Java Scientific Containers - an open source generic large data library for visualization applications

Piotr Wendykier, Bartosz A. Borucki, Krzysztof S. Nowiński
Visual Analysis Laboratory
Interdisciplinary Centre for Mathematical and Computational Modelling
University of Warsaw, Poland

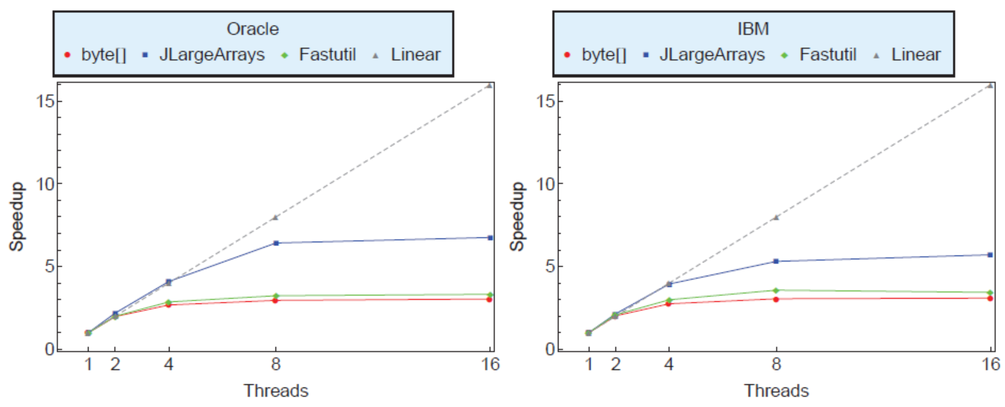
https://gitlab.com/ICM-VisLab/JSciC

Java Scientific Containers (JSciC)

- Open source pure Java library for generic representation of scientific data.
- Significantly focused on the data types for visualization systems.
- Implements the concept of a field that represents a dataset with three components: structure, geometry and multivariate values (scalar or vector).
- Provides flexible time support by independent time steps.
- Includes interpolation, slicing, arithmetical and statistical operations.
- Supports multi-threaded operations on data components.
- Implements arithmetic with physical units.
- Allows to store data components with up to 2^{63} elements.
- Includes numeric I/O data format.

Background

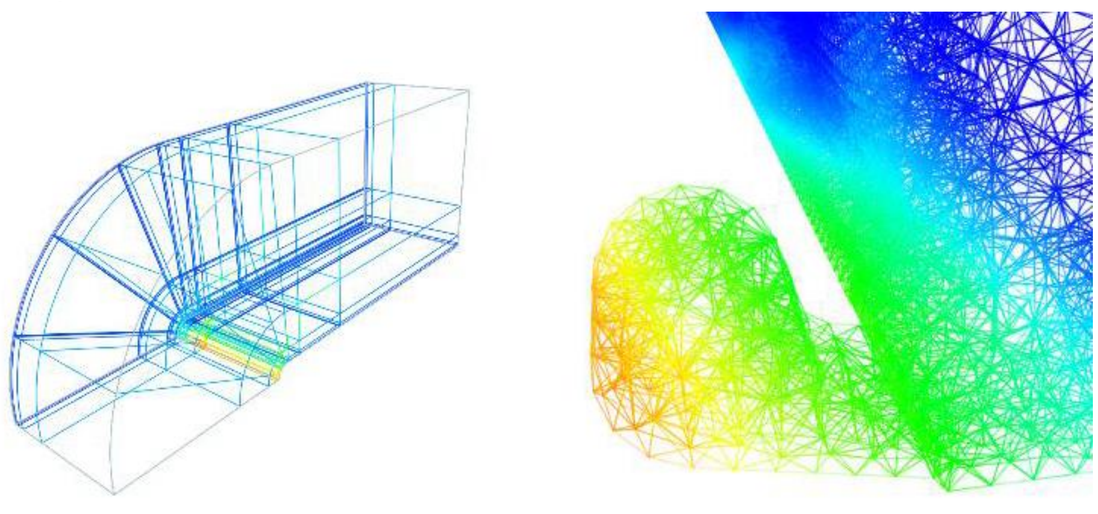
- VTK is the current standard written in C++, but there is no pure Java alternative.
- JSciC stores numerical values in 1-dimensional arrays as Java lacks true multidimensional arrays.
- JSciC uses JLargeArrays library to overcome Java limitation of 231 elements in a 1-dimensional array.
- Benchmarks results show that JLargeArrays generally outperforms both Fastutil library and native Java arrays.



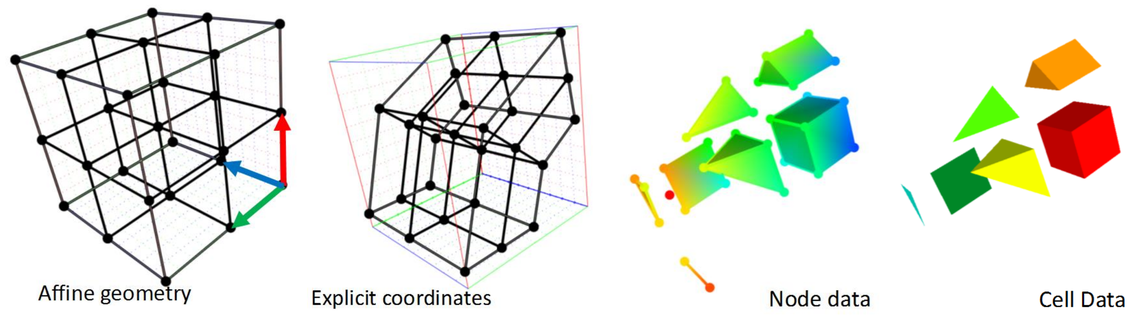
Performance of multi-threaded random read/write access on the array of bytes (array size: 2^{31}-4).

Generic Field Concept

- The geometry part of the field object describes spatial location of nodes, the structure part defines relationships between nodes, forming a data grid, and the values represent measurements on each given node.
- Two subtypes of the field: regular and irregular.
- Nodes in a regular field are organized as 1, 2, or 3-dimensional regular grid.
- The geometry of a regular field can be determined either by its origin and unit cell vectors or by explicitly given coordinate array.
- An irregular field is determined by a set of nodes and one or more cell sets. JSciC supports point, segment, triangle, quadrangle, tetrahedron, pyramid, prism and hexahedron as basic cell types. In addition to node data, cell data can be defined with different data on each cell set.
- Multivariate values, called components, can be defined over nodes or cells.
- Multiple components can be defined for one field.
- Supported types of components: logic, byte, short, int, long, float, double, complex, String, Object.
- Physical units and linear mapping to physical range are supported.
- Each data component can be scalar or vector.
- Field nodes can be masked as valid or invalid.
- Time dependency is defined independently for geometry, mask and data components.

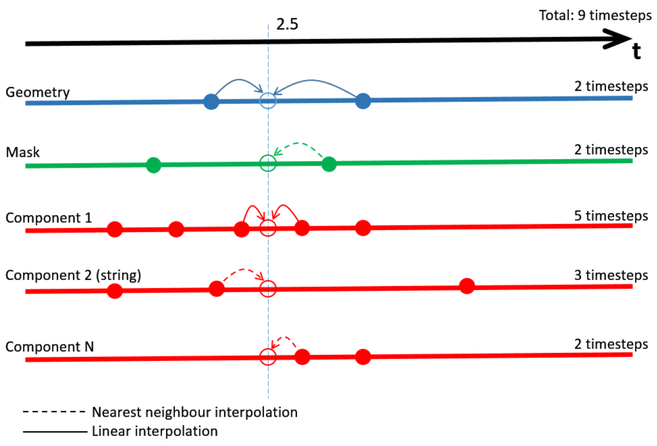


Examples of curvilinear regular field (left) and irregular field on tetrahedral cells (right) as presented in VisNow.

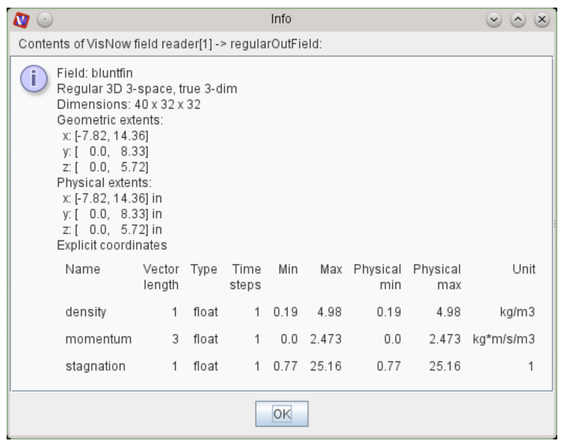


Functionality

- JSciC library provides multiple embedded structure, geometry and data processing tools as field's functionality or generic utilities:
  - interpolation,
  - triangulation,
  - slicing,
  - fingerprints,
  - I/O operations (header + data),
  - cropping,
  - downsizing.
- Data components functionality include:
  - type conversion,
  - arithmetic with physical units,
  - statistics,
  - linear algebra.
- Built-in parallelization on multicore architectures.



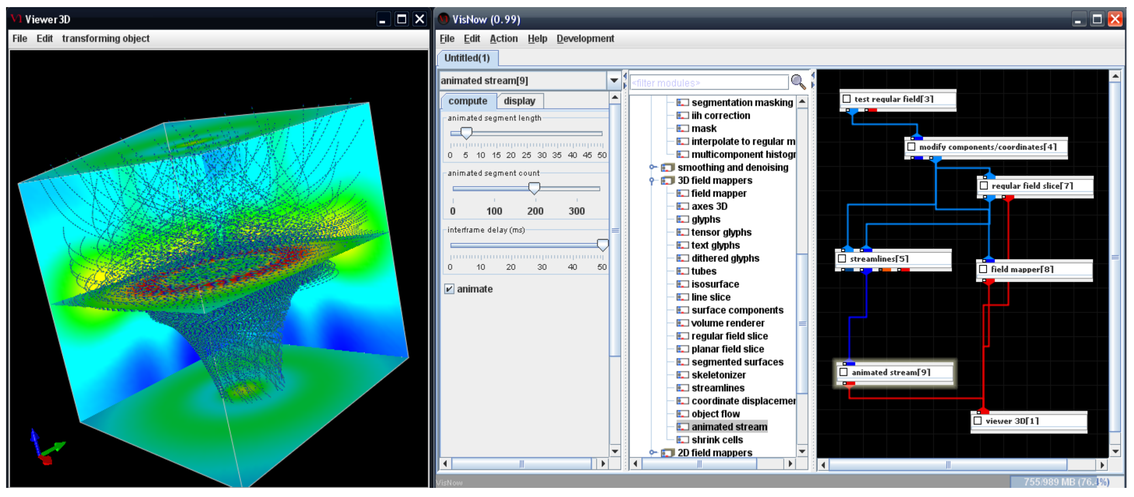
Example time dependency and interpolation schemes in JSciC



JSciC field description in VisNow.

Applications and Visualization Context

- JSciC originated from VisNow – a modular, pluggable system for data processing and visual analysis based on data-flow driven paradigm (http://visnow.icm.edu.pl).
- Currently JSciC provides internal data structures for VisNow.
- Both regular and irregular fields, serving the structure and geometry, provide the visualization system with primitives and basic structures for visualization purposes.
- Multivariate field data - multiple components – provide a choice of data for colormapping or other visual layer representation of a single field.
- Data arithmetics at component level provided by JSciC is utilized as component calculator with expression parser within VisNow module, serving as data generator or calculator for derived components.
- VisNow Field data format (VNF) being a native file format for data fields is a direct usage of JSciC provided I/O.
- JSciC can be used in applications that require generic data types defined in a real vector space such as physical simulations and engineering tools.
- Support for very large datasets makes JSciC useful in fields ranging from microscopic and medical, up to cosmological sciences.



Example view of VisNow - each module in the processing network is responsible for separate visualization task: data access, pre-processing (filtering), mapping, post-processing, rendering and presentation. Data flow (module connections) is based on JSciC concept

Contact: Laboratory of Visual Analysis, Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw
Piotr Wendykier (p.wendykier@icm.edu.pl), Bartosz A. Borucki (b.borucki@icm.edu.pl), Krzysztof S. Nowiński (k.nowinski@icm.edu.pl)
ICM UW, Pawinskiego 5a, 02-106 Warsaw, Poland, +48(22)8749100, www.icm.edu.pl