DLCL Key Technologies: March 2015 Results

Alexander Vondrous
Rainer Stotzka, Nick Kepper, Swati Chandna, Robin Dapp, Richard Grunzke, Volker Hartmann, Michael Hausmann, Jürgen Hesser, Thomas Jejkal, Ralph Müller-Pfefferkorn, Michelle Pfeiffer, Francesca Rindone, Danah Tonne, Xiaoli Yang, Eberhard Schmitt, Margund Bach, Ajinkya Prabhune, Armin Volkmann, Hjalte Raun

LSDMA

Subprojects

- ***Light Optical Nanoscopy*** (Heidelberg, Mannheim, Mainz)
- ***High Throughput Microscopy***

  Selective Plane Illumination Microscope (Karlsruhe)

  Gen Scans (Dresden: TU + MPI CBG)
- ***ANKA Tomography***

  ANKA: Ultra Fast Tomography
- ***Archeology***

  Detection of Ground Monuments
- ***DARIAH & eCodicology***

  Arts & Humanities, ESFRI DARIAH EU + BMBF DARIAH DE

Light Optical Nanoscopy

Challenges

- Heterogeneous datasets of 150 TB each
- Partially unstructured and complex metadata

Results achieved

Nanoscopy Data Repository for

→ **Base metadata schema**

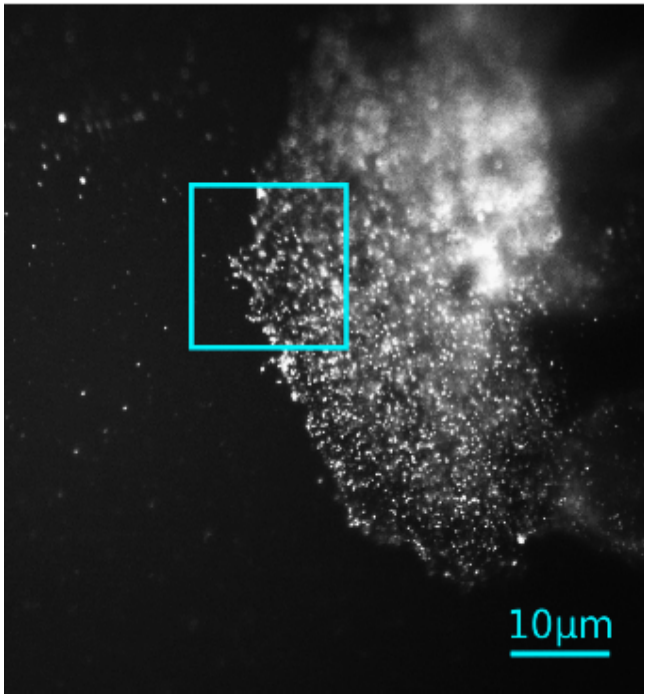
→ **Automatic metadata extraction**

→ **Data search technique**

Nanoscopy Generic Client Service API for

→ **Fail-safe data ingest and download**

→ **GUI web tool for monitoring and accessing datasets**



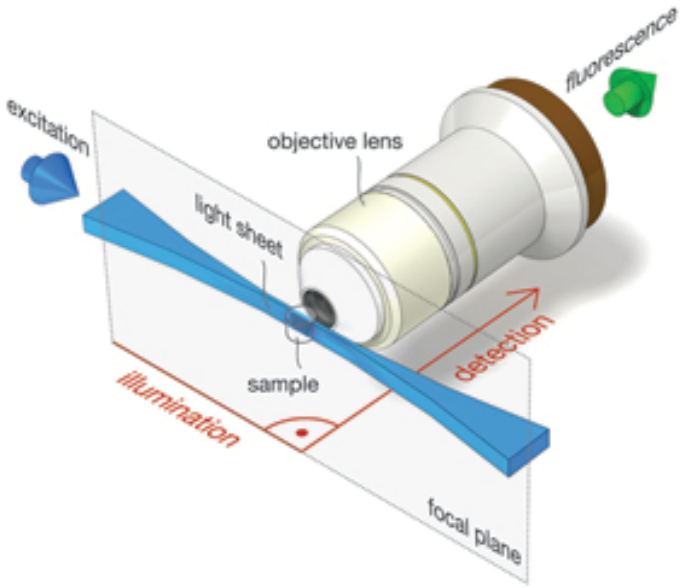
Light microscope image of a cell

High Throughput Microscopy

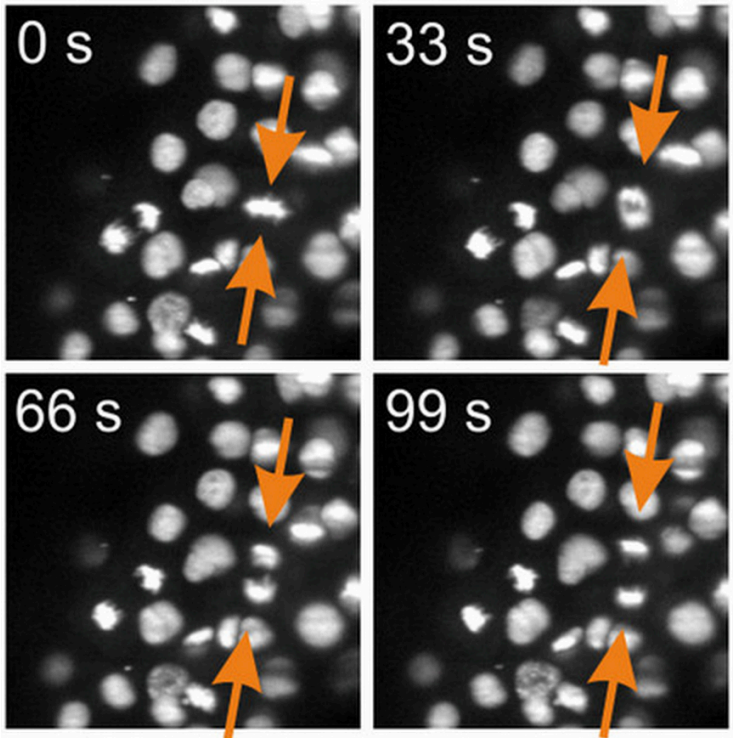
- Enabling efficient reduction and compression
- Design of a user-friendly Big Data Biology Life Cycle
- Evaluation of dCache and iRODS for data management at MPI-CBG

→ **Prototype via master thesis approaches 850 MB/s via rule-based compression**

→ **User-friendly GUI**



Selective Plane Illumination Microscope



Novel microscope to observe zebra fish embryos in vivo:

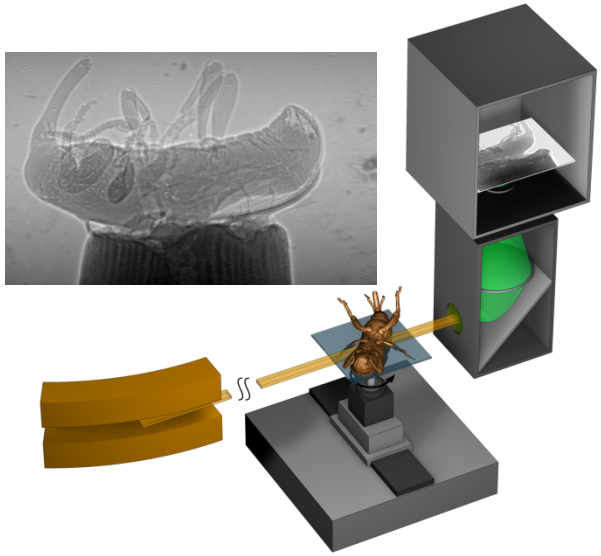
- Very high 3D resolution (~20µm)
- Short data acquisition time (20-30 seconds for full 3D stack)

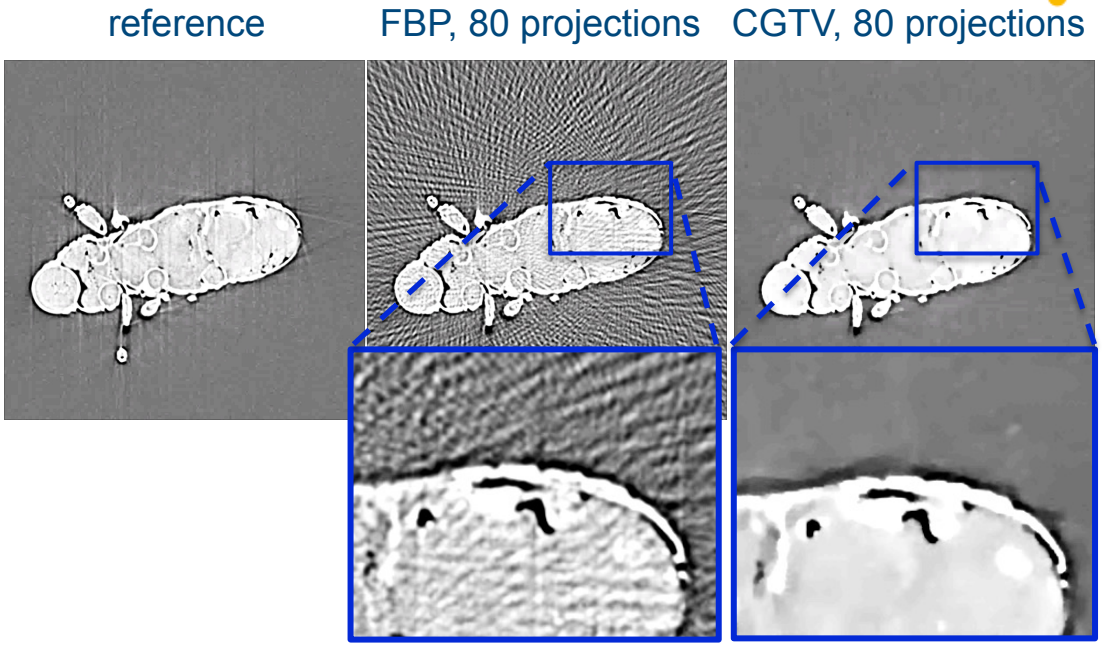
→ **220 TB of data in total was analyzed using a Hadoop cluster**

Publication

G. Nienhaus et al.; An ensemble-averaged digital model of zebrafish embryo development based on light-sheet microscopy with single-cell resolution, **Scientific Reports Nature**, 2015, DOI: 10.1038/srep08601

ANKA Tomography





Challenges

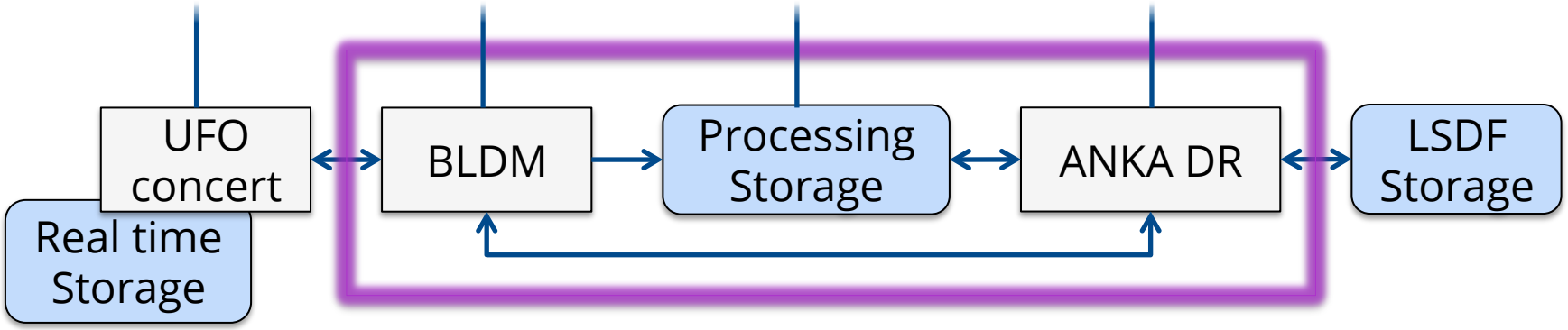
- Low number of projections (100-400)
- High reconstruction quality

→ **New reconstruction method CGTV**

Publication

X. Yang et al. ; TV-based conjugate gradient method and discrete L-curve for few-view CT reconstruction of X-ray in vivo data; **Optics Express**, 2015, DOI: 10.1364/OE.23.005368

ANKA Tomography

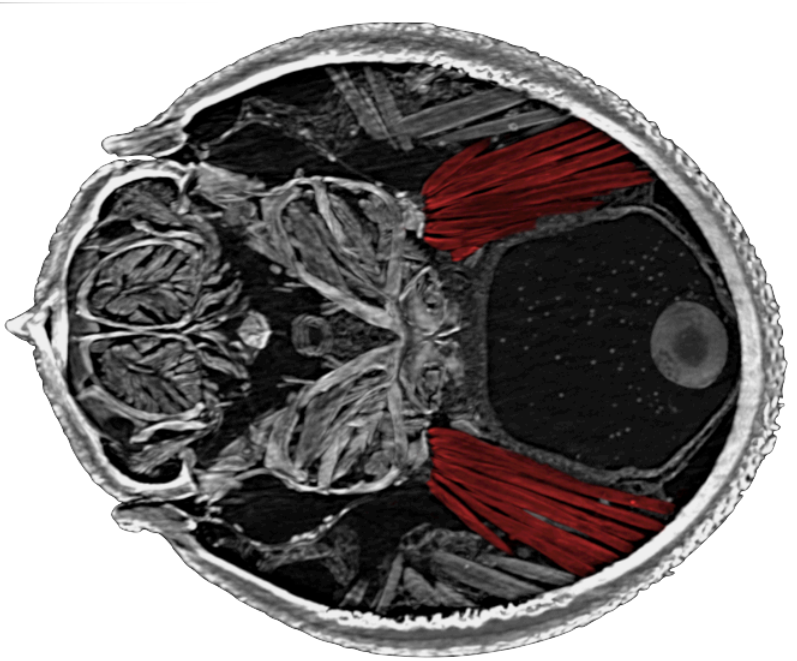


Challenges

- Identification of relevant metadata
- Understand the tomography data workflow

→ **Metadata for raw and preprocessed experiment data defined**

→ **Data workflow defined**



Archeology

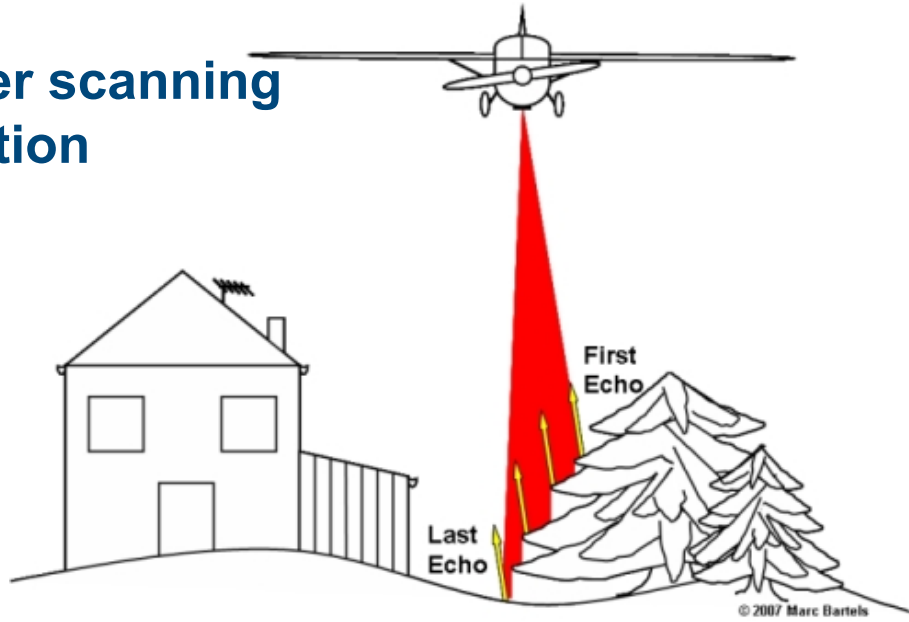
Objectives

- Building a large scale research data infrastructure for managing and processing:
  - laser scanning data
  - archeological information
- Semi-automatic extraction of cultural heritage

→ **Data repository for geospatial laser scanning data with map-based data exploration**

LiDAR Hands on Meeting
Sept. 2, 2014 – Heidelberg

LiDAR Group Open Presentations
Jan. 20.1.2015 – Heidelberg



Archeology

DARIAH & eCodicology

DARIAH-DE

DBpedia Spotlight

Software **W**orkflow for the **A**utomatic **T**agging of Medieval Manuscripts **I**mages (SWATI)

- Data handling, data preprocessing, extraction of layout features
- Interactive visualization of extracted layout features

→ ***Preconditions for data analysis fulfilled***

LSDMA

DARIAH Annotation Cluster

- Extension of DBpedia Spotlight by **GND** (Gemeinsame Normdatei) data
- Investigation of statistical backend of DBpedia Spotlight

→ ***Enhanced annotation results***

eCodicology

eCodicology

LSDMA

Machines and Manuscripts 2015

19.2.2015 - 20.2.2015

Karlsruhe

Der Name der Rose: Informatiker enträtseln alte Manuskripte

heise online

WEB.DE

ONLINE FOCUS

DIE WELT

Conclusion

Metadata are in the center of all projects