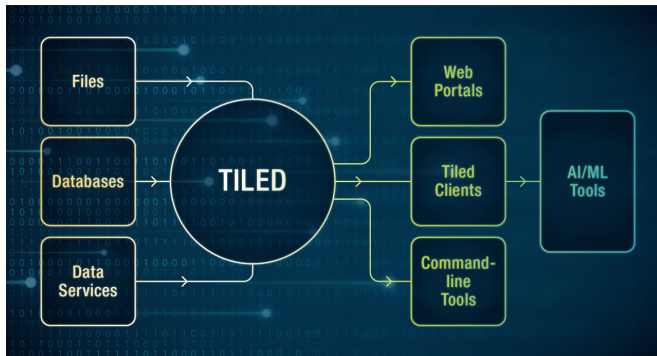


Revolutionizing data access through new software tool: Tiled

24 November 2021



Scientists can use Tiled to seamlessly access data stores across various formats such as files, data bases or other data services. Tiled allows its users to see, slice, and study their data using the most convenient tool for them. Credit: Brookhaven National Laboratory

Every time scientists study a new material for future batteries or investigate diseases to develop new drugs, they must wade through an ocean of data. Today, a whole ecosystem of scientific tools creates a wild variety of data to be explored. This exploration will now get a lot easier thanks to scientists at the National Synchrotron Light Source II (NSLS-II), located at the U.S. Department of Energy's (DOE) Brookhaven National Laboratory. Their freshly rolled-out software tool—called [Tiled](#)—allows researchers to see, slice, and study their data more conveniently than ever before. This new data access tool makes finding and analyzing the right piece of data a walk in the park compared to previous methods, paving the way for the next scientific breakthrough.

As one of the 28 DOE Office of Science user facilities across the Nation, NSLS-II welcomes nearly 2,000 scientists each year to use its ultrabright light, tackling the greatest challenges in materials and life science. These visiting researchers come from around the globe to

collaborate with experts and use the one-of-a-kind research tools at NSLS-II. They zap their samples, ranging from ancient rocks to novel quantum materials, with intense X-rays and catch outgoing signals using advanced detectors. In turn, these detectors spit out streams of data, waiting to be analyzed by scientists.

"Working with data is a central part of all research, and yet a challenge on its own. It comes in a multitude of formats, in varying sizes and shapes, and not every piece of it is useful for the researchers. This is why developing a [software tool](#) that makes accessing, seeing, and sorting through data so important," said Dan Allan, computational scientist at NSLS-II.

Tiled is a data access service for data-aware portals and data science tools. This means that Tiled sits atop databases and file systems so that scientists can access their data through, for example, a web browser or data analysis software. While the Data Science and Systems Integration (DSSI) program rolled out Tiled to all experimental stations at NSLS-II, the service, just like its cousin project Bluesky (a data acquisition software also developed at NSLS-II), can be used in any research laboratory around the globe. This is possible because Tiled is published under a popular open-source software license.

"Even though we developed Tiled in the programming language Python and, therefore, it integrates naturally with data science libraries based on Python, nothing about the service is Python-specific," said Stuart Campbell, chief data scientist at NSLS-II. "The client uses an API, or [application programming interface](#), to connect the user applications with the server. An API is basically a set of rules, or a contract that defines how different software pieces communicate with each other. The great thing about this approach is that once these rules and interfaces are defined, it provides users and developers the structure within

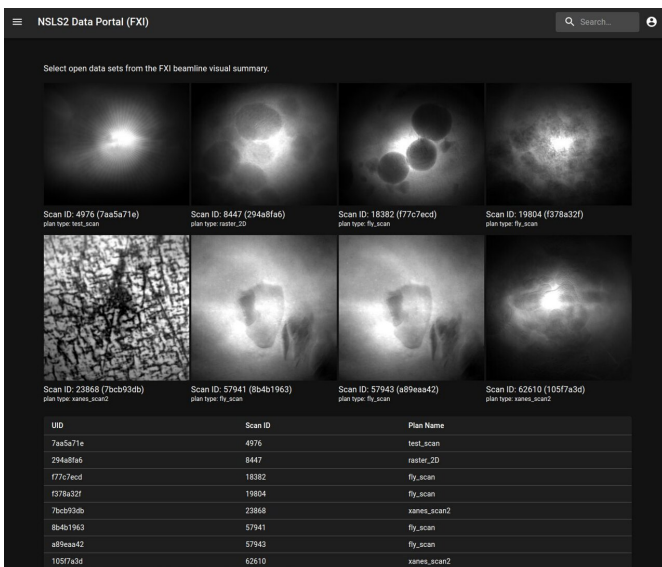
which they can build some excellent tools and expand the functionality beyond that which we had originally imagined."

Tiled's flexibility allows the service to seamlessly integrate with any database or collection of files so that it can be used on a wide range of experiments with very different techniques and data.

Getting your data needs squared away

"In the past, I used to help my Ph.D. advisor to download data from facilities like NSLS-II. It was tedious because we needed to download all of our data at once before we could sort out the useful parts. Additionally, the data were in the format of the detector—regardless of how we wanted to analyze it. This meant after a long download, we had to convert the data before we could even look at it," Allan said.

Campbell added, "If Dan had Tiled back then, he could have easily looked through the data on a web browser or data analysis application, sorted out the good parts, and shared only those of interest with his advisor through a single link."



This preview of the Tiled web client shows how different detector images from different measurements can be displayed at the same time. The preview shows the portal in dark mode. Credit: Brookhaven National Laboratory

By using Tiled, scientists can preview their data and access just the parts they want without a large download. They can also choose the format of their downloaded data or feed it directly into analysis software. At the same time, Tiled offers access control based on web security standards so that all data stay safe. Because setting up a new account can be a barrier, Tiled can be configured to allow third-party services for login, such as Google and ORCID.

"Remote capabilities are more important than ever," said Dylan McReynolds, computing systems engineer at the Advanced Light Source, a DOE Office of Science User Facility located at Lawrence Berkeley National Laboratory, who has collaborated on Tiled. "Building on open, standard web protocols advances our scientific capabilities by making it easy to move data to where it's needed."

The new software even enables a form of "airplane mode" in which the data are stored on a user's laptop so that researchers can continue to work on it offline or with a slow Internet connection.

"Our aim with Tiled is to simplify data access for everyone. If you don't need to worry about converting data formats into other formats or picking information out of file names, you can think about the more important parts, like finding the answer to your research questions," said Thomas Caswell, computational scientist at NSLS-II.

Simplifying and standardizing data access is critical to both optimizing existing workflows and enabling future workflows centered on Machine Learning, AI, and other advanced analytics. These emerging technologies critically rely on frictionless access to data, regardless of how it was collected or stored, to unlock their full potential.

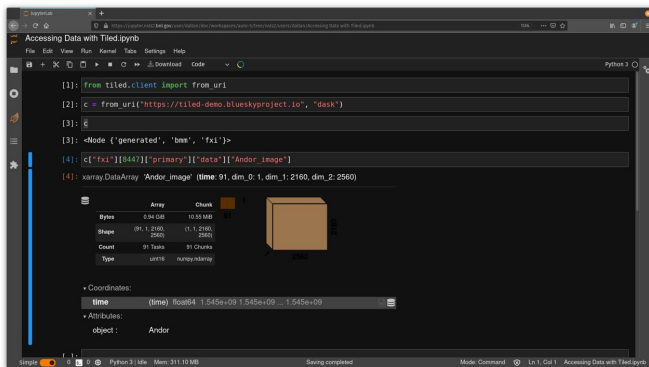
Tiled: Fits into any research puzzle

The first users of Tiled have already built some exciting and sophisticated tools to power their research.

"Tiled offers a completely new way to access the data that will simplify and streamline processing and analysis pipelines for experiments. No more

clunky downloads or wasting time importing data from a dozen formats to analyze an experiment!" said Denis Leschev, assistant physicist at NSLS-II, who tested Tiled. "In addition, Tiled will enable a more straightforward way to share the data, paving the way for more open and transparent science in the future."

The new software is not only available for NSLS-II users: the team designed the software to be adaptable to any data source. It can be deployed at a large scale for facilities like NSLS-II, but it can run just as well on a student's laptop or a research group's workstation. Other laboratories and institutions already have the opportunity to adapt this software for their own needs.



This Jupyter Notebook, a popular data analysis web application, is using Tiled to access data for calculations, processing, and visualization. Credit: Brookhaven National Laboratory

Peter Beaucage, a staff scientist at the National Institute of Standards and Technology (NIST), who is an early user of Tiled, has integrated it with his own scientific data analysis program, PyHyperScattering. He lets Tiled handle data transfer and security details, building on it to provide his users with the specific interface that they need for their work.

"The volume of synchrotron data needed for a typical analysis has expanded dramatically in the last decade, rapidly scaling beyond the capabilities of existing data transfer platforms. Tiled and similar

solutions promise to give users seamless access to the right data at the right time and accelerate discovery based on X-ray science," Beaucage said.

Beyond Beaucage, other users of Tiled also built data analysis pipelines, moving data from live experiments at NSLS-II to remote clusters and into custom software for visualizing and interrogating the data. Each step was supported by Tiled.

"Overall, we are incredibly proud to roll out Tiled. It is the culmination of our work for the last six years. It combines all the features we want in modern data access tools, and it goes hand in hand with Bluesky," said Campbell.

The road ahead

Tiled will enable a whole garden of useful tools to grow for a wide range of techniques. The team has set their eyes on building out various web applications focused on specific research techniques. The team also wants to design a public data interface so that anyone can explore real publicly available data using Tiled.

"Grants often require open data access, but it is difficult for researchers to achieve that in a way that is practical and immediately useful. Tiled lays a track to researchers' door, working with the tools they already use to help them make data findable, accessible, interoperable, and reusable, following the FAIR guiding principles for scientific data management and stewardship," added Allan.

By separating how data are stored from how they are accessed, Tiled unlocks a way to use cutting-edge storage and search technologies on the inside, while presenting researchers with time-tested and established standards. It meets them where they are and leaves them in charge of how to format and work with their data.

"Tiled aims to follow other NSLS-II software efforts in growing a friendly community of contributors and users. We are actively seeking collaboration with facilities and researchers around the world—whether in industry, academia, or government—who have similar challenges, and we are excited to see what we can build together on this platform," said Allan.

More information: Daniel Allan et al, Bluesky's Ahead: A Multi-Facility Collaboration for an a la Carte Software Project for Data Acquisition and Management, *Synchrotron Radiation News* (2019).
[DOI: 10.1080/08940886.2019.1608121](https://doi.org/10.1080/08940886.2019.1608121)

Tiled Documentation: blueskyproject.io/tiled

Tiled Demo (for programmers): tiled-demo.blueskyproject.io/

Bluesky Open Source Project Home Page:
blueskyproject.io/

Provided by Brookhaven National Laboratory

APA citation: Revolutionizing data access through new software tool: Tiled (2021, November 24) retrieved 3 December 2021 from <https://techxplore.com/news/2021-11-revolutionizing-access-software-tool-tiled.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.