

Customizing Spotfire® for AI-Driven Drug Discovery



Customizing Spotfire® for AI-Driven Drug Discovery

For the artificial intelligence (AI)-driven drug discovery company Exscientia, a unified informatics platform is essential to share algorithms, ensure maximum use of data, foster collaboration, and streamline work for its scientists.

Drawing on the power of Spotfire® visual analytics, Exscientia is able to

- track the performance of its machine-learning models
- manage chemical synthesis of new compounds while using comparisons to previous compounds to make better and faster go/no-go decisions
- analyze dose response curves and other data from biological testing
- integrate project data from both a biological and chemical point of view
- harness all this data to maintain a keen focus on productivity.

What has provided Exscientia with even further benefits is the ability to customize the highly flexible Spotfire platform to meet the precise needs of the scientists across the organization. Four main tools have proven advantageous in this customization: Python data functions, IronPython, configuration blocks, and automation services.

Python can be complicated to use.....[Spotfire] makes this complicated functionality available to users without any Python experience. So it's really democratizing access to this complex functionality." Alastair Pate, Associate Director, Data Visualisation and Analytics, Exscientia

Making Complicated Python Functionality Available to Chemists and Biologists

Driven to leverage AI, machine learning, and automation to accelerate drug discovery, Exscientia data scientists often develop analytical tools using software and data storage platforms that the biologists and medicinal chemists are not familiar with or comfortable using. Spotfire Lead Discovery Premium has solved this problem for Exscientia by providing an intuitive platform where researchers can employ the company's AI algorithms in an easy-to-use format.

This integration works smoothly because of Spotfire's ability to share data with Python. A Spotfire data table can be sent to Python for processing, where the data is handled in a Pandas data frame; when the processed data is returned, it is back in the format of a Spotfire data table. The processed data can be embedded into an existing Spotfire dashboard or be presented in a standalone format.

In one example, this capability was used at Exscientia when a computational chemist had output compounds to an AWS S3 server, but the medicinal scientists who needed to work on the compounds didn't know how to access the data. The Exscientia team set up a Python data function that reads a zipped data file from the S3 server as a Python data frame and places the data in a new table in Spotfire, which is known to the medicinal scientists. The users do not need to know anything about AWS S3; they just find the data in Spotfire in their usual workspaces. It's simple, yet powerful for the scientists, who get efficient access to data in a familiar environment to further their research.

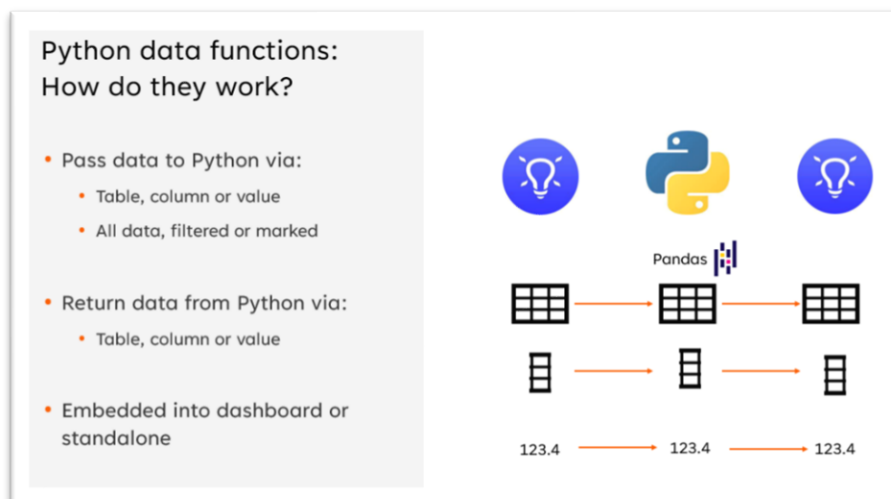


Figure 1 : Illustration depicting the sharing and transfer of data between Spotfire and Python

In another example, Exscientia data scientists had used Python to build a model for absorption, distribution, metabolism, and excretion (ADME) prediction, but the researchers who needed to use the model don't know Python. So the Exscientia team integrated the ADME model into Spotfire, where any scientist can use it. A researcher simply loads in a list of compounds—from a new SD file or from an existing Spotfire dashboard—and chooses the ADME prediction function from an $f(x)$ flyout. The Python data function converts the data and sends it to the company's model platform, where the ADME prediction calculations are run. The ADME predictions results are then sent back to the user in a clean Spotfire data table.

Enhancing Data Visualization with IronPython

IronPython is another tool that Exscientia has integrated into Spotfire Lead Discovery Premium, to enhance data visualization for its researchers. With just a few lines of code, the IT team enabled its scientists to use a slider scale to quickly resize all the chemical structures embedded in a scatter plot, or, with the click of a button, to enable or disable fold lines in correlation plots. Activating the slide or pressing the button executes the IronPython script. Such customization is easy for the IT team and has a big impact on users, who can then visualize data according to changing needs.

Creating Ease and Saving Time with Configuration Blocks

Spotfire's built-in configuration blocks have also facilitated customization at Exscientia. In one example, they used configuration blocks to give users a direct link to a bookmark that accesses public data. In another, researchers wanted to filter a dashboard to show only the latest data. The IT team prepared the filter and rolled it out to production in just two hours. The users were thrilled to start using this capability that saves them time by allowing them to quickly see what has changed since the last time they reviewed the information.

Accelerating Repeated Tasks with Automation Services

The Exscientia team has also made use of Spotfire’s automation services. This feature provides simple scripts that can easily be chained together to automatically accomplish repeated tasks. In one simple but effective example, the Exscientia team used automation services to perform regular dashboard backups. This script opens a dashboard, puts in the current date, and saves it. This ensures that timestamped dashboard backups are available whenever they are needed.

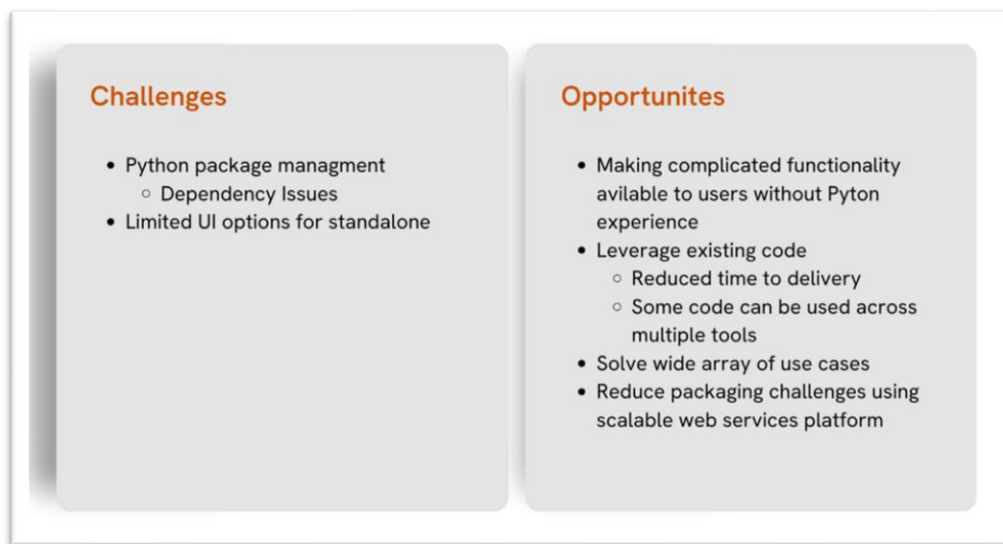


Figure 2 : Challenges and Opportunities in the implementation of Python at Exscientia

Conclusions: Customizing Spotfire/Spotfire Lead Discovery Premium to Increase Collaboration and Productivity

The high degree of customizability of Spotfire has provided big advantages for the AI-driven drug discovery company Exscientia. While keeping drug discovery chemists and biologists working in Spotfire, Exscientia IT has been able to use built-in tools like configuration blocks and automation services to apply bookmarks, filters, and automatic timestamped dashboard backups. Using IronPython, they have enabled researchers to quickly customize data visualizations. By integrating Python Data Functions, they have made complex functionality available to biologists and chemists, who now access the company’s AI algorithms right in their normal workflows, without knowing any code. The ease with which users can customize Spotfire has streamlined scientists’ work, accelerated collaboration, and increased drug discovery productivity at Exscientia.