

# Integration of CSaransh with CascadesDB

*Utkarsh Bhardwaj<sup>1</sup>, Christian Hill<sup>2</sup> and Manoj Warriar<sup>1</sup>*

1. Bhabha Atomic Research Centre, India; 2. International Atomic Energy Agency

Cascades Saransh (CSaransh) is an open source project for collision cascades analyses and visualization. We recently integrated CSaransh with IAEA's open database of collision cascades, CascadesDB. CSaransh has two modules. The first module, AnuVikar, takes the xyz files and meta information as input, and outputs a cleaned json file having defect coordinates and other properties such as defect morphologies and sizes, cascade volume and sub-cascade measure etc. for each cascade. It has a number of new algorithms that involve SaVi [1], a Graph based defect morphology identification algorithm, Unsupervised Machine Learning method to visualize and search for similar defect morphologies [2] etc. CSaransh features a single page modern web application to visualize defects and their morphologies, find patterns and correlations, and plot statistical trends. The output of AnuVikar can be loaded to CSaransh web application.

We processed over 2000 cascades from CascadesDB with AnuVikar. It processed this data with size over 140GB in under an hour and produced a 40 MB output file. The dataset that we have currently consists of all the cascades that are initially perfect and have no-surface defects. As with any large dataset, there were also cases where the data had inconsistencies and errors. The errors and warnings are written by AnuVikar in a log file after it processes the data. The errors range from general file corruption to errors in schema, to domain specific errors and inconsistencies between xyz files and meta information. These errors can either result in inability to process cascade or make it stand out as an outlier in the database. AnuVikar can estimate various parameters such as box-dimension, origin, lattice-constant if it finds the given parameters inconsistent with the Xyz file.

In this presentation we will introduce CSaransh with a live demo of the web-application from <https://cascadesdb.org/csaransh>. With integration of CSaransh one can easily browse through the CascadesDB data, and look for different patterns, statistics, and correlations. CSaransh with CascadesDB, provides an easy way to draw data-driven conclusions on collision cascades. We will then discuss the issues we found in the CascadesDB dataset after its validation with CSaransh. The integration of CSaransh with CascadesDB will help in mitigating errors in the database automatically, which becomes more crucial as the data grows.