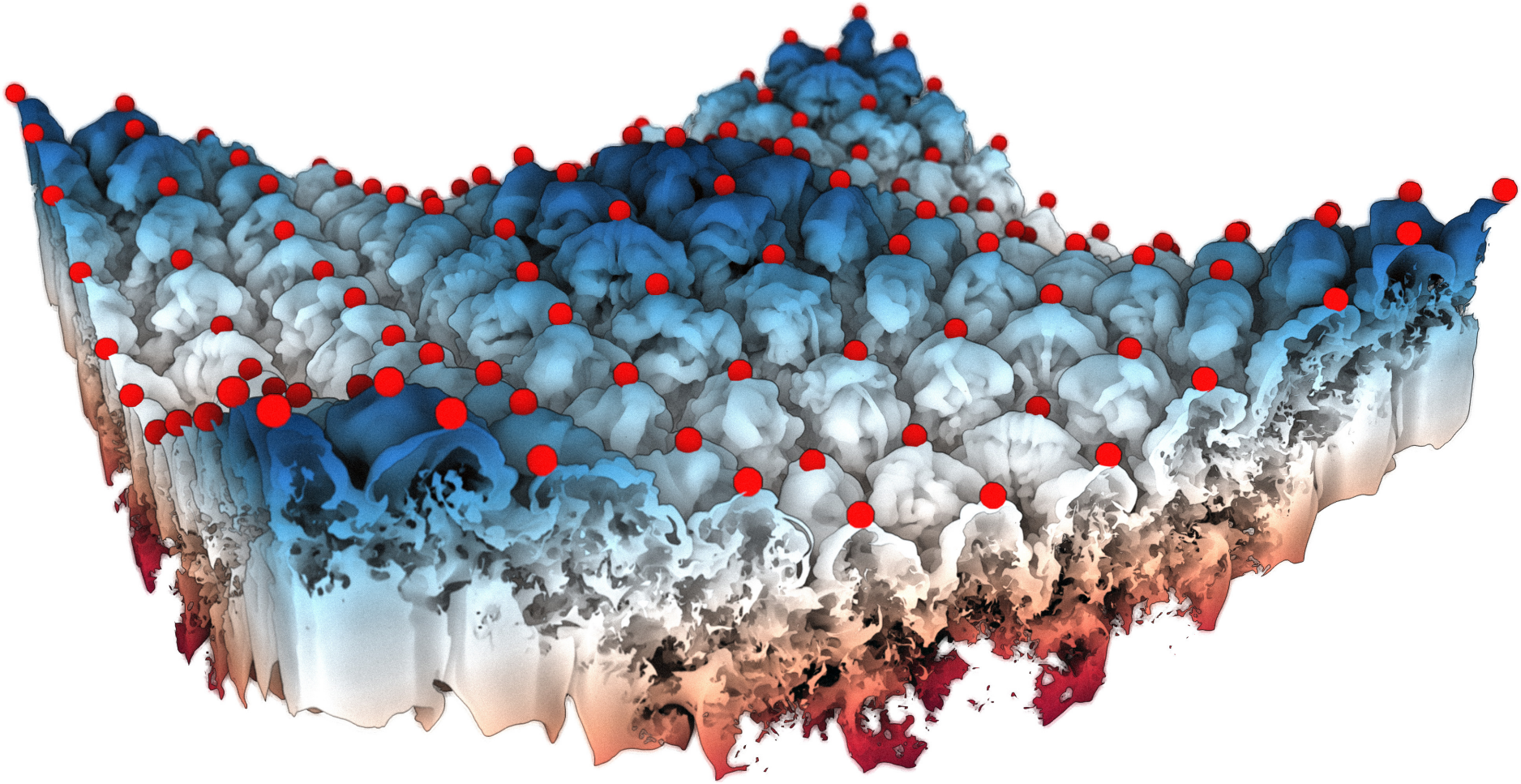
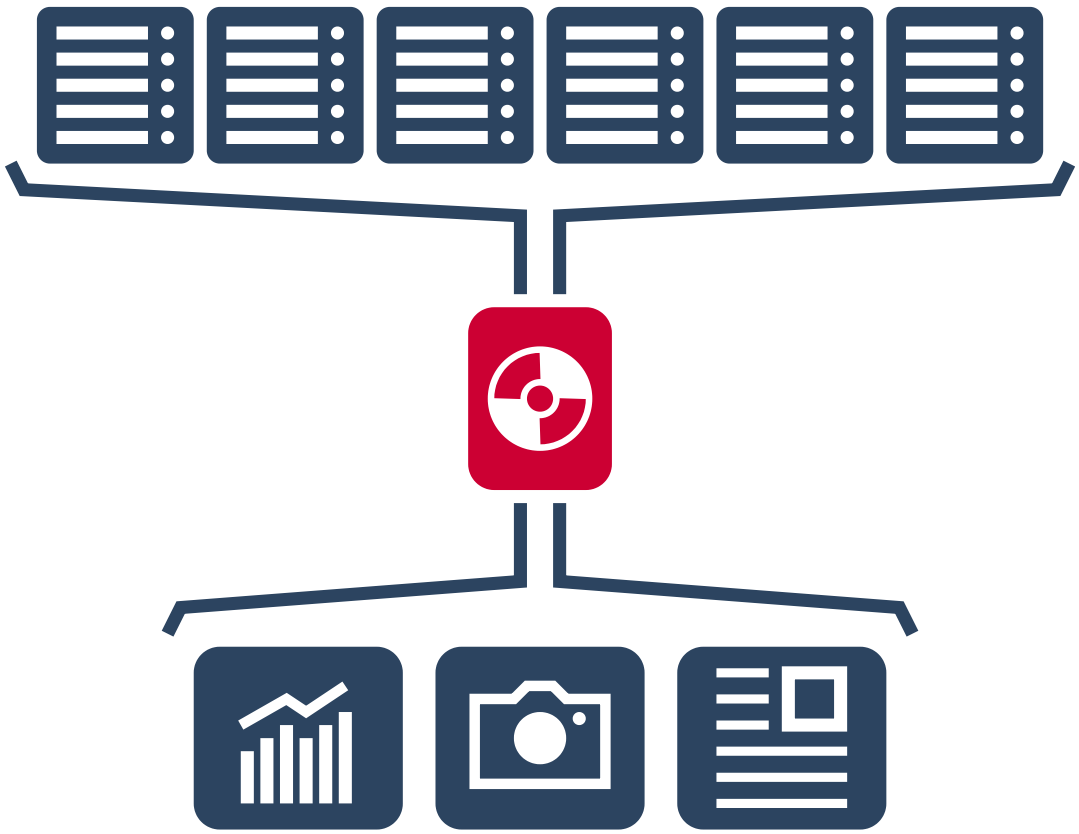
FAIR Cinema Databases

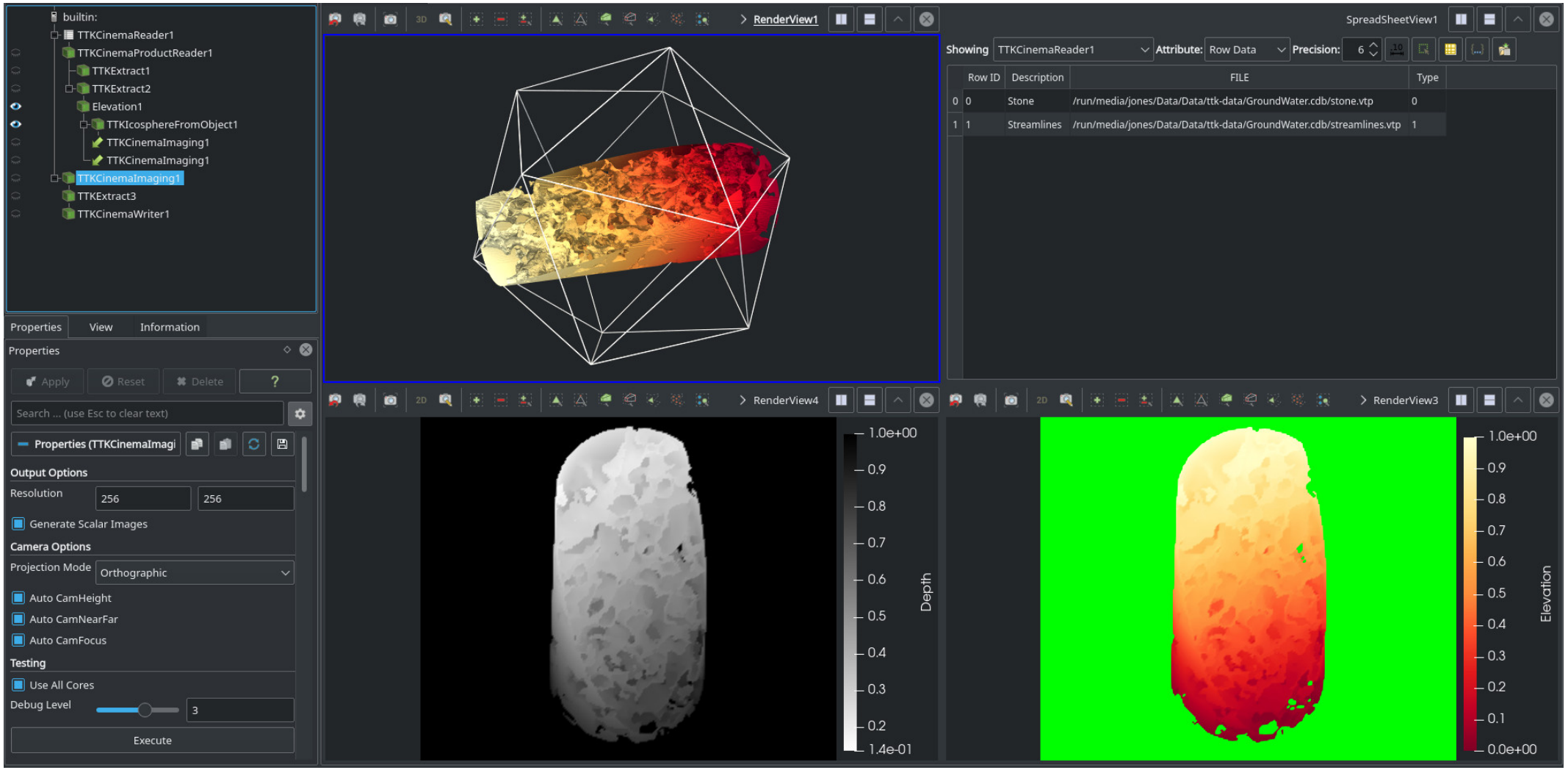


Motivation - Large-Scale Data Visualization



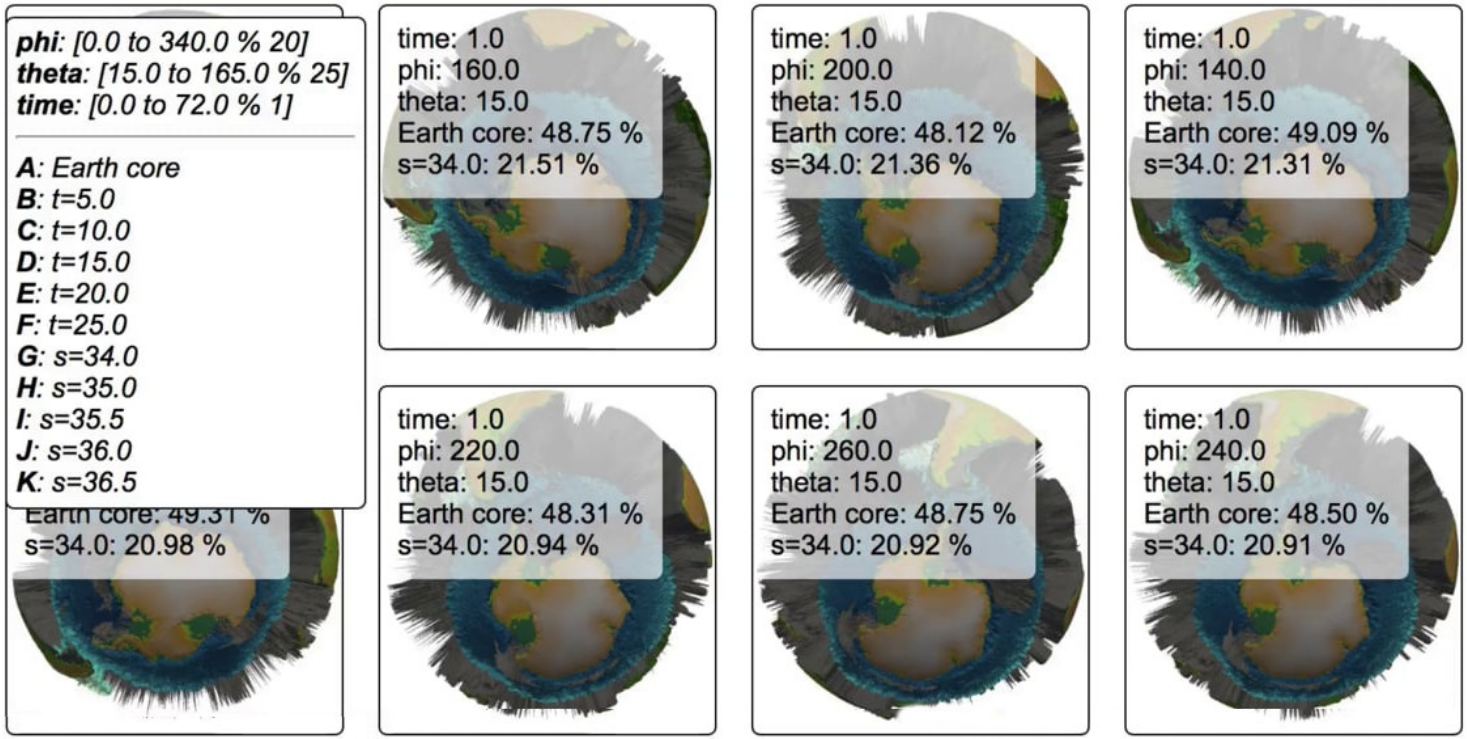
- High-Performance Computing (HPC) environments can produce massive amounts of data.
- Bandwidth constraints limit data storage.
- Store as less as possible while enabling as much flexibility as possible during post hoc exploration.

Motivation - Large-Scale Data Visualization



Motivation - Large-Scale Data Visualization

- Store **in situ** samples of the parameter and visualization space into an image database.
- Resulting databases can be interactively browsed **post hoc** along sampling axes.



[1] J. Ahrens et al. "An image-based approach to extreme scale in situ visualization and analysis". International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 424-434. IEEE Press, 2014.

Cinema Databases - Overview

A Cinema Database is just a folder with the extension ".cdb" that contains

- any kind of data product (e.g., jpeg images, vtk files, binary data, etc.); and
- a "data.csv" file that lists all data products and their associated sample values.

Folder Structure

```
Example.cdb
    ├ data.csv
    ├ image_0.jpg
    ├ image_1.jpg
    ├ image_2.jpg
    ├ image_3.jpg
    ├ image_4.jpg
    └ image_5.jpg
```

data.csv File

```
Time, Phi, Theta, FILE
   0,   0,     0, image_0.jpg
   0,  90,     0, image_1.jpg
   0, 180,     0, image_2.jpg
   0, 270,     0, image_3.jpg
   1,   0,     0, image_4.jpg
   1,  90,     0, image_5.jpg
```

[2] D. Rogers et al., "Cinema Database Specification Dietrich Release v1.2". Technical Report, Los Alamos National Laboratory, 2017.

Cinema Databases - Making them FAIR

Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

TODOs:

- Unique Identifier ⟼ DOI
- Term Ambiguity (e.g., time) ⟼ Ontologies
- Meta-Meta Model ⟼ Unclear
- Versioning ⟼ Git
- Accessibility ⟼ Git
- Provenance ⟼ Unclear
- Reusability ⟼ License