

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

The NOVA project: maximizing beam time efficiency through synergistic analyses of SR& μ ;CT data

Sebastian Schmelzle
Michael Heethoff
Vincent Heuveline
Philipp Lösel
Jürgen Becker
Felix Beckmann
Frank Schluenzen
Jörg U. Hammel
Andreas Kopmann
Wolfgang Mexner
Matthias Vogelgesang
Nicholas Tan Jerome
Oliver Betz
Rolf Beutel
Benjamin Wipfler
Alexander Blanke
Steffen Harzsch
Marie Hörnig

SPIE.

The NOVA project - maximizing beam time efficiency through synergistic analyses of SR μ CT data

Sebastian Schmelzle^{*a}, Michael Heethoff^a, Vincent Heuveline^b, Philipp Lösel^b, Jürgen Becker^c, Felix Beckmann^d, Frank Schluenzen^e, Jörg U. Hammel^{df}, Andreas Kopmann^g, Wolfgang Mexner^h, Matthias Vogelgesang^g, Nicholas Tan Jerome^g, Oliver Betzⁱ, Rolf Beutel^f, Benjamin Wipfler^f, Alexander Blanke^j, Steffen Harzsch^k, Marie Hörnig^k, Tilo Baumbach^{l,m}, Thomas van de Kamp^{**l}

^aEcological Networks, Department of Biology, Technische Universität Darmstadt, Schnittspahnstr. 3, D-64287 Darmstadt; ^bEngineering Mathematics and Computing Lab (EMCL), Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Im Neuenheimer Feld 205, D-69120 Heidelberg; ^cInstitute for Information Processing Technologies (ITIV), Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, D-76131 Karlsruhe, Germany; ^dInstitute of Materials Research, Outstation at DESY, Helmholtz-Zentrum Geesthacht, Notkestr. 85, Building 66, D-22607 Hamburg; ^eDeutsches Elektronen Synchrotron DESY, Notkestraße 85, D-22607 Hamburg; ^fInstitute of Systematic Zoology and Evolutionary Biology with Phyletic Museum, Friedrich-Schiller-University Jena, Erbertstr. 1, D-07743 Jena; ^gInstitute for Data Processing and Electronics (IPE), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany; ^hInstitute for Beam Physics and Technology (IBPT), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany; ⁱEvolutionary Biology of Invertebrates, Institute of Evolution and Ecology, Universität Tübingen, Auf der Morgenstelle 28E, D-72076 Tübingen; ^jInstitute of Zoology, University of Cologne, Zülpicher Straße 47b, D-50674 Köln; ^kCytology and Evolutionary Biology, Zoological Institute and Museum, University of Greifswald, Soldmannstr. 23, Haus 6.1, D-17489 Greifswald; ^lLaboratory for Applications of Synchrotron Radiation (LAS), Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, D-76131 Karlsruhe, Germany; ^mInstitute for Photon Science and Synchrotron Radiation (IPS), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany

ABSTRACT

Beamtime and resulting SR μ CT data are a valuable resource for researchers of a broad scientific community in life sciences. Most research groups, however, are only interested in a specific organ and use only a fraction of their data. The rest of the data usually remains untapped. By using a new collaborative approach, the NOVA project (Network for Online Visualization and synergistic Analysis of tomographic data) aims to demonstrate, that more efficient use of the valuable beam time is possible by coordinated research on different organ systems. The biological partners in the project cover different scientific aspects and thus serve as model community for the collaborative approach. As proof of principle, different aspects of insect head morphology will be investigated (e.g., biomechanics of the mouthparts, and neurobiology with the topology of sensory areas). This effort is accomplished by development of advanced analysis tools

*sebastianschmelzle@gmail.com; phone: +49 6151 16 75415; www.econetlab.net | **thomas.vandekamp@kit.edu; phone: +49 721 608-28655; www.kit.edu

for the ever-increasing quantity of tomographic datasets. In the preceding project ASTOR, we already successfully demonstrated considerable progress in semi-automatic segmentation and classification of internal structures. Further improvement of these methods is essential for an efficient use of beam time and will be refined in the current NOVA-project. Significant enhancements are also planned at PETRA III beamline p05 to provide all possible contrast modalities in x-ray imaging optimized to biological samples, on the reconstruction algorithms, and the tools for subsequent analyses and management of the data. All improvements made on key technologies within this project will in the long-term be equally beneficial for all users of tomography instrumentations.

KEYWORDS: Synchrotron X-ray micro computed tomography, cooperative data analysis, insect head, semi-automatic segmentation, web portal for scientific data, 3D web visualization, visual data browsing, Virtual Reality for scientific data, interactive interfaces, data catalog

1. INTRODUCTION

In recent years, synchrotron-based X-ray microtomography (SR μ CT) has become an invaluable technique for the 3D examination of small animals like insects and other arthropods^[1-7]. Consequently, access to dedicated imaging beamlines at synchrotron radiation facilities is requested by more and more competing research groups. Competition, however, does usually not concern the scientific questions, but rather the limited amount of available beamtime.

In this context, it is highly unfortunate that in most cases only a small part (e.g. a particular organ system) of a scanned organism is evaluated after the experiment. Different morphological structures, and the underlying SR μ CT data, are usually investigated independently and isolated from each other—although apparently linked as complex evolutionary unit^[8]. Little to none synergy thus emerged from obtained SR μ CT data sets so far and an enormous potential remains untapped.

1.1 The NOVA project

NOVA (Network for Online Visualization and synergistic Analysis of tomographic data) is a project funded by the German Federal Ministry of Education and Research (BMBF) and successor of the ASTOR project (<https://ufo.kit.edu/dis/index.php/project/astor/>), during which a virtual infrastructure for the remote analysis of tomographic data was established at KIT^[9]. Further, algorithms for the semi-automatic segmentation of tomographic datasets were developed and successfully employed by the partners^[10,11].

In NOVA, the successful ASTOR technologies are further enhanced and a similar portal for virtual analysis of tomographic data was established at DESY. By using a new collaborative approach, which includes the coordination of research on different organ systems and a regulation of the data usage by a common data policy, NOVA aims to facilitate a more efficient use of the valuable beam time at synchrotron beamlines.

NOVA is a joint research project of the Technische Universität Darmstadt (TUD), the University of Heidelberg (UHD), and the Karlsruhe Institute of Technology (KIT) with associated technical partners from the Helmholtz Zentrum Geesthacht (HZG), and the KIT groups Institute for Data Processing and Electronics (IPE; KIT), and Ångströmquelle Karlsruhe (ANKA; KIT). The associated biological project partners from the Universities of Cologne, Tübingen, Jena, Greifswald, and Darmstadt, and the Karlsruhe Institute of Technology (KIT) cover different scientific aspects of insect morphology and thus serve as a model community. Alexander Blanke combines SR μ CT data with innovative analysis methods to investigate taxonomical and evolutionary questions^[12-15] as well as functional morphological problems of insect head morphology^[16-19]. At the moment, he works on the biomechanics of insect bite dynamics to understand optimization processes and convergence in their evolution^[20]. Oliver Betz focuses on the morphology, evolution, and ecology of insects. He investigates ecological and functional morphology, ultrastructure, and bionics using SR μ CT, live-

radiography, and electron and light microscopy^[1, 21-24]. The main research topics of Rolf Beutel are large scale taxonomy and evolution of the larger hexapod groups (Polyneoptera, Acercaria, Holometabola, Coleoptera, Diptera)^[8]. He prioritizes modern techniques in a phylogenetic context. Benjamin Wipfler investigates cephalic and thoracic structures of insects considering functional, phylogenetic, and evolutionary aspects^[25]. His main research topic is the Polyneoptera (roaches, termites, mantids, and earwigs). Steffen Harzsch works on neuroanatomical and neuroethological research of different arthropods (hexapods, myriapods, and crustaceans)^[26,27]. He uses a wide range of imaging techniques with the focus on immunohistochemistry and μ CT^[28,29]. Besides general brain morphology in an evolutionary context, he is especially interested in the olfactory and visual system. Thomas van de Kamp investigates the functional morphology of insects (particularly joint systems). He is interested in the application of X-ray imaging for entomology^[30-33] and paleontology^[34-40] as well as digital visualization and 3D reconstruction of biological samples^[41-43]. The TU Darmstadt is responsible for the coordination of the project, but is also involved in data acquisition, and analysis.

1.2 Use case: collaborative research on the insect head

The scientific use case of the project is the insect head, which provides data relevant for different fields of research. It is (A) the site of food ingestion carrying the mouthparts which can be heavily modified depending on the type of food affecting associated musculature, that is situated in the head as well. (B) The head houses the animal's brain which analyses sensory input from the eyes and antennae and drives the animal's behavior^[44]. (C) These systems are not only interesting from a morphological perspective, but also in a comparative, phylogenetic and evolutionary context, and (D) from a functional and mechanical perspective: e.g. biting creates strong forces that need to be deflected. The insect head thus combines and integrates structures important for a broad spectrum of issues linked to taxonomy, ecology, neurobiology, biomechanics, and developmental biology.

In the scope of the project, we investigate the insect head on a broad taxonomic range (all associated biological partners and the TUD), and additionally focus on the heads of two insect groups, Formicidae (ants; investigated by the partners at the TUD) and Staphylinidae (rove beetles; investigated by the partners at the University of Tübingen). In the two latter, the focus lies on the functional adaptations of the mouthparts and associated musculature to the acquisition of food. Both taxonomic groups are well suited for this research, since they are very diverse (10,000 and 40,000 described species, respectively) and possess a broad feeding spectrum (for example mycophages, phytophages, predators, saprophages, and sporophages). Tomographic data are acquired at the P05 Beamline operated by HZG at the DESY synchrotron radiation source PETRA III^[45-51].

1.3 Collaborative tools

One of the project's main goals is a more economic use of tomographic data. Since tomographic datasets have a large size, appropriate tools are missing, that enable the exchange of analysis results between institutions without copying and storing large chunks of data repeatedly. We were inspired by version control systems that are standard in other domains like software development. On the technical side, the NOVA project focuses on distributing tools to manage data analysis for X-ray tomography. Core technologies are web data catalogs, interactive 3D web visualization for preview, and advanced image processing to automatically extract semantics and be able to identify relative changes during the analysis process. All developments will be integrated in the NOVA data portal as the central instance to coordinate common activities in NOVA. Consequently, on the long term we plan to provide the scientific community with an open access database. Tomographic data acquired in NOVA and its predecessor ASTOR will be stored in a tomography data catalog, which will be created in cooperation with the operators of the tomography instrumentation at PETRA III. In a later stage of the project additional scientific communities ought to be enlisted to participate in this open and cooperative way of data usage.

2. RESULTS

2.1 Data acquisition and tomographic data collection

The NOVA project is building up a catalog of terrestrial arthropods, especially insects (Fig.1) for a broad general overview on head morphology. But the catalog will also contain detailed studies of the two insect groups Formicidae (ants; Fig. 2) and Staphylinidae (rove beetles).

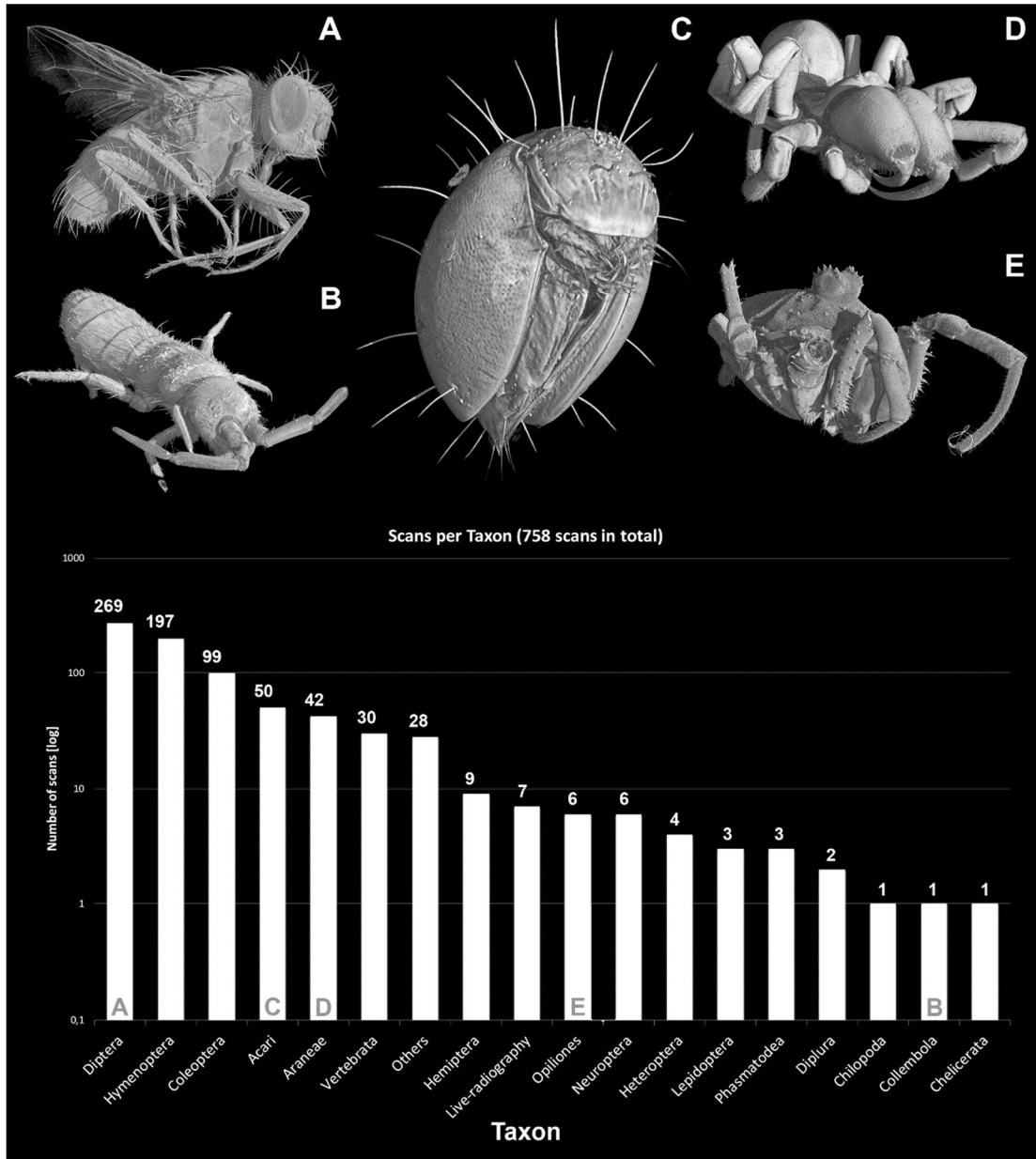


Figure 1. A - E) Examples of recorded datasets (voxel renderings). A) Fly. B) Springtail. C) Oribatid mite (*Euphthiracarus reticulatus* as use case of the preceding ASTOR project). D) Spider. E) Harvestman. Lower section: taxonomic overview of specimens recorded during beam times in April 2014 and September 2015.



Figure 2. Examples of ant species recorded in the beam time in October 2016. A, B - *Pheidole lucretii*. Comparison of soldier (A) and worker morphotype (B) of the same species. C - *Gnamptogenys striatula*. D, E - *Lasius fuliginosus*. (D) is an anterior overview, whereas (E) is a virtual cross section of the ant head showing the central nervous system (CNS; black arrow head), and muscles of the mandibles (white arrow head).

Currently we have around 2100 datasets of various taxonomic groups (Fig. 1), that will be accessible through the NOVA data catalog and will in future be accessible for all interested researchers. This data was recorded in four beam times at the ANKA Synchrotron Radiation Facility (KIT, Karlsruhe) within the preceding ASTOR project, and DESY (Hamburg). Of those datasets, two have been chosen as demonstrator organisms for the collaborative studies: the roach *Schulteria lampridiformis*, and the ant *Lasius fuliginosus* (Fig. 2D,E). Both datasets and the respective biological questions are now being analyzed by the associated biological partners.

2.2 Analysis infrastructure

For the collaborative analysis and visualization dedicated computing hardware was installed. The selected hardware has been chosen to be fully compatible with the already installed components at the high-performance computing infrastructure at DESY and related photon science resources. This allows for a smooth integration and management of the two NOVA compute nodes (each with dual socket Xeon E5-2630 processors with 40 cores in total, 256GB memory and two NVidia Quadro M6000 with 12GB GPU-memory each) by the DESY computing center. Since some of the software tools require hardware accelerated visualization tools, dedicated GPU units had to be integrated into the system.

The nodes were tested and integrated into the DESY HPC infrastructure by the end of February 2017. In contrast to the previous virtual analysis infrastructure at ANKA, we implemented a different remote access solution. Using FASTX-2 we provide user friendly SSH access offering a Linux desktop session to the users via web browser or software client solution.

To log on users need to authenticate via their DESY computer account, which is available to all consortium partners and users performing experiments at DESY facilities. Access to image data is available through the central data repository at the DESY HPC facility. This data repository (gpfs core system) is directly linked to beamline experiments at PETRA III and includes beamtime/project related data management. Tomographic data recorded at PETRA III is directly transferred to this repository providing initial access to the data to all participants registered to the experiment. This data is also available to users from outside the DESY IT network via a web access portal (gamma portal) that allows for additional rights management to grant access to other collaborators. These collaborators currently need to register with the DESY database system to get access to the data via the web portal or FTP. If these collaborators aim to use the central analysis and visualization tools running on the NOVA nodes, however, they would have to register for a DESY computer account as well. In this way, the newly set up NOVA analysis nodes are highly integrated into the already present computing infrastructure and even allow to get access and make use of the HPC infrastructure for complex computationally intense analysis. Currently standard software for image processing is provided to the network partners via licenses available from the DESY IT, DESY computing center, and the Helmholtz-Zentrum-Geesthacht, including commercial packages like AVIZO, Amira, VGStudio MAX, IDL, and MATLAB, as well as opensource and free software tools like ImageJ/Fiji and Paraview. Additional custom software tools developed within the project will be integrated and tested step by step in the upcoming months.

Work has started on a follow-up system for the ASTOR web portal^[9]. A new approach was justified by requirements that could not be satisfied by the previous system:

1. The required portal for NOVA is now focused mainly around the management of datasets and collections of datasets rather than being a control interface for virtual machines, file lists, etc.
2. The previous portal was giving the user a static view of the data, hence collaborative aspects such as editing, sharing, and commenting datasets were not foreseen, but are the main requirements of the new concept.
3. Due to the fact, that the new system has a dataset-centric view, better search capabilities and proper visualization of datasets are required.

The new system is built from scratch, but re-uses relevant components of the previous project. The architecture is split in a REST layer that manages dataset operations, and search indexing via common HTTP GET, POST, PUT, PATCH, and DELETE requests. The REST layer is written in Python using the Flask framework, SQLAlchemy for database abstraction, and Alembic for seamless database migrations. Large-scale search is implemented using Elasticsearch that allows free text search queries, and near-instantaneous search results. Besides these REST requests, Flask is used as well to serve the pages of the web client. For dynamic in-page updates such as editing of metadata, Vue.js components are used that communicate with the REST layer. A Python client can be used to interact with the system from the command line in a similar way as a modern revision control system, which means datasets can be cloned to local disk and changes can be pushed back, while keeping track of the change history. All accesses are authorized using per-user tokens that grant access on behalf of that user. From a collaborative standpoint, a dataset can be reviewed by other parties and shared with them so that they can clone it to their own workspace.

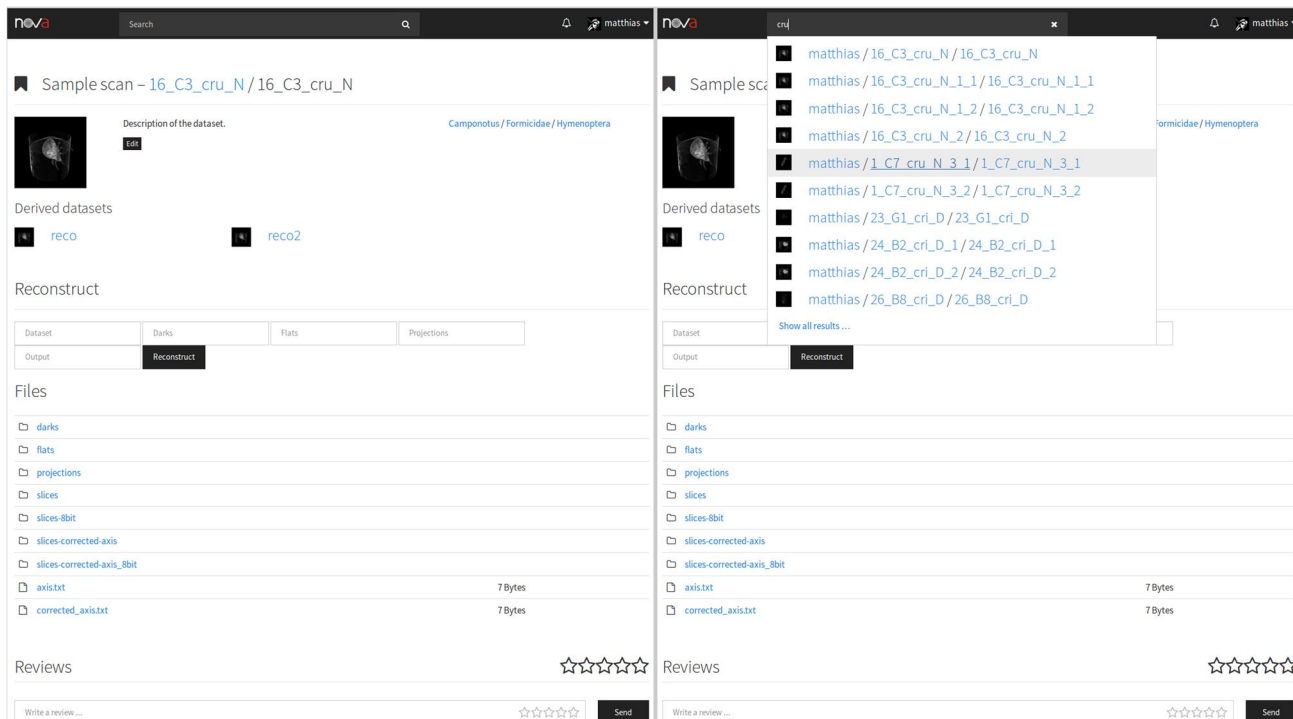


Figure 3. Screenshot of a single dataset in the new NOVA web portal (left), and demonstration of the near-instantaneous display of search results (right).

The current prototype installation provides access to 645 datasets of the previous ASTOR measurement phase with a total size of 30 TB. It allows for free text searching among the datasets, and filtering according to specimen criteria (Fig. 3). The next software iterations will feature integration of web-based visualization services, proper access restrictions, and collaborative tools to merge diverging changes of a dataset. For the latter, methods will be evaluated to deal with user interface aspects as well as strategies for merging.

2.3 Visualization services and computing

Collaborative tools for morphological data require interactive visualization of tomographic volumes (Fig. 4). In the preceding ASTOR project, the development of the WAVE web visualization toolkit has been started^[52]. We continued this approach and enhanced the WAVE framework to obtain the best image quality for the smallest amount of data transferred to the client. The WAVE framework serves as a preview tool for large volume datasets obtained with X-ray microtomography by providing surface and volume renderings.

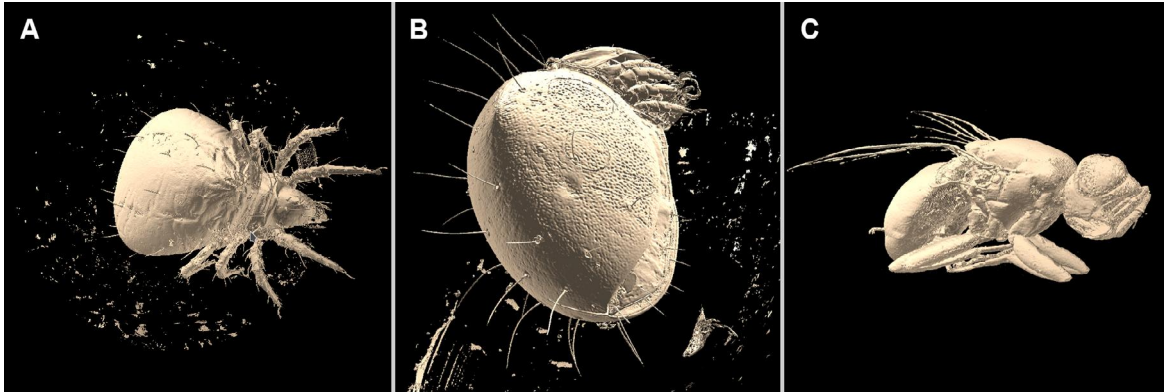


Figure 4. Surface renderings of the raw data using the WAVE web visualization toolkit: the oribatid mite *Archegozetes longisetosus*. (A), the box mite *Euphthiracarus reticulatus* (B) and an unidentified Diptera species (fly; C).

The WAVE framework reads a 3D data structure in the form of a mosaic format image, which is composed of a series of virtual sections. Like Google Maps, the framework supports progressive display by loading a low-resolution slice map first and loading the high-resolution slice map in the background. In our work, we introduce the zoom-on-demand approach allowing the user to select a region of interest in the full 3D volume to inspect details with a higher resolution (Fig. 5).

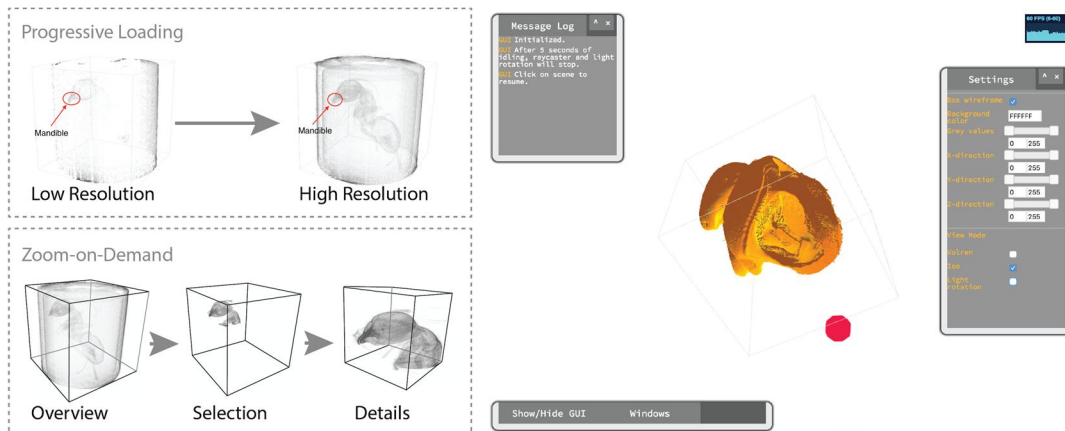


Figure 5. The features supported by the WAVE framework shown on the top left (progressive loading) and bottom left (Zoom-on-Demand). The right side of the image shows the WAVE framework integrated into the Biomedisa application.

To reduce the large data size, the WAVE framework reduces the resolution and transforms the raw data into a series of slice maps. Rather than generating a new slice map on demand for each resolution, we instead precompute a whole hierarchy of multi-resolution slice maps as caches. These multi-resolution slice maps are defined as such that each level N slice map contains $256^3 \times 2^N$ voxels. We concluded that a set of multi-resolution slice maps from level 0 to level 3 covers the typical screen sizes of most client hardware. Further levels can be included depending solely on the advancement of computer hardware. Also, we store precomputed cache data with higher resolution than the slice maps to support the zoom-on-demand feature. As first step, the client-side visualization of the WAVE framework has been integrated into the Biomedisa web application (see next chapter). The full framework will be later deployed in the collaborative analysis portal.

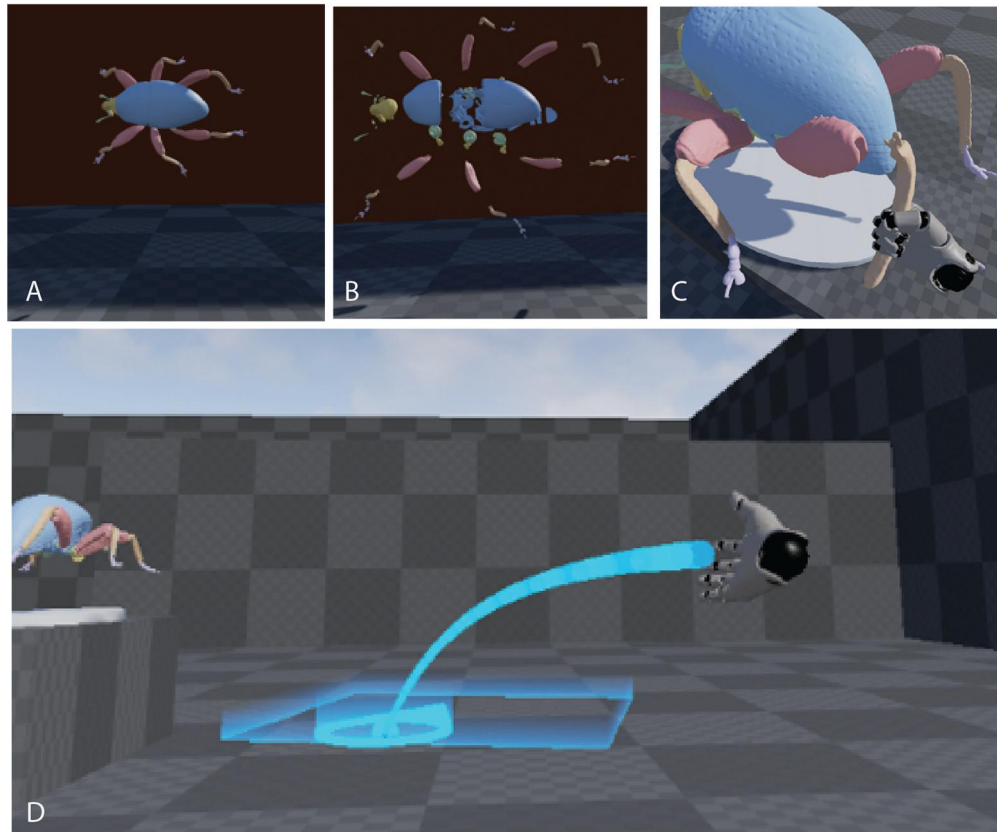


Figure 6. 3D view (A) and 3D explosion drawing (B) of the weevil *Trigonopterus vandekampi*. C - Grabbing the beetle's leg and investigating it in detail. D - The teleporter in the virtual lab.

In addition to the web-based visualization toolkit, we explore the visualization of tomographic datasets in virtual environments made possible by the development of high-quality virtual reality (VR) hardware in recent years (Fig. 6). Current VR devices are now available at moderate prices and feature high-resolution displays, precise head tracking, and lightweight hardware. In our setup, we use the HTC Vive, that is equipped with spatial-tracked hand controllers, which are essential for control and interaction in VR applications. We implemented a prototype of a virtual museum containing two models of a weevil (*Trigonopterus vandekampi*), based on SR- μ CT data from van de Kamp et al.^[42]. Both models can be grabbed using the virtual hand controllers, the first as a single continuous object, the second as separate individual objects. To get a better overview, the second specimen can be visualized as a *3D explosion drawing* of its individual segments (Fig. 6B). All 76 segments can be picked up, moved, and inspected from all sides. We implement the teleportation approach to navigate in the virtual world. A new teleport location is selected with the hand controller and visualized as a blue line or arc (Fig. 6D). In this way, it is also possible to teleport behind smaller obstacles. Further, the outline of the chaperone mesh is displayed and can be turned using the trackpad on the controller. The technology is impressive and extends the possibilities to present high-resolution data obtained by X-ray synchrotron tomography. We plan to include this technology especially in our outreach activities. The prototype has been shown as a live demo at the IVAPP conference^[53] and during the open house event at the Karlsruhe Institute of Technology.

2.4 The image segmentation app Biomedisa

The segmentation of tomographic images remains one of the most challenging tasks in computer vision. With the analysis of the data and their ever-increasing volume, further problems are arising. Available tools suffer from high computational needs, long loading time, or poor interactive response. Manual segmentation is still considered the “gold standard” in many cases. The web application Biomedisa (<https://biomedisa.de>; Fig. 7) was developed as a semi-automatic tool for the improvement and acceleration of tedious manual segmentation. It is based on a parameter-free and highly scalable diffusion algorithm which was developed during the previous project ASTOR.

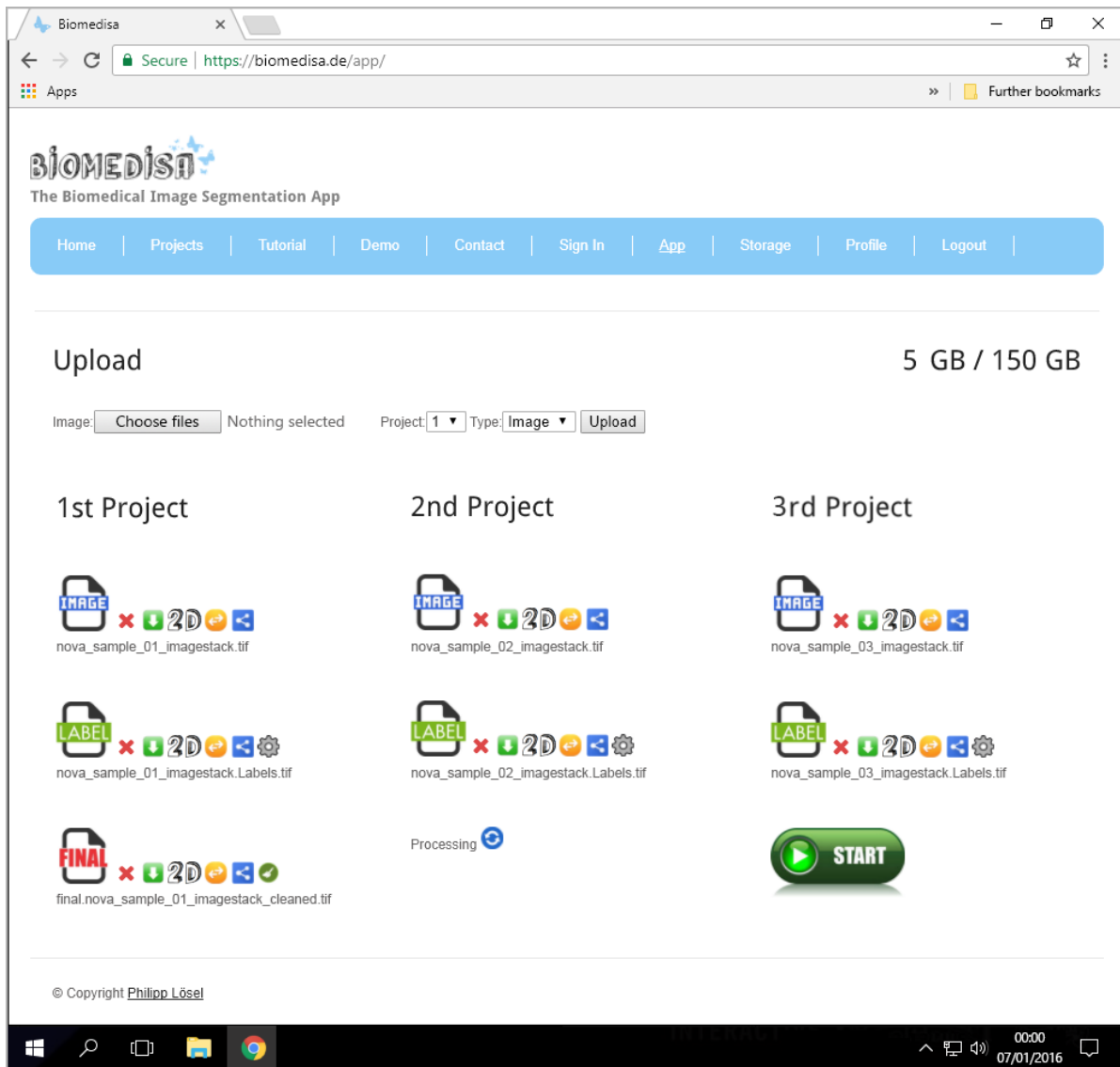


Figure 7. The Biomedisa web interface. Several functions are accessible by using buttons. Clicking the file button will activate the 3D visualization tool, the 2D symbol will open a slice viewer. In addition, it is possible to remove (red cross button), download (green button), share (blue button), and move the data from and to the storage (yellow button). It is also possible to delete outliers or undo the deletion (dark green button), and to enable labeling in all planes (settings symbol).

The segmentation uses weighted random walks^[10], which begin in a few manually labeled reference slices and extrapolate the information contained in these slices to the remaining volume. Over time, the voxels are hit by numerous random walks. Based on the respective number of hits, it is possible to approximately determine the probability that a voxel can be assigned to one particular segment. Segmentation can thus be performed by assigning each voxel to the region from which most of the hits originate. The weights are defined without the use of hyperparameters. Thus, it is not necessary to search for proper values, which eliminates the need for an elaborate and tedious configuration. Due to the random walks' independence, their calculation is highly scalable, which allows for the use of massively parallel computer architectures (e.g., graphics processors) and thus for an evaluation of large image data in a short time. In addition, a quick and easy correction of the segmentation results is possible by adding additional reference slices.

This semi-automatic approach has already proven to be extremely effective within the ASTOR project. Since the input data can be created with any common segmentation software, such as Amira or Fiji/ImageJ, the users can continue to work in their familiar environment. However, the segmentation process itself is significantly accelerated, and the quality of the resulting segmentation is improved. The web application Biomedisa was established to provide users with the required computing power. Users can upload their tomograms and labeled slices, run the segmentation process, and visualize their image data and results using a 2D slice viewer, and the 3D rendering software developed in the project. The procedure of the segmentation is documented in an online tutorial and can be tested in a demo session. In addition to the segmentation application, the users can store their data and share it with other users. Once a new user accepts the data, it is physically copied into the second user's memory. It is also possible to share data with unregistered persons by creating a password-protected download link.

2.5 Evaluation of the NOVA semi-automatic segmentation

To evaluate the semi-automatic segmentation, we took every n^{th} slice (for $n = 5, 10, 15, 20, 30, 40, 50, 100$) from a complete manually segmented dataset of the box mite *Euphthiracarus reticulatus* BERLESE (Fig. 8A,C) and used these as basis for the semi-automatic segmentation. We then calculated 3D models based on the resulting segmented image stacks with Amira 5.6.0 (Fig. 8B-H). A rough visual comparison shows little to no differences (Fig. 8B-F). The larger exoskeletal elements look nearly identical. Only a detailed review shows differences in minute structures (Fig. 8G,H). A comparison of surface properties (number of triangles, surface area and volume) shows increasing differences between the 3D models depending on the number of skipped slices (table 1; Fig. 8). The maximum difference between surface area (1.95%) and number of triangles (3.85) is compared to the difference in volume (10.82%) relatively low. The volume difference, however, stays nearly the same (around 6.5%) through every 5th to every 50th slice, changing only by 1.09%. The same can be said about the number of triangles and the surface area, although the numbers of differences are even lower. To further characterize the differences, we used the *shortest edge distance module* in Amira (table 2, Fig. 9). Overall, the mean error remains smaller than the effective voxel size of 1.22 μm . A threshold value of 1.22 μm reveals that the congruence of the 3D models lies between 85.6 and 91.7%, so even taking only every 100th slice results in a difference of just 14.4%, but results in a massive amount of saved time. The maximum error in this scenario, however, is 184.2 μm . A visual comparison reveals, that the long hairs are the source of this large maximum error (Fig. 9). Since selection of pre-segmented slices was done based on a fixed distance, a content-based selection of pre-segmented slices should increase the quality of semi-automatic segmentation by a large factor. Manual postprocessing of the semi-automatically segmented slices can of course be done to optimize the results. Additionally, the amount of time the total process of semi-automatic segmentation takes is only a fraction of the time needed for a complete manual segmentation. This in turn not only enables work on single specimens to be done much faster, but also facilitates studies that need several to dozens of specimens, for example ecological studies^[54], that would not have been feasible relying on manual segmentation alone. In combination with the NOVA data catalog, that already contains hundreds of datasets and will constantly be expanded, quantitative biology on high-res SR μ CT data can be realized in a to date unknown dimension.

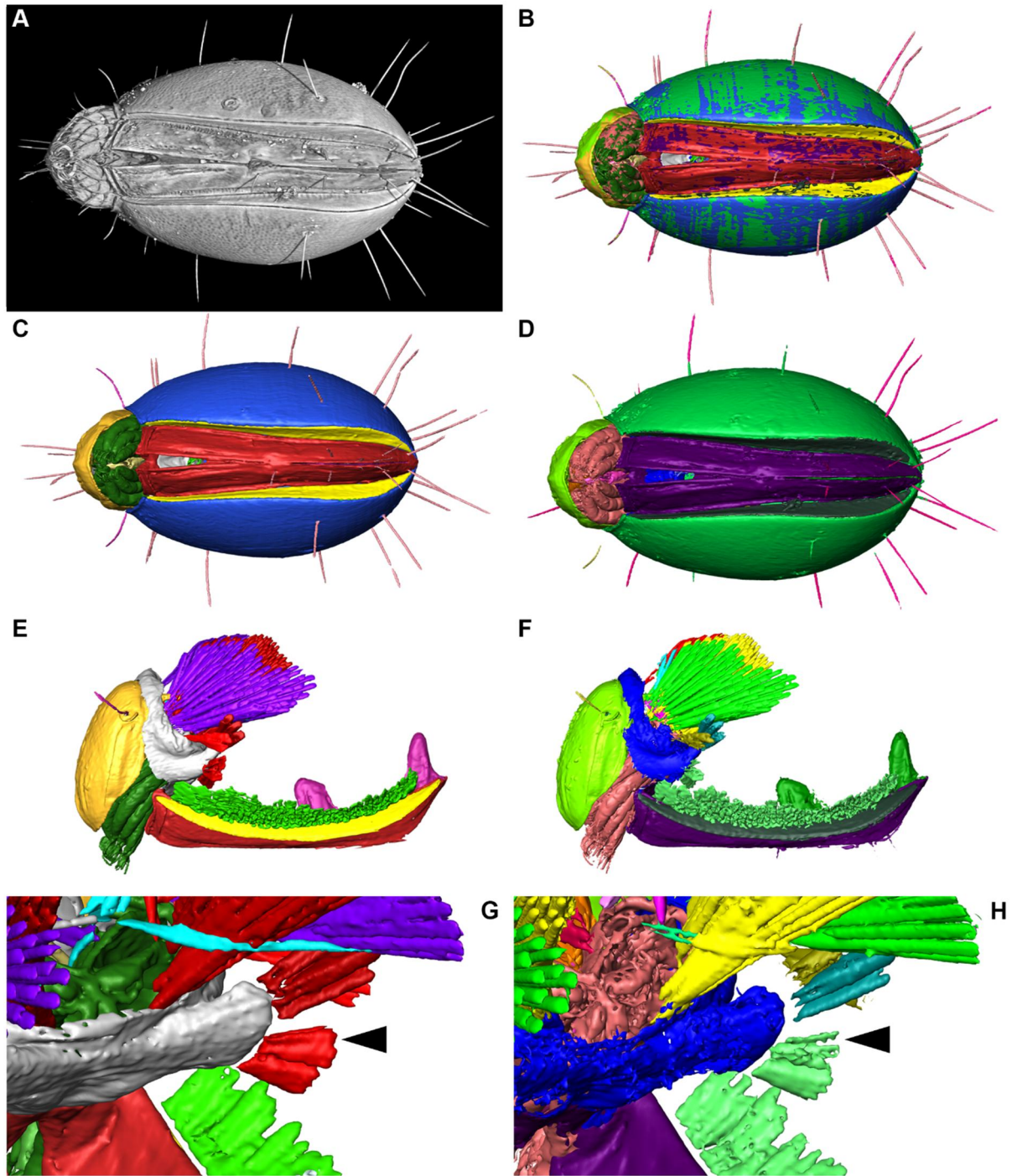


Figure 8. A) Voxel rendering of a dataset of *Euphthiracarus reticulatus* recorded at ANKA, Karlsruhe, Germany. B) Overlay of 3D models of manual and semi-automatic segmentation (cf. C and D). C, E, G) Result of the manual segmentation. D, F, H) Result of the semi-automatic segmentation based on pre-segmentation of every 50th slice. Arrowheads in G and H point to a region with clear differences.

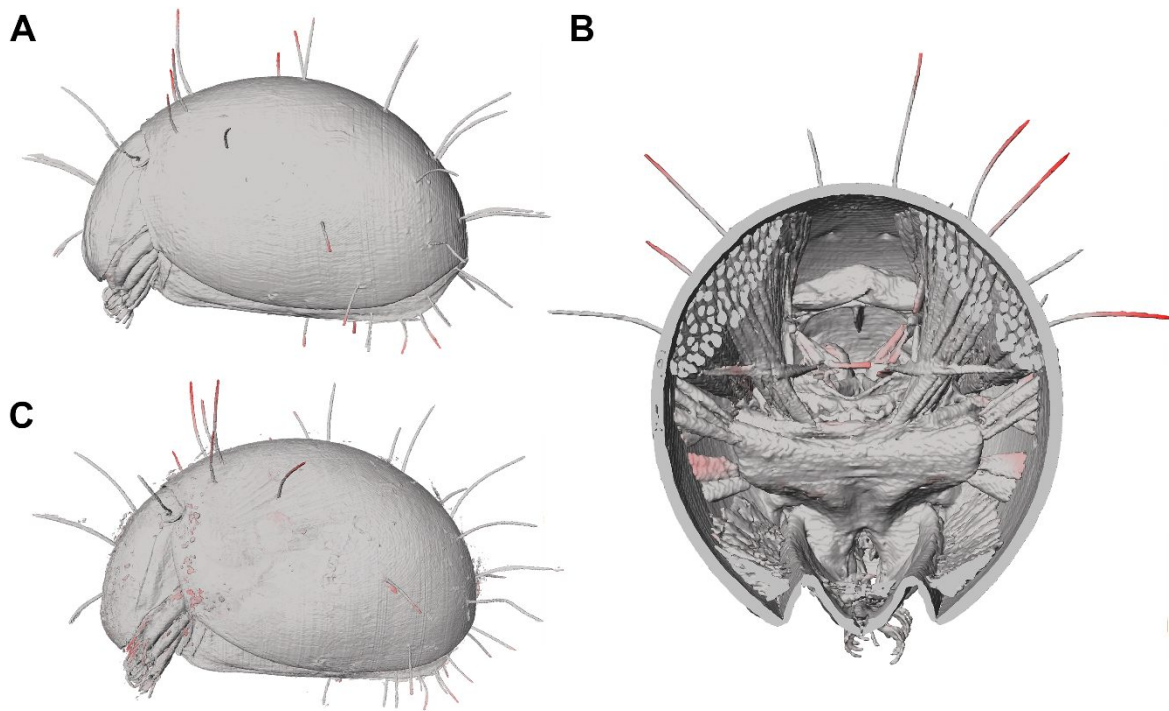


Figure 9. Comparison of manual and semi-automatic segmentation of the ptychoid box mite *Euphthiracarus reticulatus*. A) Lateral view of the manually segmented 3D model and projected error intensity (shades of red; based on every 20th slice pre-segmented). B) Virtual cross section of (A). C) Lateral view of the manually segmented 3D model and projected error intensity (shades of red; based on every 100th slice pre-segmented).

Table 1. Relative differences of three measured variables of semi-automatic vs. manual segmentation.

Slice distance	5	10	15	20	30	40	50	100
Number of triangles [%]	0,38	0,6	0,47	0,96	0,86	1,34	2	3,85
Surface area [%]	0,15	0,33	0,17	0,58	0,43	0,78	1,27	1,95
Volume [%]	-6,4	-6,41	-6,44	-6,54	-6,68	-6,97	-7,49	-10,82

Table 2. Congruence, mean error, standard deviation and maximum error of semi-automatic segmentation related to manual work load and distance of segmented slices (all: manual segmentation of all slices).

Slice distance	All	5	10	15	20	30	40	50	100
Manual work [h]	50	10	5	3,33	2,5	1,67	1,25	1	0,5
Computing* [min]	0	120	63	43	32	22	16	13	7
Congruence [%]	100	91,7	90,1	89,1	89,0	88,5	88,8	87,4	85,6
Mean error [μm]	0	0,31	0,32	0,37	0,36	0,48	0,53	0,59	1,12
Standard deviation [μm]	-	1,10	1,30	2,01	1,87	3,10	3,47	3,21	5,46
Maximum error [μm]	1,22	38,5	75,6	127,6	125,7	136,5	156,4	137,6	184,2

* Calculation was done on six GPUs (4xK40, 1xK20, 1xGridK2).

3. CONCLUSION

The project addresses core challenges of data-intensive sciences. In many fields, the amount of data has increased in such a way that it renders currently available tools unsuitable. The scientists in NOVA are convinced that not only new tools are required but also a collaborative attitude is required to succeed in the new age of data-intensive sciences. In this sense, the results from NOVA on one community will serve as an example that could also be transferred to other communities. While each community needs its specific methods, many of the collaborative tools can be considered as generic, e.g. general segmentation algorithms, or scalable visualization system dedicated to large datasets. The comprehensive data catalog with high-resolution tomograms of various arthropods is not only valuable to scientists, but also for the society in general.

4. ACKNOWLEDGEMENTS

We thank the German Federal Ministry of Education and Research (www.bmbf.de) for funding of ASTOR and NOVA (grants 05K13VK5 and 05K16VKB, respectively). DESY (proposal numbers I-20160035, I-20160657, and I-20170141) and ANKA are acknowledged for providing beamtime.

REFERENCES

- [1] Betz, O., Wegst, W., Weide, M., Heethoff, M., Helfen, L., Lee, W.-K. and Cloetens, P., "Imaging applications of synchrotron X-ray phase-contrast microtomography in biological morphology and biomaterials science. I. General aspects of the technique and its advantages in the analysis of millimetre-sized arthropod structure," *J. Microsc.* 227(1), 51-71 (2007).
- [2] Friedrich, F. and Beutel, R.G., "Micro-computer tomography and a renaissance of insect morphology," *Proc. SPIE* 7078, 70781U-1-70781U-6 (2008).
- [3] Schmelzle, S., Helfen, L., Norton, R.A. and Heethoff, M., "The ptychoid defensive mechanism in Euphthiracaroida (Acari: Oribatida): A comparison of muscular elements with functional considerations," *Arthropod Structure and Development* 38 (6), 461-472 (2009). doi:10.1016/j.asd.2009.07.001
- [4] Schmelzle, S., Norton, R.A. and Heethoff, M., "Mechanics of the ptychoid defense mechanism in Ptyctima (Acari, Oribatida): one problem, two solutions," *Zoologischer Anzeiger - A Journal of Comparative Zoology* 254, 27-40 (2015). doi:10.1016/j.jcz.2014.09.002
- [5] Alberti, G., Heethoff, M., Norton, R.A., Schmelzle, S., Seniczak, A. and Seniczak, S., "Fine Structure of the Gnathosoma of *Archezogozetes longisetus* Aoki (Acari: Oribatida, Trhypochthoniidae)," *Journal of Morphology* 272,1025-1079 (2011).
- [6] van de Kamp, T., Ershov, A., dos Santos Rolo, T., Riedel, A. and Baumbach, T., "Insect imaging at the ANKA Synchrotron Radiation Facility," *Entomologie heute* 25, 147-160 (2013).
- [7] Hartmann, K., Laumann, M., Bergmann, P., Heethoff, M. and Schmelzle, S., "Development of the Synganglion and Morphology of the Adult Nervous System in the Mite *Archezogozetes longisetosus* Aoki (Chelicerata, Actinotrichida, Oribatida)," *Journal of Morphology* 277, 537-548 (2016).
- [8] Beutel, R.G., Friedrich, F., Hörschemeyer, T., Pohl, H., Hünefeld, F., Beckmann, F., Meier, R., Misof, B., Whiting, M.F. and Vilhelmsen, L., "Morphological and molecular evidence converging upon a robust phylogeny of the megadiverse Holometabola," *Cladistics* 26, 1-15 (2011).
- [9] Mexner, W., Bonn, M., Kopmann, A., Mauch, V., Ressmann, D., Chilingaryan, S., Tan Jerome, N., van de Kamp, T., Heuveline, V., Lösel, P., Schmelzle, S. and Heethoff, M., "OpenGL® API based analysis of large datasets in a cloud environment; in: Das, P.K. and Deka, G.C. (eds.), *Design and use of virtualization technology in cloud computing*," IGI Global, Hershey, in press (2017).
- [10] Lösel, P. and Heuveline, V., "Enhancing a diffusion algorithm for 4D image segmentation using local information," *Proc. SPIE* 9784, (2016). doi:10.1117/12.2216202
- [11] Lösel, P. and Heuveline, V., "A GPU Based Diffusion Method for Whole-Heart and Great Vessel Segmentation; in: Zuluaga, M., Bhatia, K., Kainz, B., Moghari, M. and Pace D. (eds), *Reconstruction, Segmentation, and Analysis of Medical Images. Lecture Notes in Computer Science*, vol 10129," Springer, Cham, (2017). doi:10.1007/978-3-319-52280-7_12
- [12] Blanke, A., Wipfler, B., Letsch, H., Beckmann, F., Beutel, R. and Misof, B., "Revival of Palaeoptera—head characters support a monophyletic origin of Odonata and Ephemeroptera (Insecta)," *Cladistics* 28(6), 560-581 (2012).
- [13] Blanke, A., Beckmann, F. and Misof, B., "The head anatomy of *Epiophlebia superstes* (Odonata: Epiophlebiidae)," *Org. Divers. Evol.* 13, 55–66 (2013).
- [14] Blanke, A., Greve, C., Mokso, R., Beckmann, F. and Misof, B., "An updated phylogeny of Anisoptera including formal convergence analysis of morphological characters," *Syst. Entomol.* 38, 474-490 (2013).

- [15] Blanke, A., Koch, M., Wipfler, B., Wilde, F. and Misof, B., "Head morphology of *Tricholepidion gertschi* indicates monophyletic Zygentoma," *Front. Zool.* 11, 16 (2014).
- [16] Blanke, A. and Wesener, T., "Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda)," *Arthropod Struct. Dev.* 43, 63-75 (2014).
- [17] von Reumont, B.M., Blanke, A., Richter, S., Alvarez, F., Bleidorn, C. and Jenner, R.A., "The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin?," *Mol. Biol. Evol.* 31, 48-58 (2014).
- [18] Blanke, A., Machida, R., Szucsich, N.U., Wilde, F. and Misof, B., "Mandibles with two joints evolved much earlier in the history of insects: dicondily is a synapomorphy of bristletails, silverfish and winged insects," *Syst. Entomol.* 40, 357-364 (2015).
- [19] Blanke, A., Rühr, P.T., Mokso, R., Villanueva, P., Wilde, F., Stampanoni, M., Uesugi, K., Machida, R. and Misof, B., "Structural mouthpart interaction evolved already in the earliest lineages of insects," *Proc. R. Soc. B.* 282, 20151033 (2015).
- [20] Blanke, A., Greve, C., Wipfler, B., Beutel, R.G., Holland, B.R. and Misof, B., "The Identification of concerted convergence in insect heads corroborates Palaeoptera," *Syst. Biol.* 62, 250-263 (2013).
- [21] Weide, D. and Betz, O., "Head morphology of selected Staphylinoidea (Coleoptera: Staphyliniformia) with an evaluation of possible groundplan features in Staphylinidae," *J. Morph.* 270, 1503-1523 (2009).
- [22] Weide, D., Thayer, M.K., Newton, A.F. and Betz, O., "Comparative morphology of the head of selected sporophagous and non-sporophagous aleocharinae (Coleoptera: Staphylinidae): Musculature and hypopharynx-prementum complex," *J. Morph.* 271, 910-931 (2010).
- [23] Weide, D., Thayer, M.K. and Betz, O., "Comparative morphology of the tentorium and hypopharyngeal-premental sclerites in sporophagous and non-sporophagous adult Aleocharinae (Coleoptera: Staphylinidae)," *Acta Zool.* 95, 84-110 (2014).
- [24] Schmitt, C., Rack, A. and Betz, O., "Analyses of the mouthpart kinematics in *Periplaneta americana* (Blattodea, Blattidae) using synchrotron-based X-ray cineradiography," *J. Exp. Biol.* 217, 3095-3107 (2014).
- [25] Wipfler, B., Schneeberg, K., Löffler, A., Hünefeld, F., Meier, R. and Beutel, R.G., "The skeletomuscular system of the larva of *Drosophila melanogaster* (Drosophilidae, Diptera) - A contribution to the morphology of a model organism," *Arthropod Structure and Development* 42(1), 47-68 (2013). doi:10.1016/j.asd.2012.09.005
- [26] Kenning, M., Müller, C., Wirkner, C.S. and Harzsch, S., "The Malacostraca (Crustacea) from a neurophylogenetic perspective: New insights from brain architecture in *Nebalia herbstii* Leach, 1814 (Leptostraca, Phyllocarida)," *Zoologischer Anzeiger* 252, 319-336 (2013). doi:10.1016/j.jcz.2012.09.003
- [27] Meth, R., Wittfoth, C. and Harzsch, S., "Brain architecture of the Pacific White Shrimp *Penaeus vannamei* Boone, 1931 (Malacostraca, Dendrobranchiata): correspondence of brain structure and sensory input?," *Cell and Tissue Research*, 1-17 (2017). doi:10.1007/s00441-017-2607-y
- [28] Sombke, A., Lipke, E., Michalik, P., Uhl, G. and Harzsch, S., "Potential and limitations of X-ray micro-computed tomography in arthropod neuroanatomy – a methodological and comparative survey," *The Journal of Comparative Neurology* 523(8), 1281-1295 (2015).
- [29] Hörnig, M.K., Sombke, A., Haug, C., Harzsch, S. and Haug, J.T., "What nymphal morphology can tell us about parental investment – a group of cockroach hatchlings in Baltic Amber documented by a multi-method approach," *Palaeontologia Electronica* 19.1.5A, 1-20 (2016).

- [30] Onelli, O.D., van de Kamp, T., Skepper, J.N., Powell, J., dos Santos Rolo, T., Baumbach, T. and Vignolini, S., "Development of structural colour in leaf beetles," *Sci. Rep.* 7, 1373 (2017).
- [31] Wulff, N.C., van de Kamp, T., dos Santos Rolo, T., Baumbach, T. and Lehmann, G.U.C., "Copulatory courtship by internal genitalia in bushcrickets," *Sci. Rep.* 7, 42345 (2017).
- [32] dos Santos Rolo, T., Ershov, A., van de Kamp, T. and Baumbach, T., "In vivo X-ray cine-tomography for tracking morphological dynamics," *PNAS* 111(11), 3921-3926 (2014).
- [33] van de Kamp, T., dos Santos Rolo, T., Baumbach, T. and Greven, H., "X-ray radiography of a spraying stick insect (Phasmatodea)," *Entomologie heute* 27, 37-44 (2015).
- [34] Riedel, A., dos Santos Rolo, T., Cecilia, A. and van de Kamp, T., "Sayrevilleinae Legalov, a new subfamily of fossil weevils (Coleoptera, Curculionoidea, Attelabidae) and the use of synchrotron microtomography to examine inclusions in amber," *Zool. J. Linn. Soc.* 165(4), 773-794 (2012).
- [35] Vršanský, P., van de Kamp, T., Azar, D., Prokin, A., Vidlička, L. and Vagovič, P., "Cockroaches likely cleaned up after dinosaurs," *PLoS ONE* 8(12), e80560, 1-11 (2013).
- [36] Butcher, B.A., Zaldivar-Riverón, A., van de Kamp, T., dos Santos Rolo, T., Baumbach, T. and Quicke, D.L.J., "Extension of historical range of Betylobracninae (Hymenoptera: Braconidae) into Palaearctic Region based on a Baltic amber fossil, and description of a new species of *Mesocentrus Szépligeti* from Papua New Guinea," *Zootaxa* 3860(5), 449-463 (2014).
- [37] van de Kamp, T., dos Santos Rolo, T., Baumbach, T. and Krogmann, L., "Scanning the past – synchrotron X-ray microtomography of fossil wasps in amber," *Entomologie heute* 26, 151-160 (2014).
- [38] Schwermann, A.H., dos Santos Rolo, T., Caterino, M.S., Bechly, G., Schmied, H., Baumbach, T. and van de Kamp, T., "Preservation of three-dimensional anatomy in phosphatized fossil arthropods enriches evolutionary inference," *eLife* 5, e12129, 1-17 (2016).
- [39] Schwermann, A.H., Wuttke, M., dos Santos Rolo, T., Caterino, M.S., Bechly, G., Schmied, H., Baumbach, T. and van de Kamp, T., "The fossil insects of the Quercy region: a historical review," *Entomologie heute* 28, 127-142 (2016).
- [40] Zuber, M., Laaß, M., Hamann, E., Kretschmer, S., Hauschke, N., van de Kamp, T., Baumbach, T. and Koenig, T., "Augmented laminography, a correlative 3D imaging method for revealing the inner structure of compressed fossils," *Sci.Rep.* 7, 41413 (2017).
- [41] van de Kamp, T., Vagovič, P., Baumbach, T. and Riedel, A., "A biological screw in a beetle's leg," *Science* 333(6038), 52 (2011).
- [42] van de Kamp, T., dos Santos Rolo, T., Vagovič, P., Baumbach, T. and Riedel, A., "Three-dimensional reconstructions come to life – interactive 3D PDF animations in functional morphology," *PLoS ONE* 9(7): e102355, 1-7 (2014).
- [43] van de Kamp, T., Cecilia, A., dos Santos Rolo, T., Vagovič, P., Baumbach, T. and Riedel, A., "Comparative thorax morphology of death-feigning flightless cryptorhynchine weevils (Coleoptera: Curculionidae) based on 3D reconstructions," *Arthropod Struct. Dev.* 44(6), 509-523 (2015).
- [44] Sandeman, D., Kenning, M. and Harzsch, S., "Adaptive trends in malacostracan brain form and function related to behavior; in: Derby, C., and Thiel, M. (Eds.), *Crustacean Nervous System and their Control of Behaviour. The Natural History of the Crustacea. Vol. 3.*," Oxford University Press, (2014).
- [45] Beckmann, F., Dose, T., Herzen, J., Uteke, S., Lippmann, T., Schell, N., Schreyer, A., Stock, S.R. and Haibel, A., "The GKSS Beamlines at PETRA III and DORIS III," *Developments in X-ray tomography VI.* San Diego, USA, (2008).

- [46] Haibel, A., Beckmann, F., Dose, T., Herzen, J., Ogurreck, M., Müller, M. and Schreyer, A., "Latest developments in microtomography and nanotomography at PETRA III," Powder diffraction 25, 161-164 (2010).
- [47] Haibel, A., Ogurreck, M., Last, A., Mohr, J., Beckmann, F., Dose, T., Wilde, F., Herzen, J., Müller, M., Schreyer, A., Nazmov, V. and Simon, M., "Micro- and nano-tomography at the GKSS Imaging Beamline at PETRA III," SPIE Optical Engineering + Applications. San Diego, California, (2010).
- [48] Hipp, A., Beckmann, F., Lytaev, P., Greving, I., Lottermoser, L., Dose, T., Kirchhof, R., Burmester, H., Schreyer, A. and Herzen, J., "Grating-Based X-Ray Phase-Contrast Imaging at PETRA III," SPIE Optical Engineering + Applications. San Diego, California, (2014).
- [49] Greving, I., Burmester, H., Müller, M., Beckmann, F., Wilde, F., Ogurreck, M., Herzen, J., Hammel, J.U., Hipp, A., Friedrich, F., Lottermoser, L. and Dose, T. "P05 imaging beamline at PETRA III: first results," Developments in X-Ray Tomography IX 9212, 92120O (2014).
- [50] Hipp, A., Herzen, J., Hammel, J., Lytaev, P., Schreyer, A. and Beckmann, F., "Single-grating interferometer for high-resolution phase-contrast imaging at synchrotron radiation sources," SPIE Optical Engineering + Applications 9967, 996718-996718-996717 (2016).
- [51] Wilde, F., Ogurreck, M., Burmester, H., Müller, M., Schreyer, A., Greving, I., Hammel, J.U., Beckmann, F., Hipp, A., Lottermoser, L., Khokhriakov, I., Lytaev, P. and Dose, T., "Micro-CT at the Imaging Beamline P05 at PETRA III," Proceedings of the 12th International Conference on Synchrotron Radiation, (2015).
- [52] Tan Jerome, N., Chilingaryan, S., Shkarin, A., Kopmann, A., Zapf, M., Lizin, A. and Bergmann, T., "WAVE: a 3D online previewing framework for big data archives," Proc. VISIGRAPP 2017, 152-163 (2017).
- [53] Bergmann, T., Balzer, M. Hopp, T., Kopmann, A., van de Kamp, T., Tan Jerome, N. and Zapf, M., "Inspiration from VR Gaming Technology: Deep Immersion and Realistic Interaction for Scientific Visualization," International Conference on Information Visualization Theory and Applications - IVAPP 2017, Porto, Portugal, (2017).
- [54] Kühnel, S., Brückner, A., Schmelzle, S., Heethoff, M. and Blüthgen, N., "Surface area-volume ratios in insects," Insect Science, 1-13 (2016). doi:10.1111/1744-7917.12362