DLCL Key Technologies: Status + Future

**Rainer Stotzka, Nick Kepper,
Richard Grunzke, Volker Hartmann, Michael Hausmann, Jürgen Hesser, Thomas Jejkal, Ralph Müller-Pfefferkorn, Halil Pasic, Rasmus Schröder, Danah Tonne, Xiaoli Yang**

LSDMA

Status

5 (+3) subprojects:

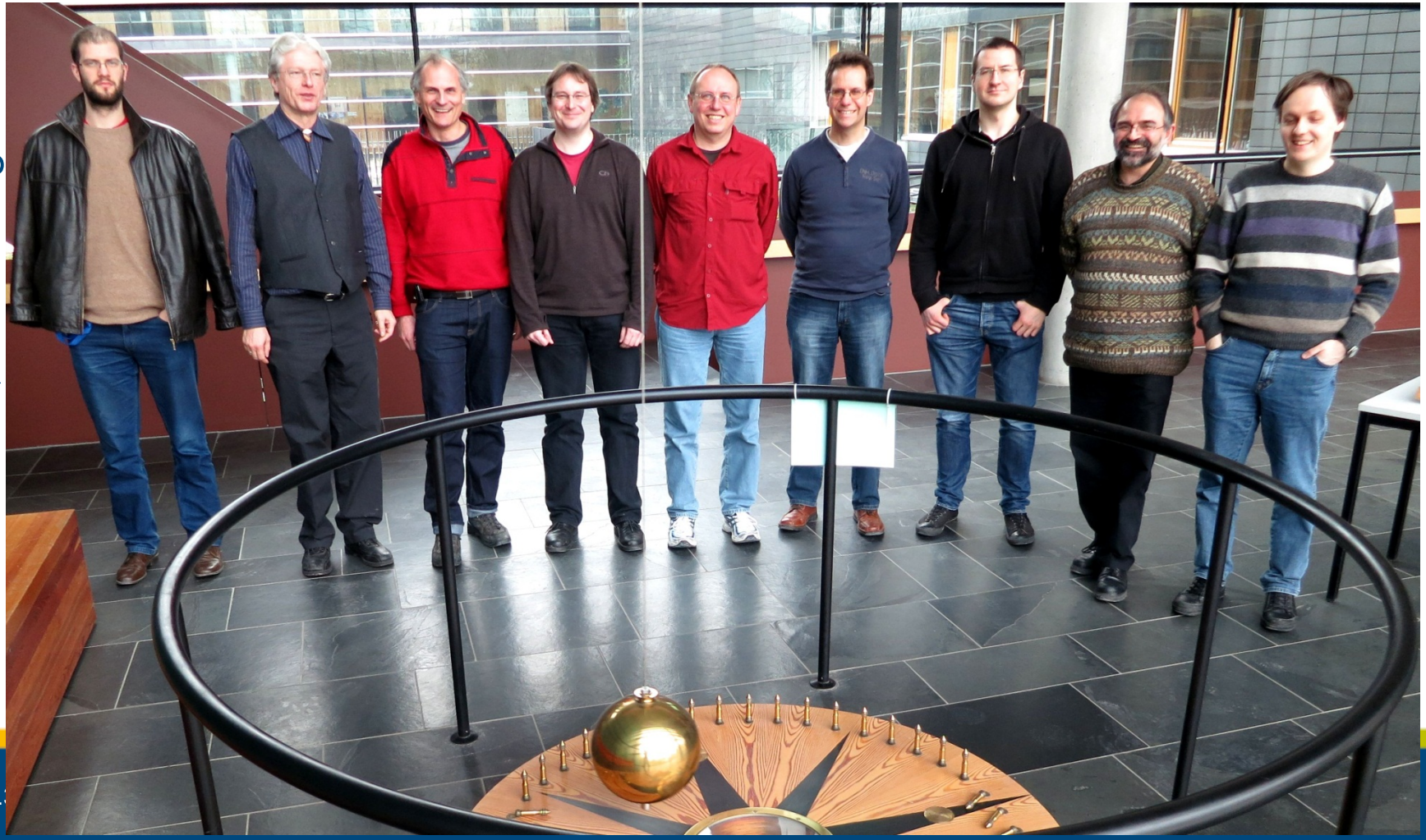
- ***Lightoptical Nanoscopy*** (Heidelberg, Mannheim)
- High Throughput Microscopy: ***Gen Scans***
  (TU Dresden + MPI of Molecular Cell Biology and Genetics)
- High Throughput Microscopy: ***Selective Plane*** Illumination Microscope (Karlsruhe)
- ***Electron*** Microscopy (Heidelberg)
- Material Research: ***Tomography*** Beam Line at ANKA (Karlsruhe)
- (Arts & Humanities): ESFRI ***DARIAH*** + BMBF ***eCodicology***
- *µTomography: in preparation (Institut für Angewandte Materialien, Karlsruhe)*

Organization

Speakers: Rainer Stotzka, Nick Kepper

Monthly phone conferences (each 2. Thursday): November 12, January 13

Face-2-Face meeting:
14. Feb. 2013, Heidelberg



Publications (Paper/Book, Talk, LSDMA)

[1] P. Kilpatrick, P. Milligan, R. Stotzka (eds.), Proceedings of the 21st International Euromicro Conference on Parallel, Distributed and Network-based Processing, IEEE Conference Publishing Services, DOI 10.1109/PDP.2013, 2013

[2]+[3] Danah Tonne, Jedrzej Rybicki, Stefan E. Funk, and Peter Gietz, Access to the DARIAH Bit Preservation Service for Humanities Research Data, 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, pp. 9-15, DOI 10.1109/PDP.2013.12, 2013

[4]+[5] Xiaoli Yang, Thomas Jejkal, Halil Pasic, Rainer Stotzka, Achim Streit, Jos van Wezel, and Tomy dos Santos Rolo, Data Intensive Computing of X-Ray Computed Tomography: 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, pp. 86-93, DOI 10.1109/PDP.2013.21, 2013

[6] Danah Tonne, Rainer Stotzka, Ingtegration von eCodicology in die DARIAH-Dienstewelt, Tagung Digitale Rekonstruktion Mittelalterlicher Bibliotheken, 18.1.2013, eingeladener Vortrag, Trier

[7] Haas, D.; Mexner, W.; Spangenberg, T.; Cecilia, A.; Kopmann, A.; Balzer, M.; Vogelgesang, M.; Pasic, H.; Chilingaryan, S., Status of the ultra fast tomography experiments control at ANKA. 9th Internat.Workshop on Personal Computers and Particle Accelerator Controls (PCaPAC-2012), 2012

[8] Halil Pasic, Managing Large Scale Scientific Data at ANKA, Data Rate Initiative of the HGF Programme PNI, 2.3.2013

[9] Xiaoli Yang, Thomas Jejkal, Halil Pasic, Rainer Stotzka, Achim Streit, Jos van Wezel, Tomy dos Santos Rolo, Algebraic Reconstruction of X-ray Tomography on parallel computing Architecture of LSDF, Data Rate Initiative of the HGF Programme PNI, 2.3.2013

[10] Rainer Stotzka, Marcus Hardt, Facilitating Data Upload, Access and Deployment, eIRG, Amsterdam, 2012

[11] Sutter, M.; Hartmann, V.; Götter, M.; Wezel, J.; Trunov, A.; Jejkal, T. & Stotzka, R. File systems and access technologies for the large scale data facility Remote Instrumentation for eScience and Related Aspects, Springer, 2012, 239-256

[12] Vanscheidt, P.; Rapp, A. & Tonne, D. (2012), 'Storage Infrastructure of the Virtual Scriptorium St. Matthias"Digital Humanities 2012', Poster, 2012

[13] DLCL-Key Aktivitäten Quartal 3/2012, October 2012

[14] DLCL-Key Aktivitäten Quartal 4/2012, February 2013

Publications (Paper/Book, Talk, LSDMA)

[15] Grunzke et al.,A Data Driven Science Gateway for Computational Workflows, in: UNICORE Summit 2012 Proceedings, pages 35-49, 2012

[16] Gesing et al., A Science Gateway Getting Ready for Serving the International Molecular Simulation Community, in: Proceedings of Science, 2012

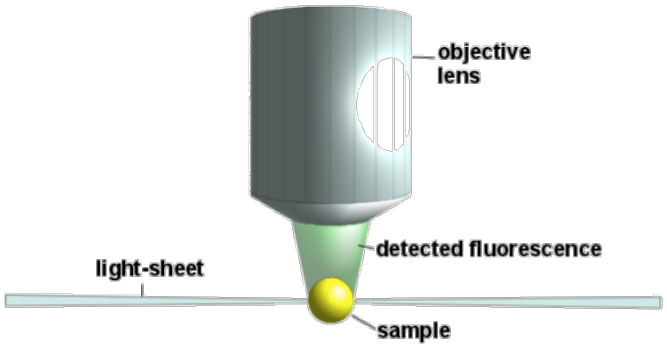
[17] Gesing, Grunzke, et al, A Single Sign-On Infrastructure for Science Gateways on a Use Case for Structural Bioinformatics (2012), in: Journal of Grid Computing, 10:4(769-790)

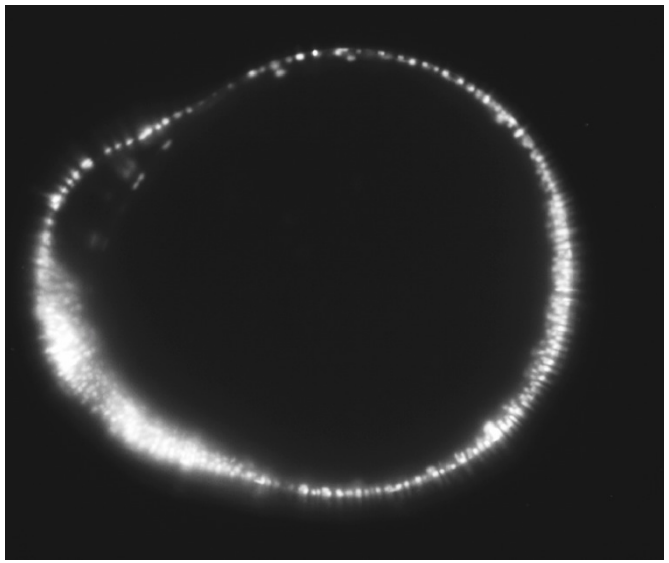
[18] Herres-Pawlis et al., Workflow-enhanced conformational analysis of guanidine zinc complexes via a science gateway, Studies in Health Technology and Informatics, 175:142-151, IOS Press., 2012

[19] Pyatykh, Stanislav, Zheng, Lei and Hesser, Jürgen (2012), "Efficient method of pixel neighborhood traversal", Journal of Visual Communication and Image Representation, 23, 1: 719-728, 2012.

[20] Teuber, T., Remmele, S., Hesser, J. and Steidl, G. (2012), "Denoising by second order statistics", Signal Processing, 92, 12: 2837-2847, 2012.

Single Plane Illumination Microscope





Novel microscope to observe zebra fish embryos in vivo:

- Very high 3D resolution
- Short data acquisition time

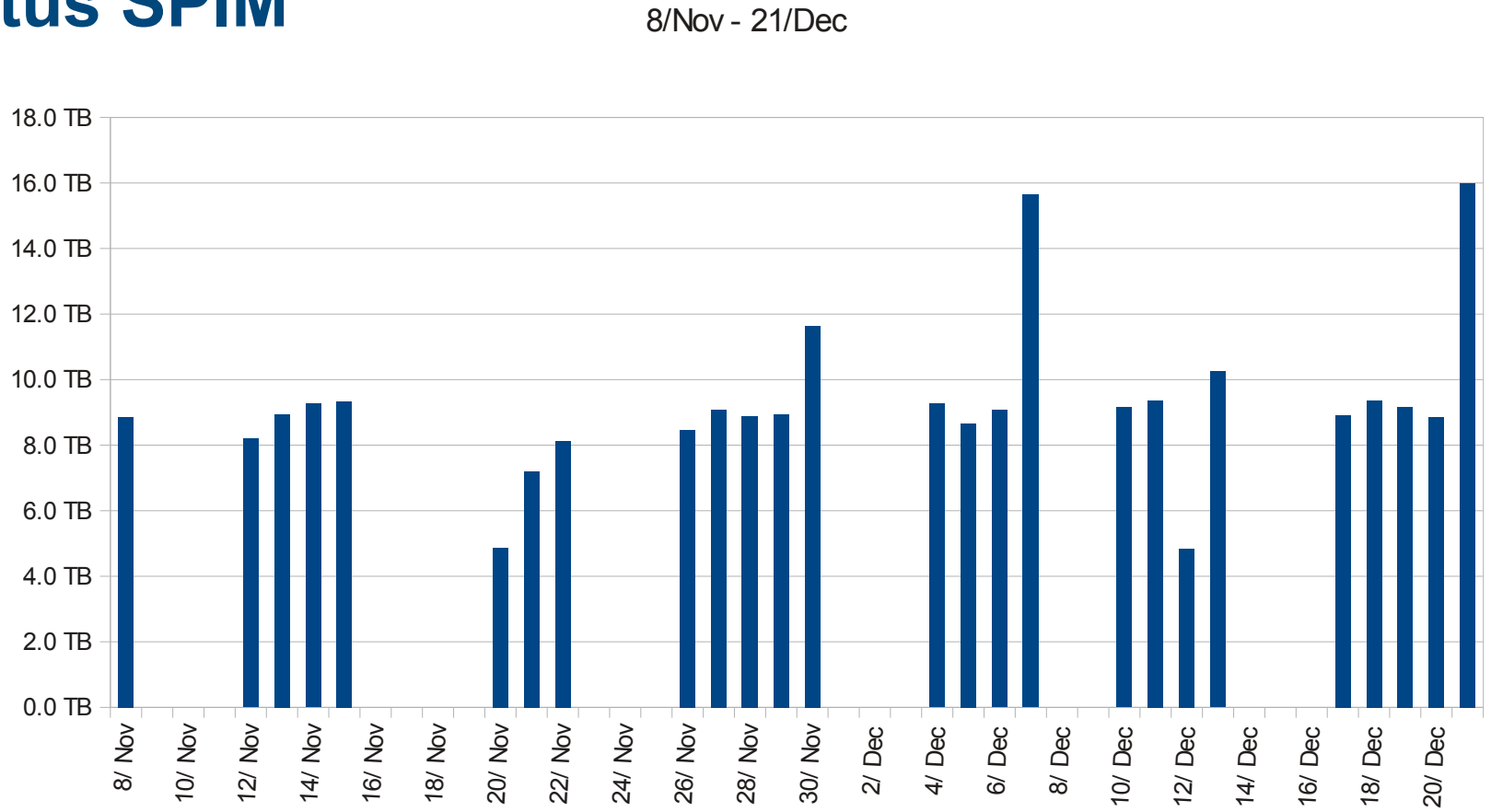
Each day one observation of the development of an embryo:

Data: 7 TB/10 h → ∑ 300 TB

Challenges:

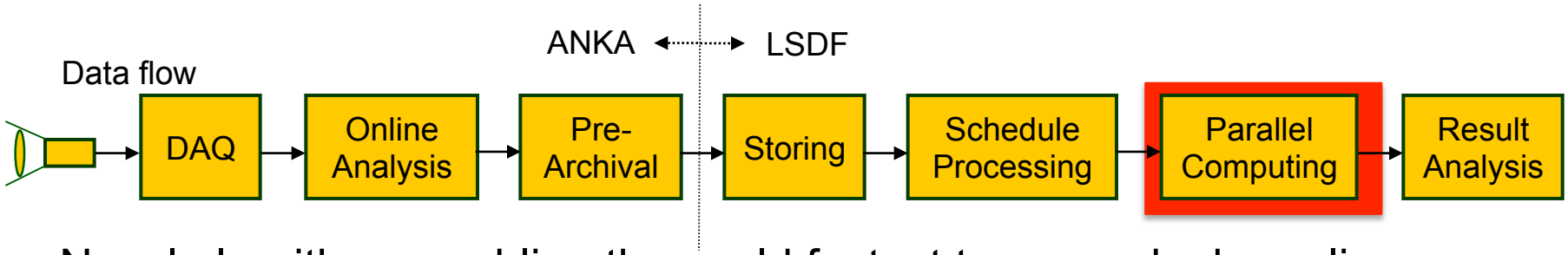
- Meta data extraction
- ***Data transfer*** source to LSDF > 400 MB/s
- Data intensive ***analysis*** < 24 h: registration, fusion, segmentation, tracking, statistics

Status SPIM



Transfer:		**Transfer rate:**	
Total amount:	253 TB	Maximum:	442 MB/s
# Files:	53,333	Minimum:	366 MB/s
Total transfer time:	7d 15h	***Average:***	***403 MB/s***

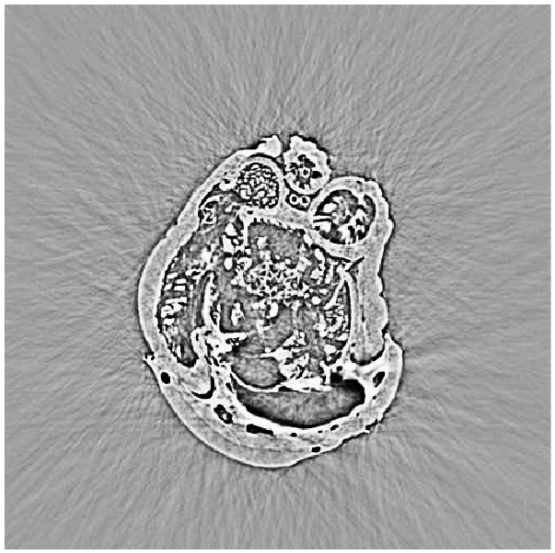
Material Research: Tomography Beam Line at ANKA



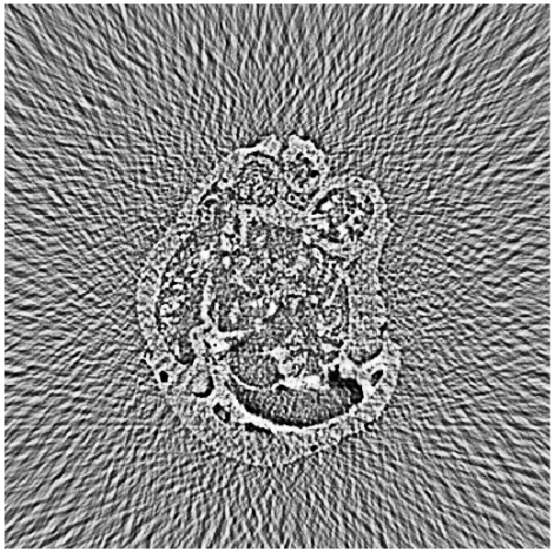
Novel algorithms enabling the world fastest tomography beamline

- Compressive Sampling Algebraic Reconstruction (***CS-ART***)

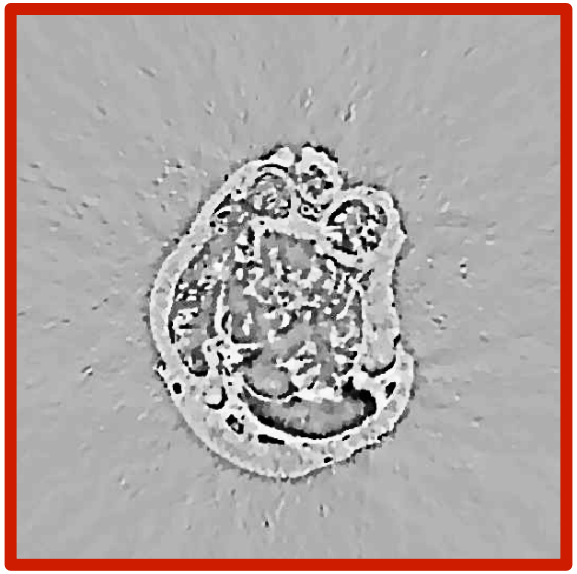
Standard: 1500 projections



Standard: 60 projections



CS-ART: 60 projections

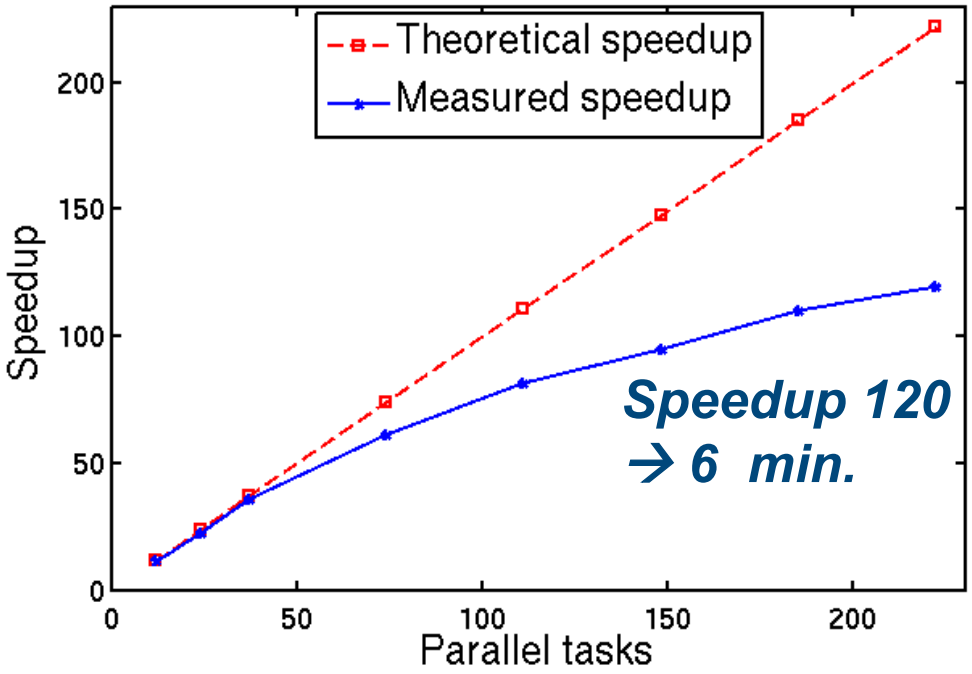


Material Research: Tomography Beam Line at ANKA

- CS-ART implemented in Matlab
- Computational expensive → ***11 h !!!! >> 1 min.***

- Parallelized using DSIT DIC, speedup of 120

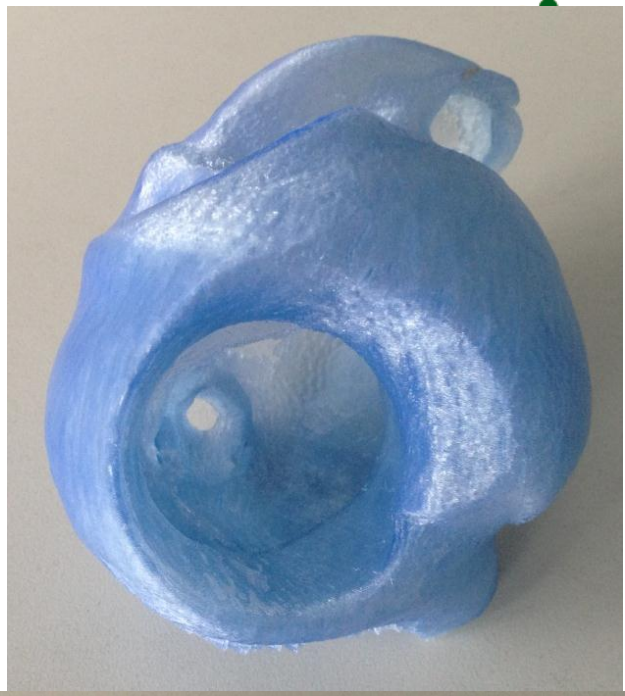
→ CS-ART: ***better*** images
→ ***Automatic*** DIC workflow 2014
→ ***Enables*** world's fastest tomography beamline
→ ***General***, not limited to ANKA



Material Research: Data to Finger

Printing reconstructed volumes in 3D

HECTOR Seminar: highly skilled MINT students

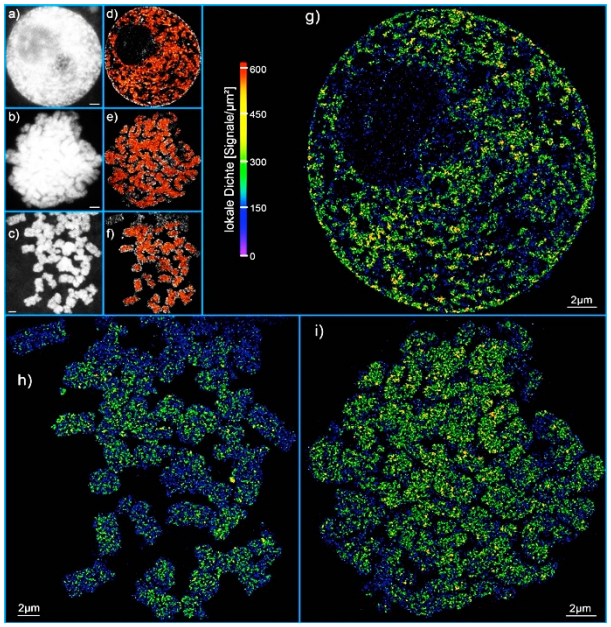


Lightoptical Nanoscopy

Projects considered:

- Changes density and packaging in hetero- and euchromatin by radiation damage via high resolution SPDM images
- Analysis genome architecture of breast cancer marker Her2/neu as marker for the tumorigenesis
- Micro- and nanoarchitecture of Her2/neu receptor clusters before and after amplification of genes in tumor cells

SPDM of nucleosomes



Tools:

- ***Automatic analysis*** with parallelized MatLab programs: localization, distribution, and classification

Visualization:

- Scientific ***visualization*** with ParaView

Important requirements:

- Data storage facilities
- Data life cycle definition
- Application dependent evaluation
- Meta data handling

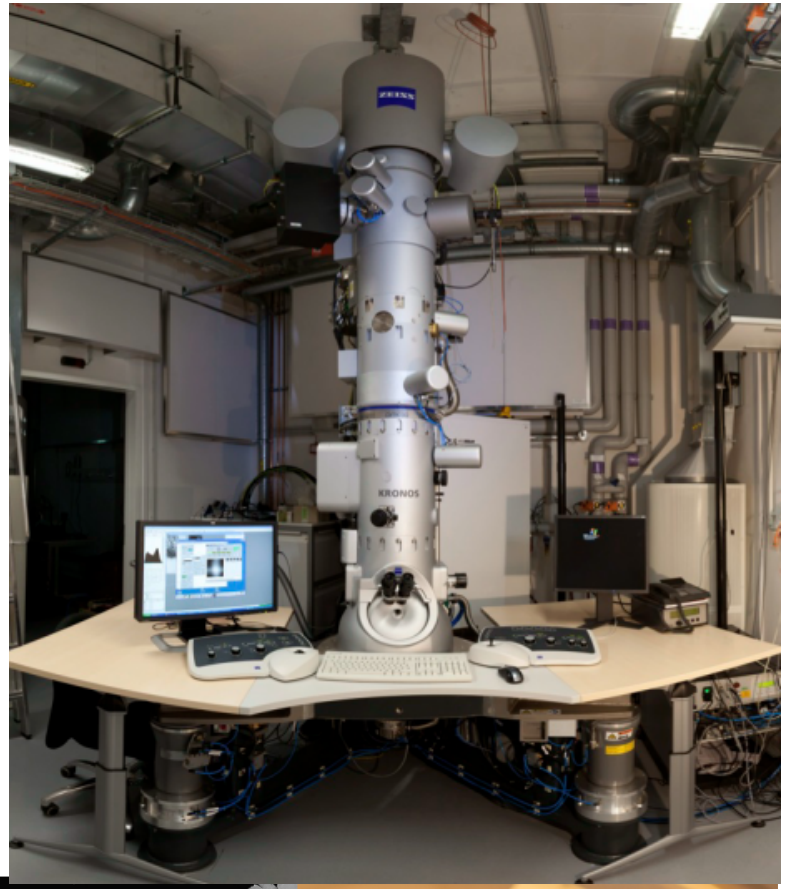
Electron Microscopy

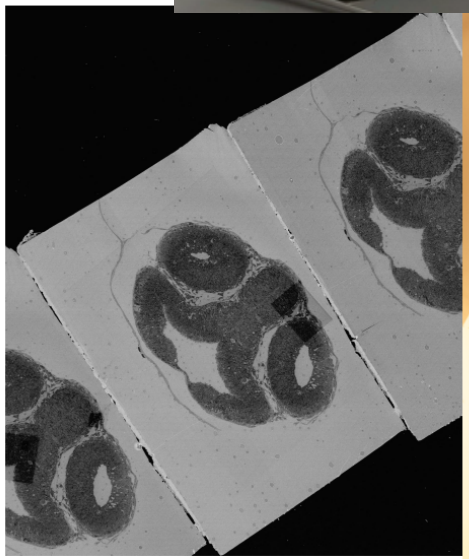
To do:

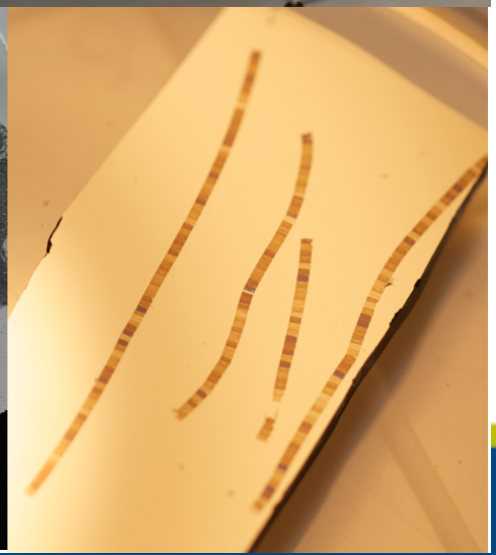
- Acquiring data: automation, workflow, electronic lab book (EMEN2)
- Image data data base structure (LEGINON)
- Storage (local for processing?)
- Processing pipeline (registration, segmentation) needs to be developed
- Data browsers, 3D-Viewers? Data reduction?
- Archiving?

30 GB / image
Depth in 3rd dimension → 10^5 images
1 eye: 3 PB

Image sources: Rasmus R. Schröder
CellNetworks, BioQuant, CAM, Universität Heidelberg



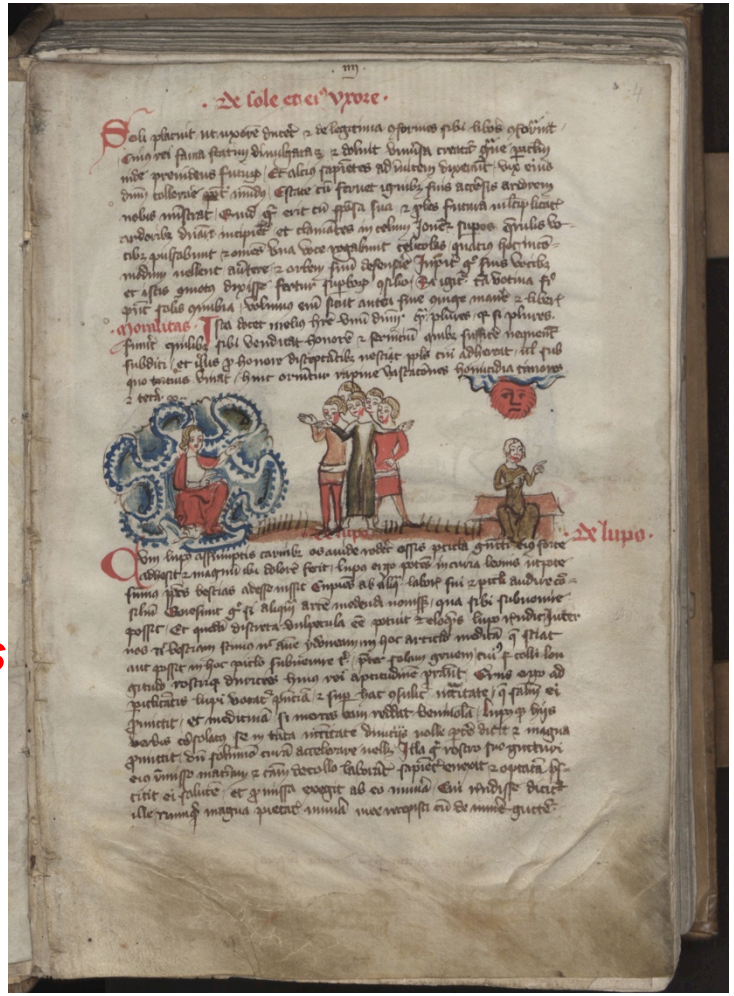




DARIAH + eCodicology

DARIAH-DE
Digital Research Infrastructure for the Arts and Humanities

ESFRI project, preservation infrastructure



eCodicology: BMBF joint project (Karlsruhe, Trier, Darmstadt), PhD students + ***code camps***

- Automatic ***analysis*** of 500 medieval manuscripts, 150,000 pages

Most important requirements:

- Meta data handling
- Reproducibility
- Performant data access
- Data intensive workflows

DSIT Services

- Federated Data Access
- Meta Data Catalogues and Repositories
- Data Intensive Computing

Some GENERAL Results

Similarities found in nearly all sub-projects:

- ***Meta data is important !***
- Automatic meta data extraction required
  File format → standard meta data format

Experiment view

- EMEN2
- „by hand“
- proposal DB

Instrument view

- OME
- Tango
- Labview
- LEGINON

- Data provenance

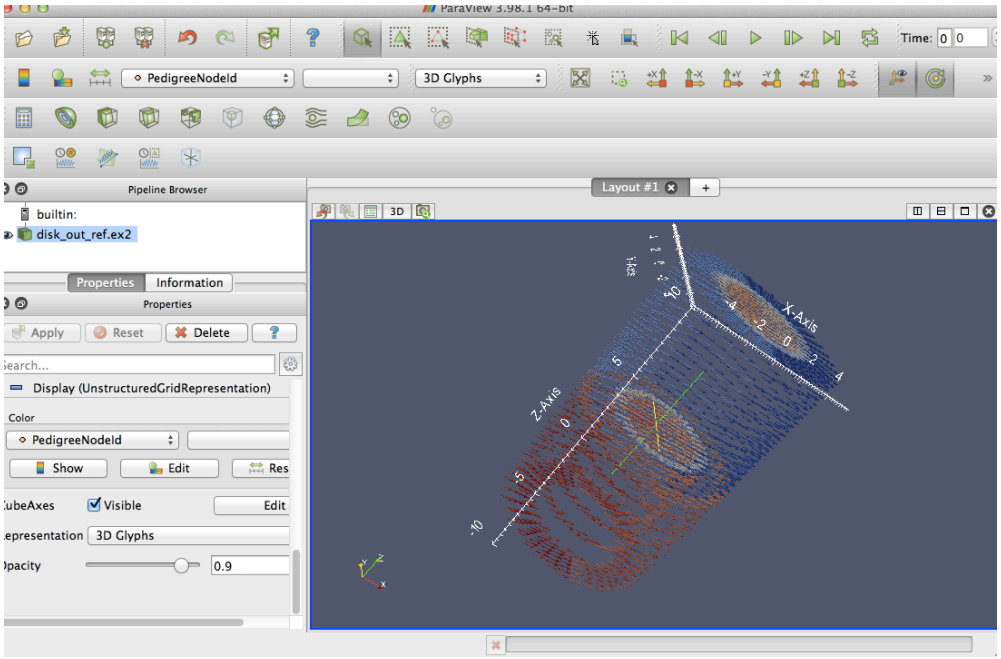
- Where to store the data?
  → KIT LSDF as the platform for tool development & testing
  → TU Dresden

Some GENERAL Results

Established knowledge exchange

Common tools defined:

- ***Data transfer client*** requires federated AAI !!!
- ***Meta data*** management tools
- Data ***analysis*** tools:
  - Digital image processing lib. consisting of basic functions, e.g. denoising
  - ***Visualization***



Additional Funding

- **eCodicology**: 1 (+2) PhD positions
  Relevance to LSDMA:
  - Data life cycle and data sharing
  - Meta data
  - Image analysis
  - Data intensive computing

- 1 PhD student **scholarship** (4 years)
  Relevance to LSDMA:
  - Algorithms for volume reconstruction & analysis
  - Data intensive computing

- **Hector** Seminar: 1 year's students project
  Relevance to LSDMA:
  - Novel data representations

- **MoSGrid** (→ ER-flow : 2 years)
  Relevance to LSDMA:
  - Science gateway
  - Workflow management & sharing
  - Data interoperability in workflows

- **MoSGrid** (→ SCI-BUS: 3 years)
  Relevance to LSDMA:
  - Science gateway
  - Meta data repository
  - Result visualization

- BMBF Hausmann: „Fokussierte Fluoreszenzmarkierung (COMBO-FISH)"

- BMU Hausmann: „3D/4D Architektur von chromosomalen Bruchpunktregionen in Zellkernen nach Bestrahlung von Krebs- und nicht-Krebszellen"

DLCL-Key: Planned Publications

- Mikut, R.; Geurts, P.; Hamprecht, F.; Kausler, B. X.; Marée, R.; Mikula, K.; Pantazis, P.; Ronneberger, O.; Stotzka, R.; Strähle, U. & Peyriéras, N., Automated processing of zebrafish imaging data - a survey, *Zebrafish,* **13.3.2013**
- J. Lux, …, M. Hausmann, Reconstruction and optical sequencing of a DNA repeat expansion unit by location nanoscopy.
- Xiaoli Yang, A New Forward Model for Improved Tomographic Reconstruction, IEEE Big Data || PDP 2014 || Tomography, July 2013
- Halil Pasic et al., Managing Large Scale Scientific Data at ANKA
- Grunzke et al., Standard-based Metadata Management for Molecular Simulations, 2013
- Krüger, Gesing, Grunzke, et al. , The MoSGrid Science Gateway, 2013
- Herres-Pawlis et al.,User-friendly Workflows in Quantum Chemistry, 2013
- Pyatykh, Stanislav, Zheng, Lei and Hesser, Jürgen (2013), "Fast noise variance estimation by principal component analysis", acc. for pub. IS&T/SPIE Electronic Imaging 2013, SPIE.

Conclusions

a. DLCL-Key: broad ***spectrum*** of scientific projects

b. Heterogeneous team, highly motivated

c. DSIT services will be required

d. Results: established knowledge exchange
→ common ***tools*** defined

Next steps:

→ Establish software exchange, common tool development and integration