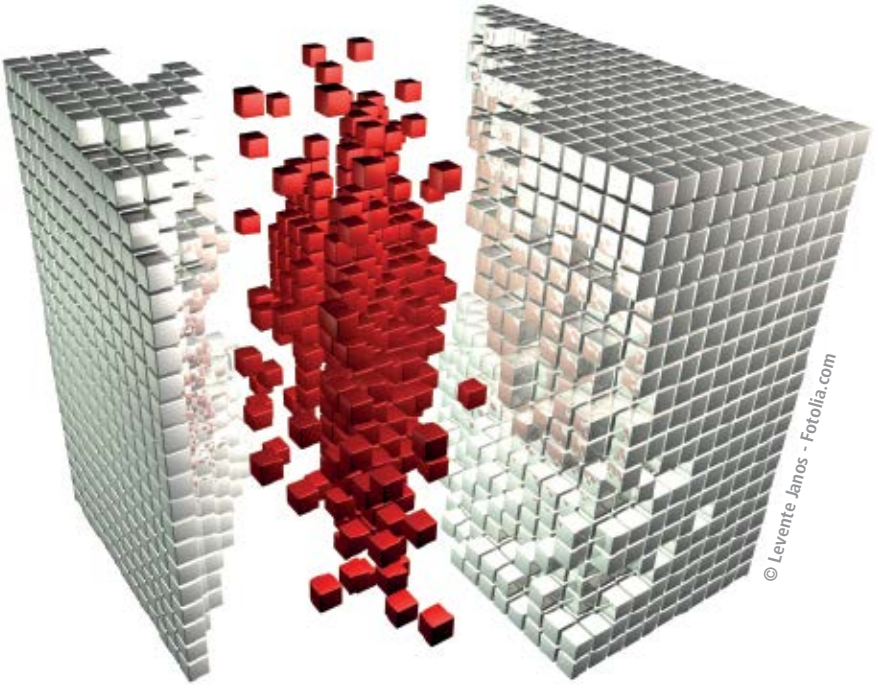
Big Data in Microscopy

Data Analysis & Management in Complex Microscopy Experiments

We introduce a strategy to deal with big data in microscopy and to provide laboratories access to image analysis of large multidimensional image data sets. Modern microscopes acquire big amount of data and provide much deeper insight in detailed biological issues. Progress in image acquisition shifted the bottle neck in research to image data visualization, analysis, computing capacity and availability of experts. One solution is cooperation with a service provider offering flexible image analysis solutions via online data transfer or sending of hard disks via mail.



David Haumann,
Head of Sales, da-cons GmbH,
Eggenstein-Leopoldshafen,
Germany

Why Is Image Data in Microscopy so Big?

Modern X-Ray, Electron and Light Microscopes generate large volume datasets for multidimensional illustration of samples - e. g. the monolithic Digital Scanned laser Light sheet fluorescence Microscope (mDSLM), the Selective Plane Illumination Microscopy (SPIM), developed by Prof. EHK Stelzer at the Buchmann Institute for Molecular Life Sciences (BMLS), or the new Light Sheet Fluorescence Microscope (LSFM) from the company Carl Zeiss [1, 2].

Data volume growth is extraordinary, because samples are illuminated and recorded with multicolor channels at the same time. For every channel, z-stacks of 2D images are acquired for every single time point. Furthermore, the sample can be illuminated and observed from different angles. These image data sets need to be processed via registration and fusion into a three dimensional image, called volume data set. These volume data sets are 2 GB-100 GB per time point in size.

A serial sequence of volume data sets allows researchers to observe development and dynamic changes in living biological samples in three dimensions. Dynamic processes can be tracked in 3D videos and cell behavior becomes observable. Changes in position, size, number, distances and migration velocity can be measured [3].

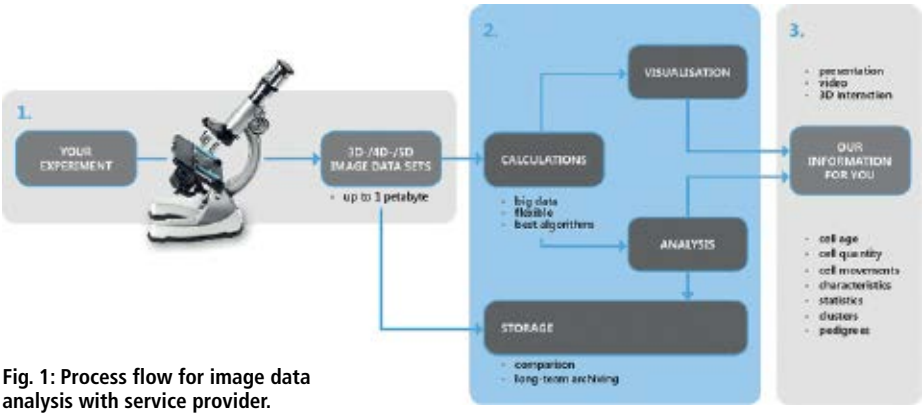


Fig. 1: Process flow for image data analysis with service provider.

Information about size and shape of cells, nuclei and sub-cellular structures in roots of plants [4], *in vivo* embryogenesis [5] or 3D-microtumor models of cleared objects [6] can be obtained. Via grid recording, a technique that samples and fuses partial images of larger samples, it is possible to visualize large three dimensional areas in the mm^3 scale. This is interesting for example to visualize effects of cosmetics on structure and capillary network in human skin.

Support with Handling & Analysis of Large Volume Data Sets

Single volume data sets and especially time series of large volume data sets (time lapse) are challenging concerning data handling, computing and extraction of all available information. Furthermore it is difficult to visualize and archive such huge data sets over long time periods. During long observation periods up to hundreds of terabyte data can be generated.

If man power, expertise in multidimensional image analysis or computing capacity is not adequate on-site, the produced data can be transferred to a service

Keywords

Multidimensional Big Data Sets, Image Analysis, Visualization, Storage, Data Management

provider as shown in figure 1. The researcher generates large image data sets via microscope or X-ray microscopy, µMRT, high content screening, simulations etc., symbolized in grey box 1. (left side). Collaboration with service provider offers the researcher access to computing capacity, analysis know how with complex algorithms, visualization, data handling and archiving (blue box 2. in the middle). The researcher gets back the desired extracted information from the image data sets (grey box 3. right side).

Cooperation with service providers allows laboratories with limited budget and companies to analyze huge data sets. Data need to be transferred from research location to the service provider.

Transfer of Big Image Data Sets to a Service Provider

Transportation of big data is of central importance in data management. Data transport depends on a reliable file transfer and a complementary meta-data framework to keep track of the transferred data. Conventional data transfer works well only for low data volumes, small file sets and a reliable internet connection. However manually controlled transfer of large amounts of data is error prone and consumes an inordinate amount of time of the experimenter.

Alternatively data can be moved by transfer technologies such as GridFTP used to move petabytes of data in the Worldwide Large-Hadron-Collider Computing Grid (WLCG) and the Large Scale Data Facility (LSDF) at the Karlsruhe Institute of Technology (KIT). An efficient way of transporting large datasets can be shipping storage media.

Analysis of Large Volume Data Sets

Image analysis usually means to process hundreds to thousands of files. Computer facilities such as the LSDF allow direct access to the computing farm. For special cases the data can be loaded into the hadoop file system pairing fast data access from commodity worker nodes and massive cluster computing with an excellent price performance ratio (fig.1). Development at da-cons and KIT aims to improve data management and its interaction with the workload management and storage systems of the LSDF.

Example: Analysis of a Spheroid

Figure 3 shows a cellular spheroid of BxPC3 human pancreatic cancer cells. For sample preparation, nuclear staining with DRAQ5 and volume data acquisition see [8, 9]. Image acquisition has been done with a selective plane illumination microscope (SPIM).

On the left side in figure 3a, the original data set with lots of noise in the image-background is shown. The noise was filtered in the volume data set (not shown) followed by segmentation of the nuclei in figure 3b. The term segmentation describes an analysis process with special algorithms that identifies objects - in this case the nuclei including position and morphological features (e.g. volume, surface area, roundness, relations and distances between nuclei, standard

deviation of nuclei, average volume of nuclei [10].

In figure 3c two 3D prints of the spheroid with different size and color are shown. The original spheroid volume is 135 thsd µm³. The red 3D print is 6 cm in diameter with a volume of 113 cm³, about one billion times larger than the original spheroid in figure 3a. In this way, researchers can hold their objects in the hand.

Conclusion

New image acquisition methods like light sheet-based fluorescence microscopy, X-ray microscopy and high-content screening methods generate large image volume data sets, imposing high requirements on storage capacity, computing capacity, know how for data handling and analysis. Cooperation with service provider enables every smaller and large laboratory or company to analyze huge image data sets, overcoming the lack of computing capacity, man power shortage, as well as getting access to a variety of analysis algorithms.

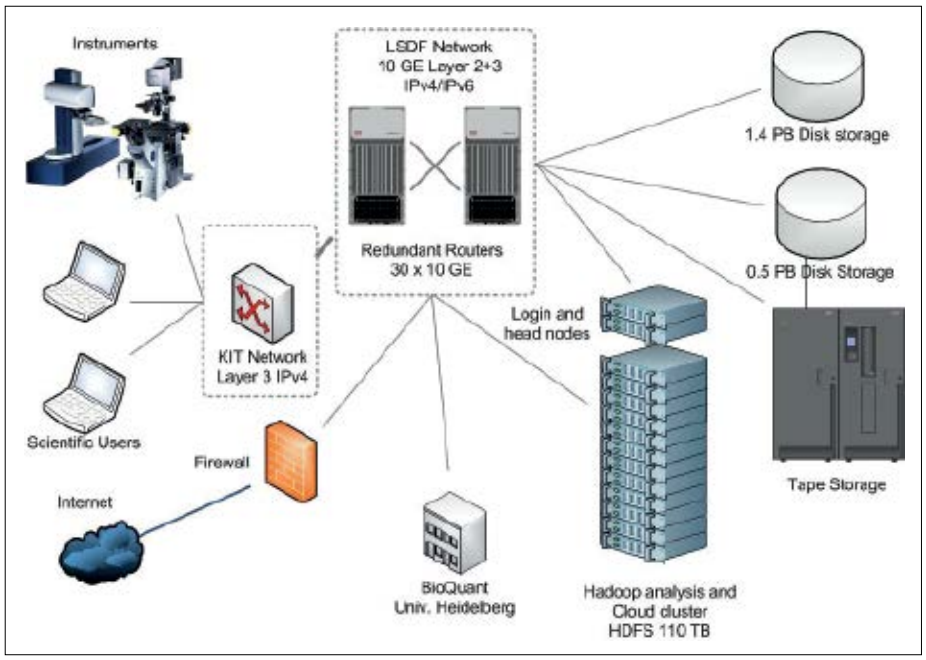


Fig. 2: The LSDF at KIT tightly integrates massive storage and archives with online computing.

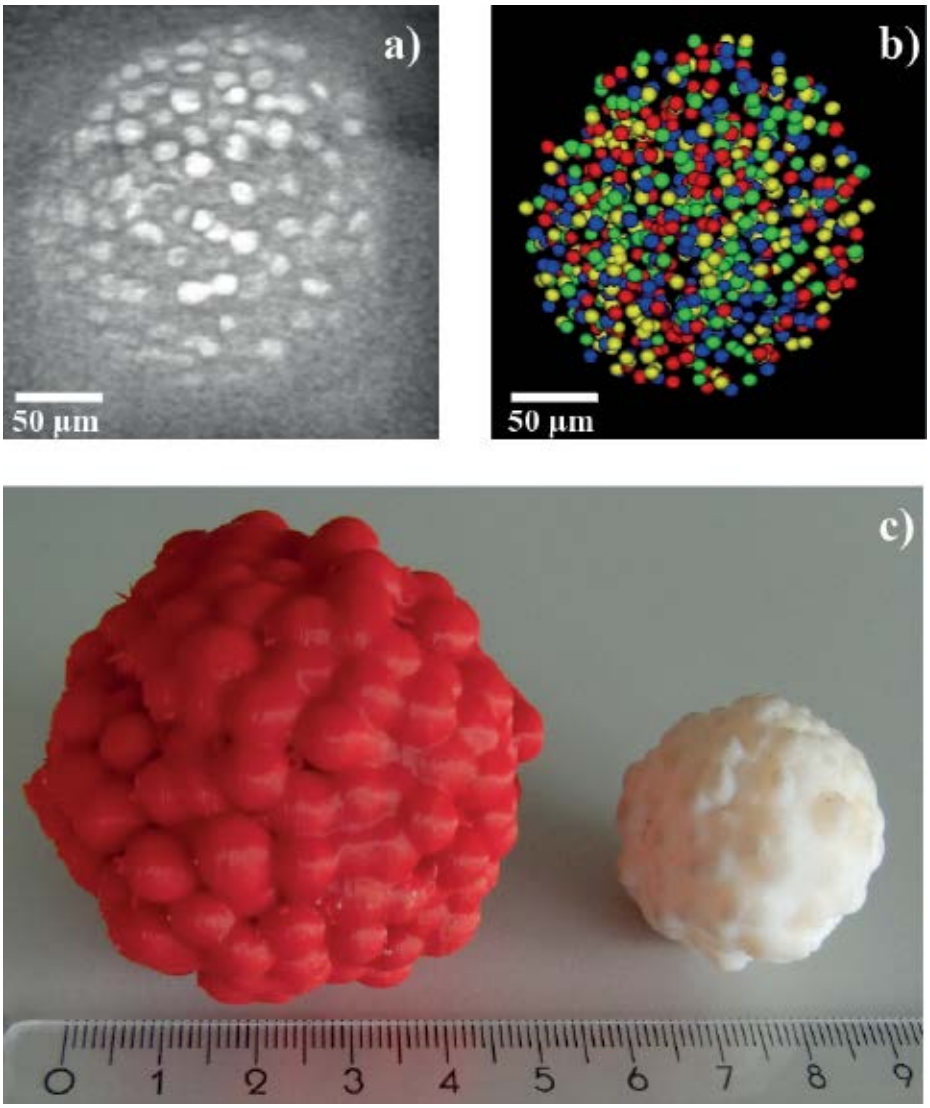


Fig. 3: Data set from a spheroid a) original, b) processed and c) 3d-printed.
3D animation movie of spheroid:

References

[1] Stelzer E.H.K.: Journal of Microscopy, 15–24 (1998)

[2] Huisken J. *et al.*: Science, 1007–1009 (2004)

[3] www.youtube.com/watch?v=VwoSMTZ1pz4

[4] Maizel A. *et al.*: The Plant Journal, 377-385 (2011)

[5] Keller P.J.: Methods (2013)

[6] Weber P. *et al.*: Journal of Biomedical Optics, 1015081–1015085 (2012)

[7] Garcia A.O. *et al.*: International Conference on Digital Information Management, 235–241 (2011)

[8] Pampaloni F. *et al.*: Recent Patents on Biotechnology 3.2, 103–117 (2009)

[9] Pampaloni F. *et al.*: Cell and Tissue Research, 1–17 (2013)

[10] www.youtube.com/watch?v=AnnX6bXE0Y8

Authors

David Haumann, da-cons GmbH, Germany
Jos van Wezel, Karlsruhe Institute of Technology, Steinbuch Centre for Computing, Karlsruhe, Germany
Thomas Schenker, da-cons GmbH, Germany

Contact
David Haumann
da-cons GmbH
Eggenstein-Leopoldshafen, Germany
info@da-cons.com
www.da-cons.com