCIe, <15GB/s)
=> streaming to GPU for processing

**Applications:**

Helmholtz-Zentrum Geesthacht
Centre for Materials and Coastal Research

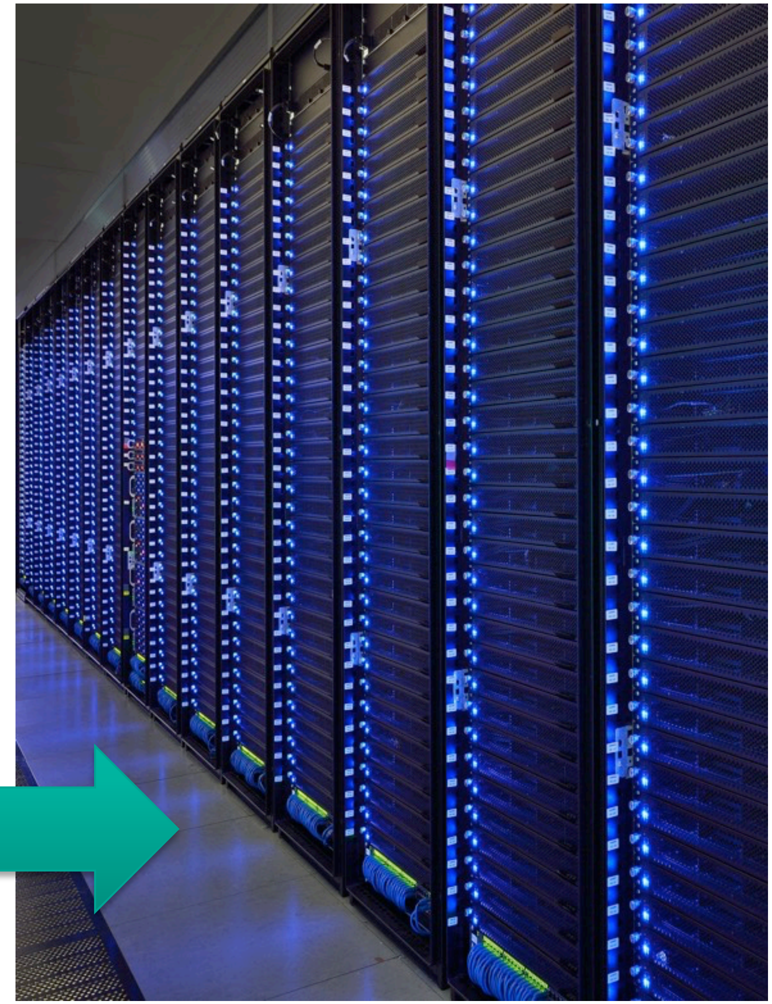
- Phase contrast tomography P05/07
- KAPTURE electron beam characterization, ATP *see M Caselle*

**Enabling HPC technologies for DAQ systems**

Karlsruhe Institute of Technology

# But why online processing?

- Improve data quality
  - Optimize experimental setup
- Reduce data size
  - Apply pre-analysis steps, when possible
- Design new experiments
  - Automatic control

Minerva, Max Planck Institute for Gravitational Physics

# Data-Intense Applications in Photon Science and  Machine Physics

- Facilities
  - PETRA3
  - XFEL
  - Acc. Test Platform ATP: ANKA, FLUTE, ..
- Data-intense methods
  - Tomography
  - Scattering
  - Electron Beam monitoring
  - …
- Detectors
  - Commercial cameras
  - X-ray detectors
  - Fast digitizers


High-speed cameras
(e.g. pco.dymax)


Large-area X-ray
detectors
(e.g. Pilatus 6M)

Custom detectors
(e.g. KAPTURE)

# Detector Rates in Photon Science and Machine Physics

| Device | Rate MB/s | Rate TB/8h | Ch. | ADC bit | Rate Hz | Beamline |
|---|---|---|---|---|---|---|
| "CL Base" | 255 | 7 | | | | e.g. pco.dymax |
| "CL Full" | 800 | 22 | 4M | 16 | 100 | e.g. pco.edge |
| Pilatus 6M | 223 | 6 | 6M | 20 | 10-12,5 | P11 |
| UFO Camera | 1125 | 31 | 20M | 12 | 30 | P05/07 |
| KAPTURE | 3815 | 105 | 4 | 12+4 | 500M | ANKA |
| KAPTURE2 | 15259 | 419 | 4-8 | 12+4 | 1G | |
| KALYPSO2 | 4883 | 134 | 512 | 14 | 5MHz | |

- CL = Camera Link
- All devices without compression

# “Real-life” Data Rates (2015-16)

Karlsruhe Institute of Technology

PETRA3

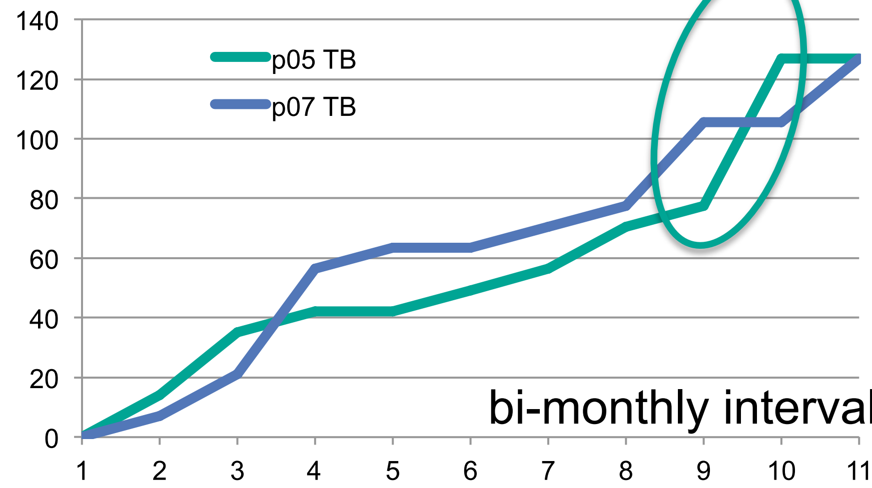
- Tomography (P05/P07)
  - Mean 75 TB/a
  - Max 300 TB/a (eq. 75h op)

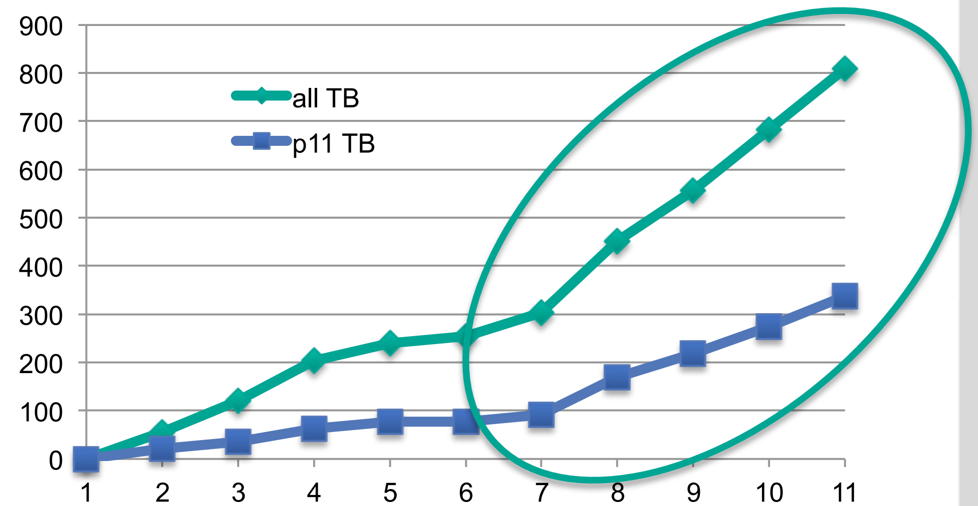
- Imaging / Diffraction (P11)
  - Mean 200 TB/a
  - Max 370 TB/a (eq. 500h op)

ANKA

- Tomography (TopoTomo)
  - Mean 40 TB/a (eq. 42h op.)
- KAPTURE
  - Mean 10 TB/a (eq. 1h op.)

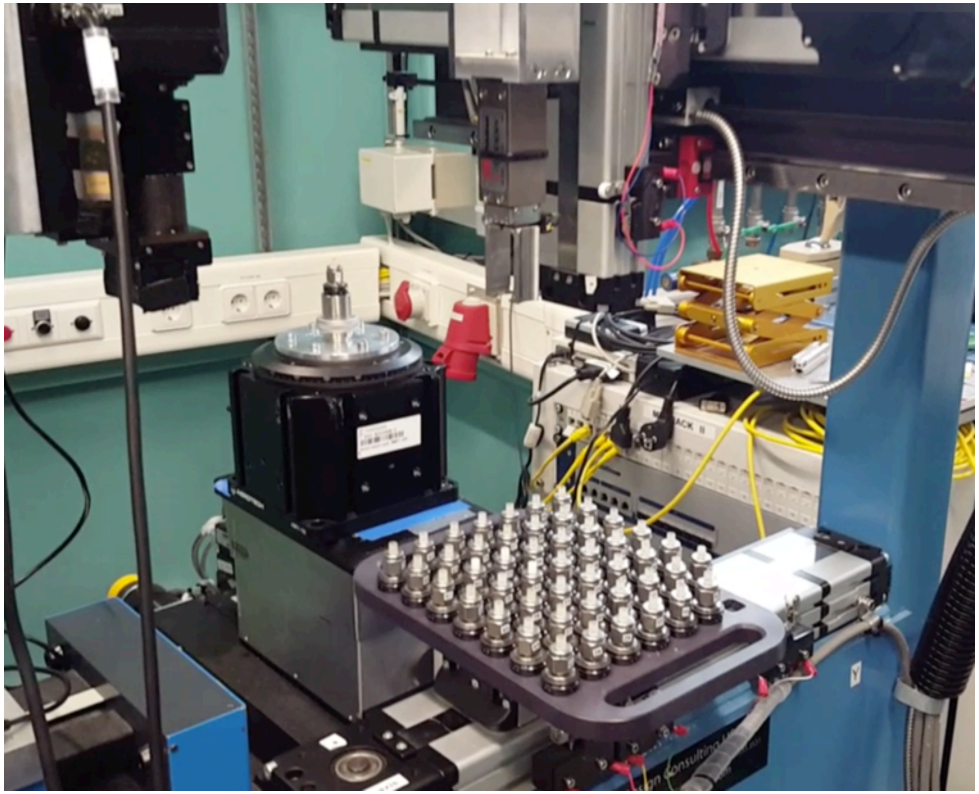
extracted from previous chart



# Beamline Automation

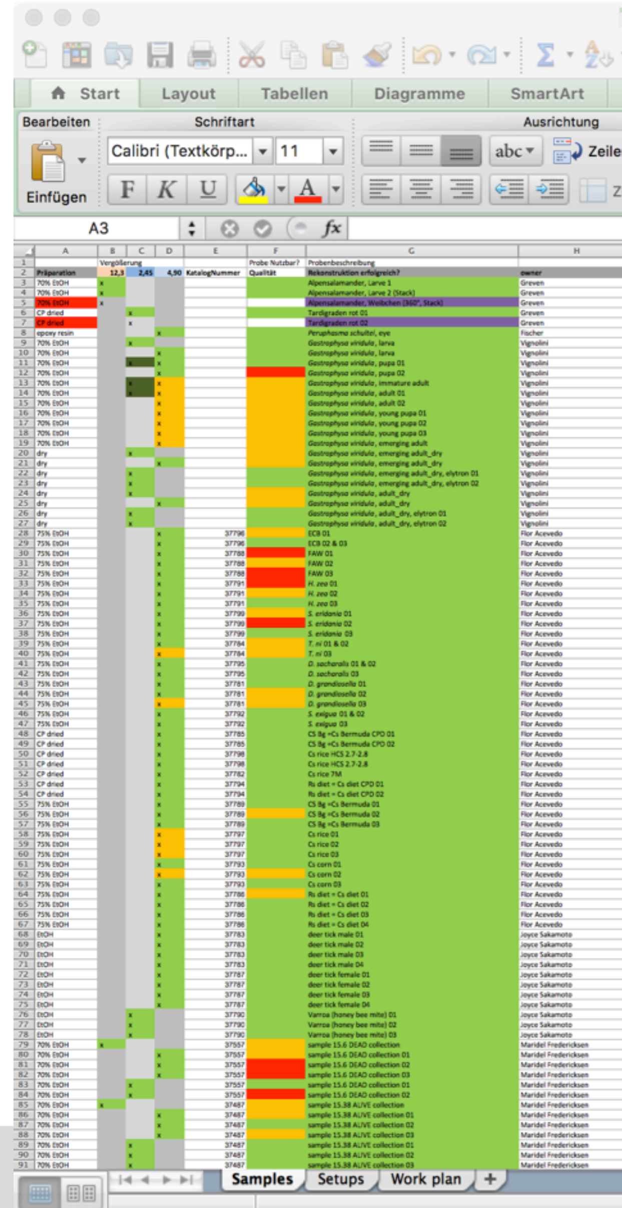
- Robots for sample changers
- Sample identification (e.g. by QR code)
- Camera revolver, lenses
- Advanced control systems



# Tomography Services for Morphology

Karlsruhe Institute of Technology

- ANKA domain expert collects samples
  - Check of sample quality
- Measurement at ANKA
  - Beamtime 6 days (e.g. Nov 2015)
  - Up to 300 samples from 14 scientists
  - Methods: Radiology, Tomography, Cine-Tomography
  - Samples: in alcohol, dried, amber, …
  - More than 90% of the samples are usefull
  - Interested scientists might join
- Datasets are automatically transfered to processing storage
  - Analysis by users



# Quantitative Studies of Fly Puparia

- Digitalization of 30 million-year-old phosphatized insects
- Little-known fossil type!

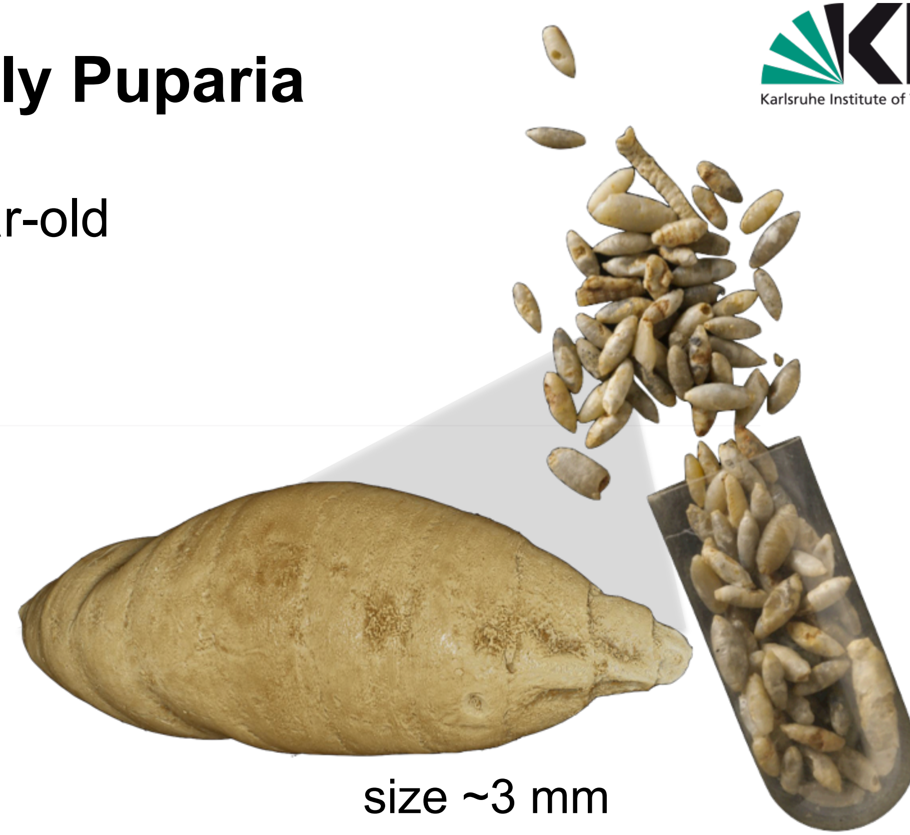
Scanning:

- February 2016: 29 puparia
- August 2016: 1379 puparia

**Result: 1408 scanned puparia!**

**Datasets: ~60 TB**

**Automation enables quantitative studies**

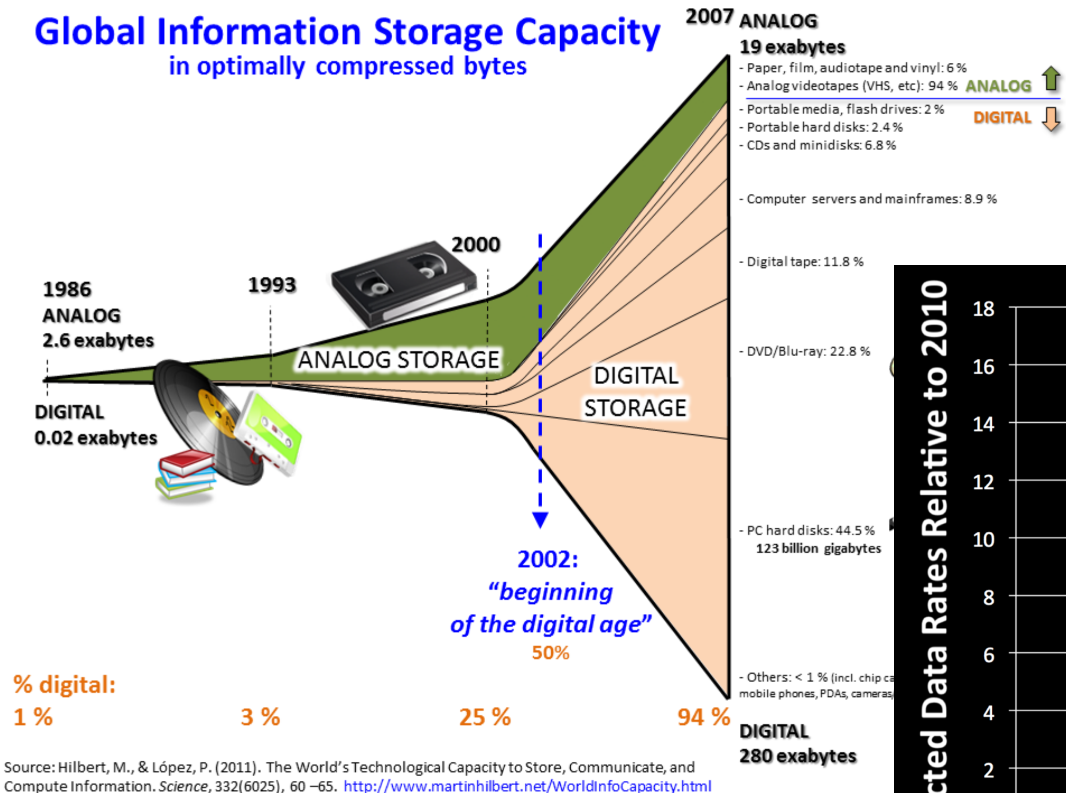


size ~3 mm

*T. van de Kamp, ANKA*

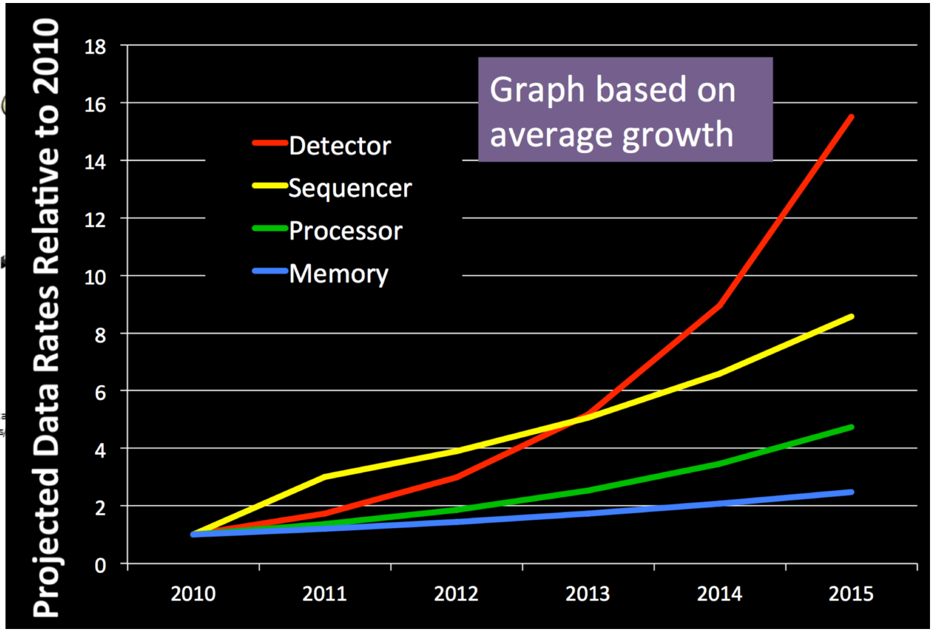
# Digital transformation and large datasets

- "Big datasets are so large or complex that traditional data processing applications are inadequate" (Wikipedia)




*Nersc: "Data growth is outpacing Computer growth"*

Source: Yelick K, LBNL/Nersc




*According to Moore's law: factor 2 every 2a*

# Data Challenges:
# 1. Reduce Data as Early as Possible

- Adopt pre-processing, reconstruction to detector rates
- Avoid user interaction, define workflows
- Make use of trigger to filter relevant data
- Assess quality of data

- Heterogeneous FGPA-GPU computing is very powerful
- But tasks are often domain specific

- Solution:
  Modularity and frameworks with core functionality

# 2. Adopt Analysis for Large Datasets

- Huge amount of data
  - User often do not have the necessary resources (computers, memory, software) available in order to process the datasets at their home institutions

Verbundforschung:

astor --> nova

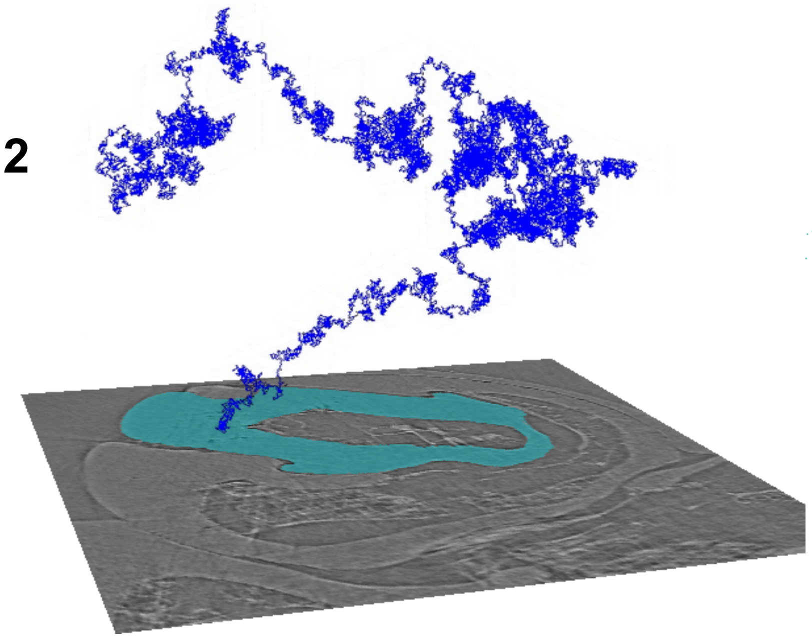
- Data analysis is still very time-consuming
  - Segmentation of tomographic datasets is usually done manually

- Solutions:
  - Establish analysis centers at the facilities
  - Involve more scientists by collaborative analysis and open data
  - Re-evaluate known algorithms with recent computing performance

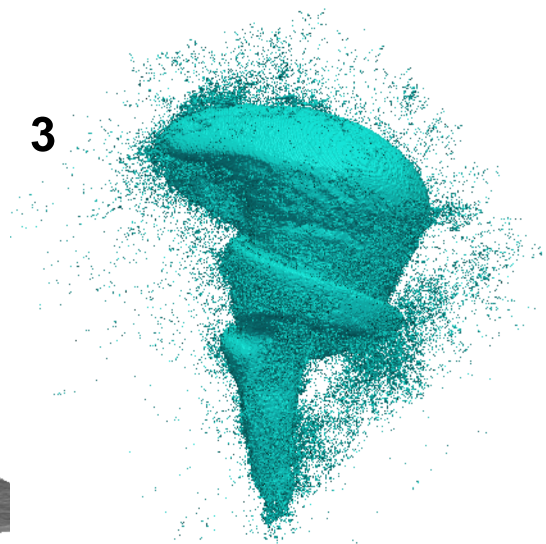
# 3D Segmentation Algorithms

- Algorithm – *requires performance of GPUs*
  1. Preprocess: Label the segments in some well chosen slices.
  2. Start a great number of weighted random walks at each labeled pixel.
  3. The number of hits by random walks which were started in the same segment leads to the probability that a voxel belongs to this segment.
  4. Post processing using Active Contour Method in 3D.

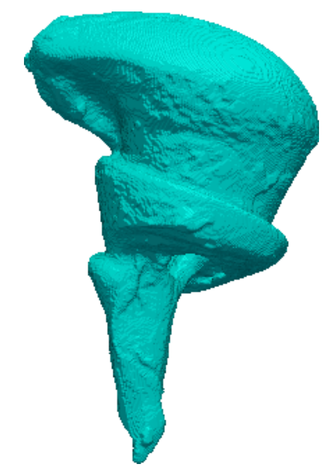
**2**



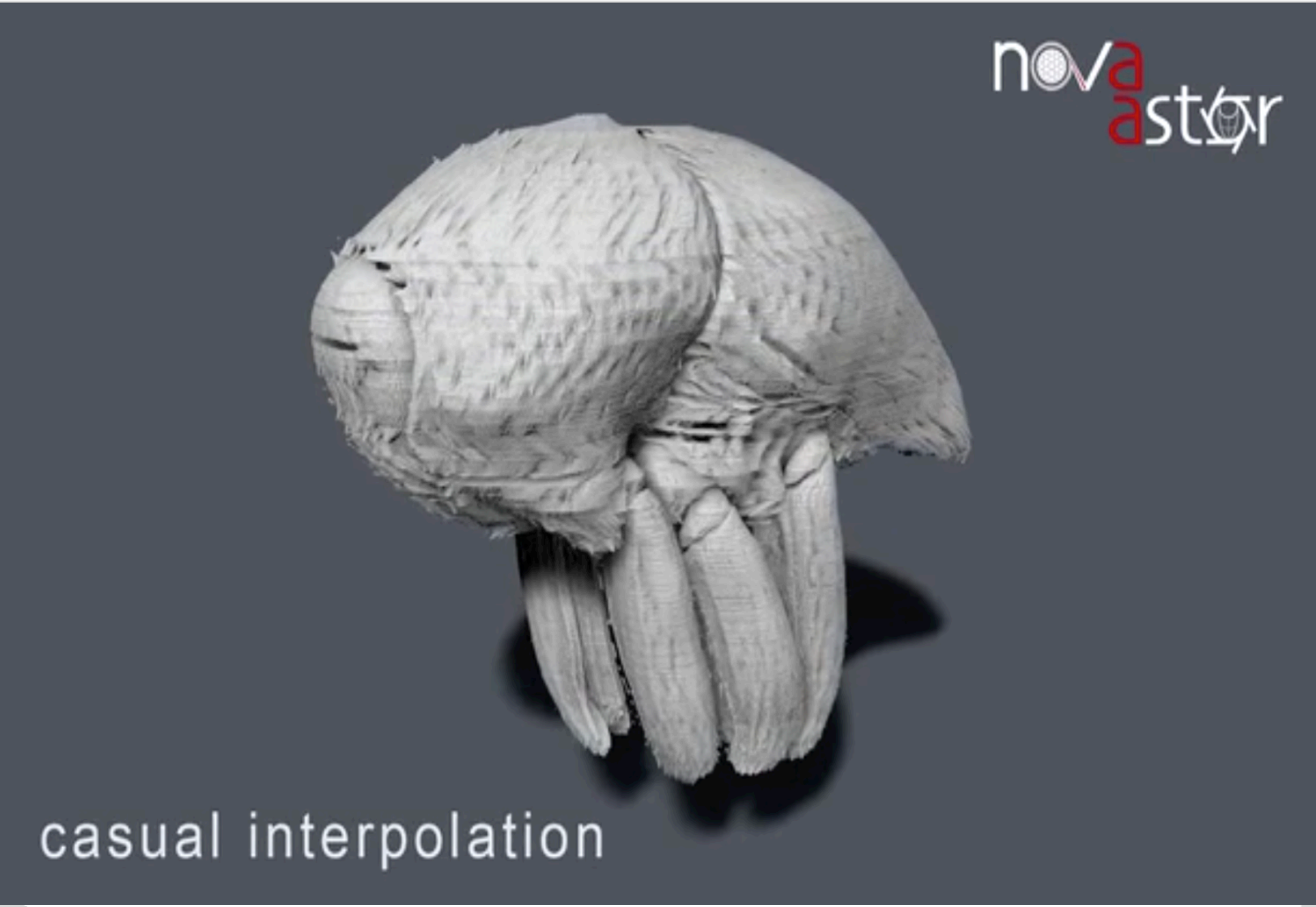
**3**

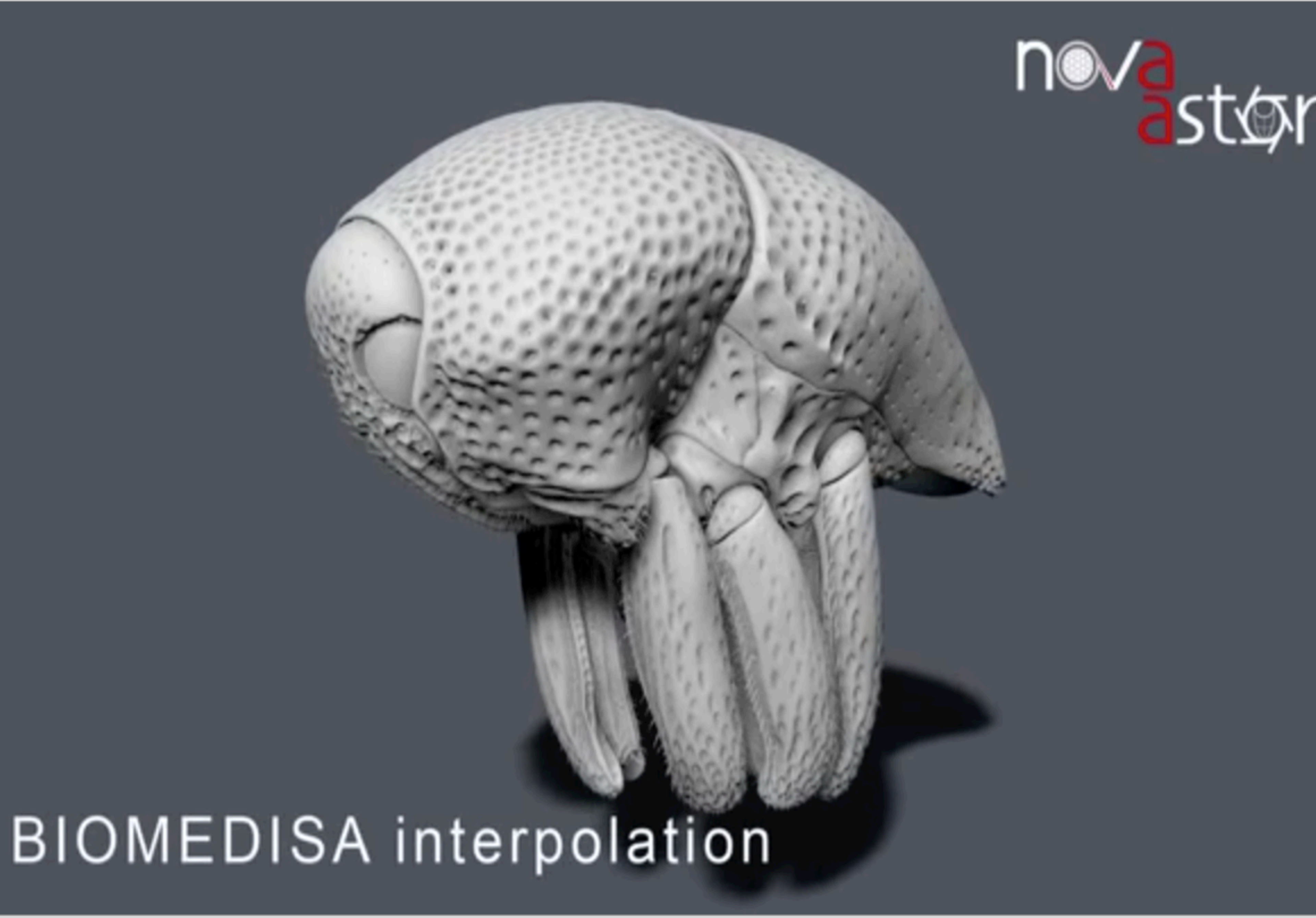


**4**



P Lösel, V Heuveline, SPIE 2016, Paper 9784-92
U Heidelberg

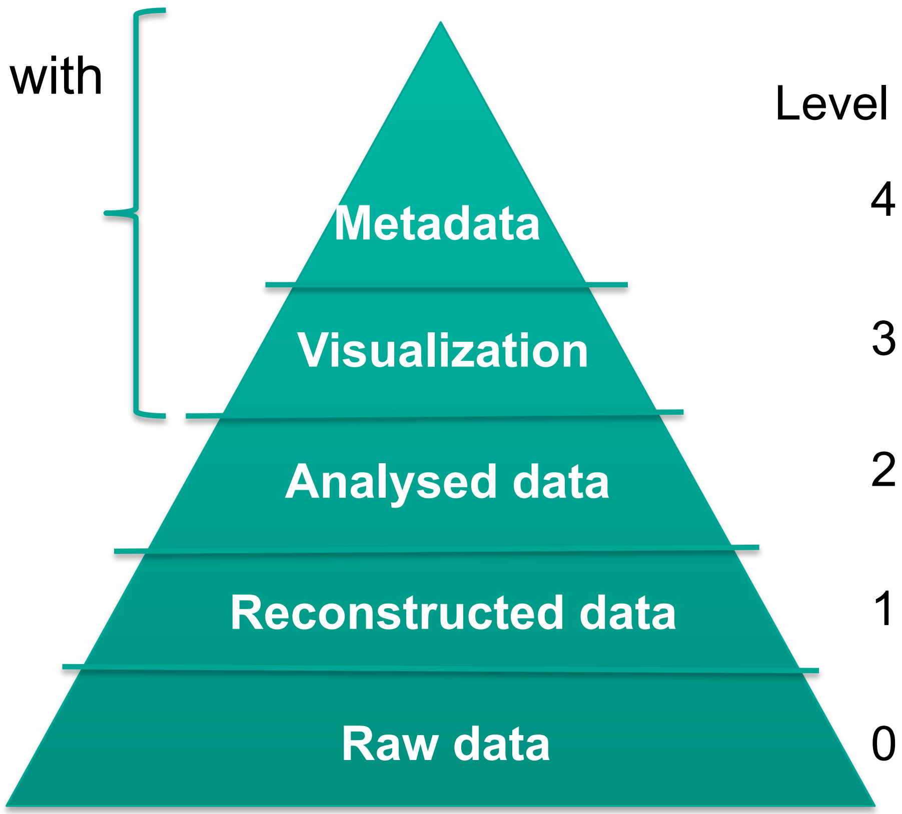

casual interpolation


BIOMEDISA interpolation

# 3. Hierarchical Organization of Datasets

KIT
Karlsruhe Institute of Technology

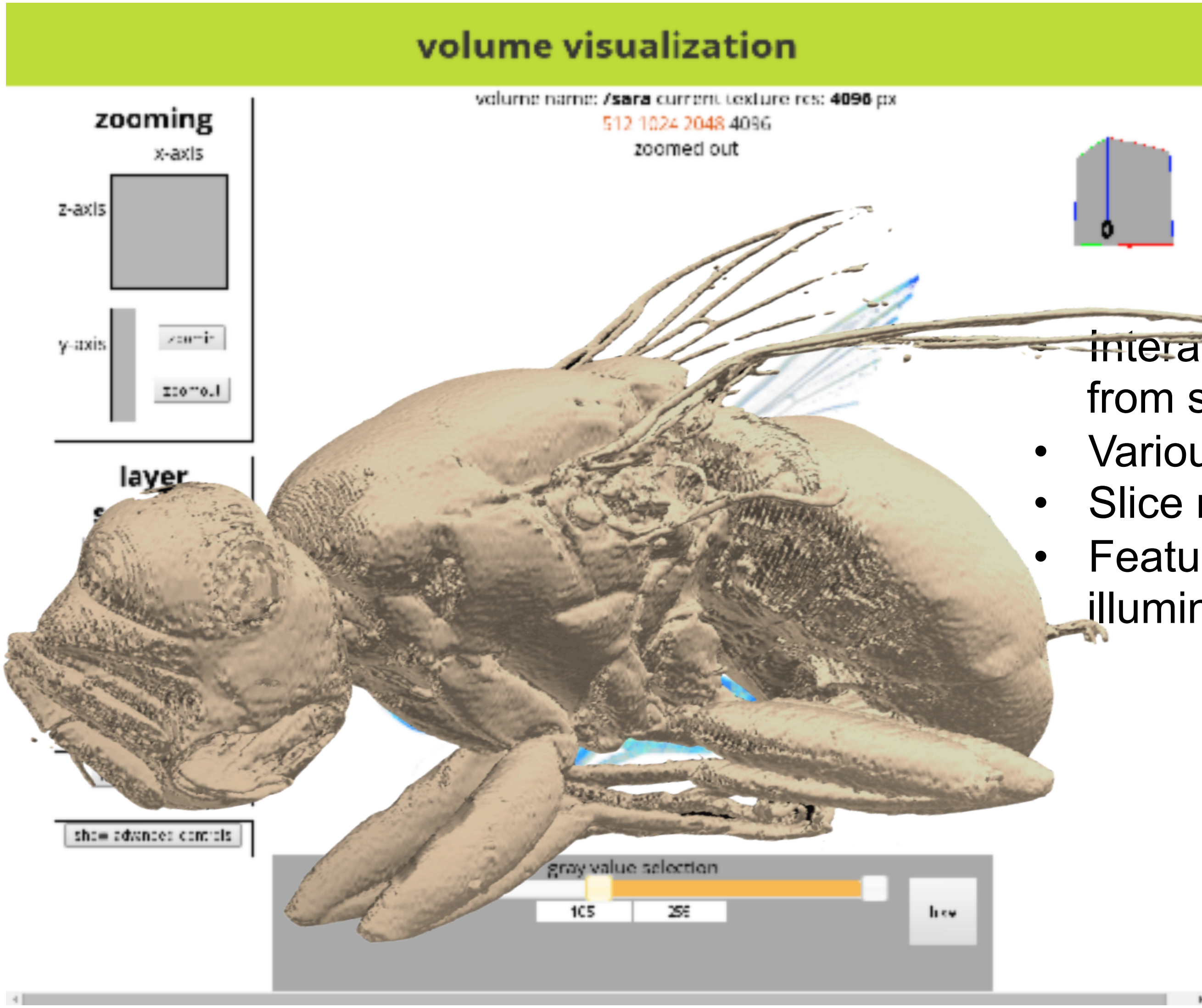
- **Long-term archive have long access times**
- Organize large dataset in levels with decreasing size is needed
- Only metadata is available in data catalogs



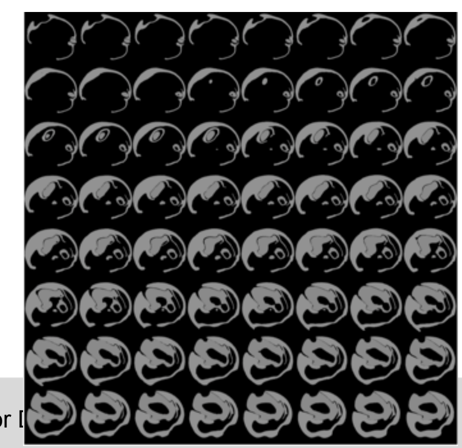

**Solution:**

- **Extend classical metadata by visualization data**
- **Develop 3D-visualization for data catalogs**

# Web Data Catalog with 3D Visualization




- Interactive 3D visualization from slice maps
- Various levels of resolution
- Slice maps
- Features: cuts, thresholds, illumination models , …



# Conclusion

- Data rates will reach 1PB/a soon
- Exascale data center will come

- **Still datasets need to be analyzed**
  - Online data assessment
  - Improve Analysis methods for large datasets
- **Data management matters**
  - Scalable data formats with extended metadata
    + scalable visualization
  - Dependencies between datasets
    e.g. higher resolution data should replace older

*Solutions are often domain specific and*
*will not appear automatically by generic "Big Data" research*