DATASHEET

PIPELINE PILOT OVERVIEW

Pipeline Pilot is a high-performance scientific informatics platform that optimizes the research innovation cycle, increases operational efficiency, and reduces costs for both research and IT. Pipeline Pilot provides integrated and controlled access to the large volumes of disparate research data locked in silos, automates the scientific analysis of that data, allows users across the enterprise to rapidly explore, visualize and report research results, and facilitates collaboration throughout the wider scientific community.

Pipeline Pilot has the unique ability to comprehend both simple data formats (text and numeric data) as well as complex scientific data types (image, chemical structures, and biological sequences). The breadth and depth of science, along with Accelrys' ecosystem of partners, allows us to meet the varied needs of many research-based industries, including pharmaceuticals, biotechnology, materials science, consumer packaged goods, oil and gas, automotive, aerospace, energy and defense.

WITH PIPELINE PILOT YOU CAN:

- Complete R&D projects significantly faster (10x or more) by leveraging existing research and intellectual property and automating routine data gathering and analysis processes.
- Uncover the knowledge hidden in your data by quickly aggregating and processing massive volumes of structured and unstructured data from multiple disparate research areas in a single environment.
- Encapsulate and deploy best practices to ensure compliance and enable collaboration across your research and development organizations.
- Reduce direct research costs by developing models for discovery based on predictive science.
- Rapidly build and deploy high quality scientific solutions based upon standard technologies and widely accepted science.

Accelrys Pipeline Pilot optimizes the research innovation cycle by providing tools for scientific analysis (in green) and allowing for the automation and standardization of manual, repetitive data preparation and collation tasks (in red). This allows scientists and engineers to spend more time applying their skills and knowledge where innovation takes place.

- Improve decision making with better, faster experimental results through real-time reporting dashboards.
- Reduce R&D support costs by integrating incompatible, point solution software and databases.

POWERFUL DATA PROCESSING

At the heart of Pipeline Pilot, Accelrys' scientific informatics platform, is a powerful data pipelining engine that uses configurable protocols in a rapid application development environment to automate scientific data management, analysis and reporting processes. This powerful capability eliminates highly repetitive, often complex manual processes that consume high-value resources and drive up costs.

Pipeline Pilot is a configurable platform that offers a standardized technology option for research and development IT organizations, lowering the total cost of ownership while enabling scientific innovation.

The combination of a powerful data pipelining engine and a flexible, scientifically-aware architecture enables Pipeline Pilot to uniquely address the challenges faced by research and development organizations. Pipeline Pilot delivers:

- Data management, analysis and reporting for text, numeric, and complex scientific data, including chemical structures, biological sequences, and scientific images
- A rapid application development environment for engineers, developers and domain experts, to assist requirement gathering and prototyping, and lead directly to a final production system
- Configurable components for data retrieval, manipulation, computational filtering and display
- "Build-your-own" component capability leveraging standard technologies including SOAP/Web Services, Perl, Java and command line access
- A flexible, configurable, service-oriented architecture
- Validated scientific components and best practice workflows that cover a broad range of scientific disciplines
- Integration with third-party applications, databases and existing scientific infrastructures including Microsoft SharePoint, Microsoft Office applications, chemical registration databases, image data management systems, online data sources such as PubMed, PubChem, patent databases, RCSB (bio sequence data and protein structures) and BLAST (sequence database). Pipeline Pilot's open architecture provides the ability to access many additional third-party data sources, including those that are ODBC or JDBC and web services compliant.

FLEXIBLE INTEGRATION OF DISPARATE DATA AND APPLICATIONS

Pipeline Pilot allows you to transform disparate data into organized information necessary for better informed decisions across the organization. Using a number of standard technologies, you can bring data together from in-house databases, files, and instruments, collected from across the enterprise, as well as from external sources. You can read chemistry, sequence, text, imaging, and numeric data from all popular formats and analyze data from multiple sources in real-time. For highly efficient on-demand data processing and reporting, you can dynamically create data marts using the Pipeline Pilot extract, transform and load capabilities.

CAPTURE AND DEPLOYMENT OF BEST PRACTICES

With Pipeline Pilot you can encapsulate, annotate and version your institutional best practices, allowing you to document and reproduce the steps used to achieve a particular result. Your resulting protocols can be published and shared with others to facilitate cooperative development and knowledge transfer. To make your automated process available to an even wider community, you can make Pipeline Pilot protocols available via the browser-based Pipeline Pilot Web Port interface, portal technologies such as SharePoint, as well as many other deployment environments.

CUSTOMIZED REPORTING AND WEB APPLICATION DEVELOPMENT

You can create customized reports that summarize your data analysis and mining protocols with a full range of text, tables, charts, and images. With complete control over the layout and content, you can easily interpret and communicate your results. By displaying multiple tables, charts and images in a single report, you can see different views of your data, including side-by-side comparisons from different sources and data that was processed in alternate ways. You can deploy your reports in a wide variety of ways, including HTML, PDF, Word, Excel, and PowerPoint.

To make your reports more dynamic you can add interactivity, allowing you to link within, between, and outside reports. You can further extend the interactivity to create web applications that link several protocols together into a functional unit. This greatly

extends the versatility of your protocols, allowing end-users to provide input data and determine the way the data are presented, without writing protocols for themselves.

PIPELINE PILOT CLIENTS

Pipeline Pilot Professional Client

The Pipeline Pilot Professional Client is the authoring tool for creating data pipelining protocols, providing users with the full power of Pipeline Pilot. The Professional Client can be used to create and modify protocols for your own use, and also to publish them for others to use. The Professional Client can also be used to create new components, or edit existing ones, to meet the varied needs of your organization. By configuring integration or scripting components to integrate with external tools, you can standardize access to third-party data sources and applications, making them available to a wider set of users.

Pipeline Pilot Web Port

Pipeline Pilot Web Port is a browser-based environment allowing the power of Pipeline Pilot to be extended to a broad community of users. The same protocols that Pipeline Pilot Professional users create and run can be made accessible to Pipeline Pilot Web Port users, thereby exposing the full capabilities of Pipeline Pilot to a wide audience through a simple web-based interface. Web Port users can browse available protocols and parameterize them to run their own data to suit their unique needs, without having to understand how the protocols were built. Users of Web Port can also access and execute protocols via any other web-based interface, such as Microsoft SharePoint, RSS feeds, web-links, and custom interfaces developed with any of the Pipeline Pilot client SDKs.

PIPELINE PILOT COMPONENT COLLECTIONS

Pipeline Pilot's Component Collections are the "building blocks" of the scientific informatics platform and are grouped by category of science or function. The collections contain numerous components which allow researchers, developers and IT professionals to perform both science-specific and generic data-processing functions. By graphically combining components, you can construct workflows for data retrieval, filtering, analysis, and reporting.

Feature	Professional Client	Pipeline Pilot Web Port
Use protocols published by others	✓	✓
Execute existing protocols	✓	✓
Edit protocol parameters	✓	✓
Monitor protocol execution progress	✓	✓
View protocol layout and logic	✓	
Compose new protocols	✓	
Modify existing protocols	✓	
Manage jobs and results	✓	
Create new components	✓	
Edit component interface	✓	
Integrate external tools	✓	
Employ script languages (incl. PilotScript)	✓	
Publish protocols for use by others	✓	
Command-line access for scheduling	✓	

CHEMISTRY

Chemistry Component Collection

Analyze, profile, and manage the compounds in your corporate database with chemically intelligent filters and learning. Compare your compounds directly with a vendor's flat files to identify novel molecules with appropriate profiles determined by calculated properties. Enumerate and analyze combinatorial libraries at high speed using reactions or Markush library definitions. Identify common cores and group hitlists by chemical families. Execute chemical analysis steps by seamlessly using Pipeline Pilot and third-party processes together to automate a complete analysis.

ADMET Component Collection

Calculate predicted absorption, distribution, metabolism, excretion and toxicity (ADMET) properties for collections of molecules such as synthesis candidates, vendor libraries, and screening collections. The collection includes models for human intestinal absorption, aqueous solubility, blood brain barrier penetration, plasma protein binding, cytochrome P450 2D6 inhibition, and hepatotoxicity.

Cheminformatics Component Collection

The Cheminformatics Collection is designed to make it easier for you to build database-focused search and browse applications using Pipeline Pilot. The collection is comprised of two sub-systems: List Management and Query Services (LMQS) - a database search and hit list management system, and Pipeline Pilot Chemistry Cartridge - a high performance data cartridge that enables you to store and search chemical structures and reactions in Oracle. Together, these powerful systems extend the capabilities of Pipeline Pilot to provide you with a strong foundation on which to build a variety of enterprise cheminformatics applications.

BIOLOGY

Gene Expression Component Collection

Visualize, analyze, annotate, and report on gene expression experiments, including the individual target genes. Core functionality is based on BioConductor—the open source software for the analysis and comprehension of genomic data. By harnessing the graphical protocol building capabilities of Pipeline Pilot, the Gene Expression Collection allows you to construct complex workflows without writing code, while also making it easy for you to couple gene expression analysis with other Pipeline Pilot-based processes, such as sequence analysis, text analytics, and reporting.

Sequence Analysis Component Collection

Analyze, annotate, and compare biological sequences in an environment where modular tools can be graphically linked together to create practical bioinformatics routines. From over 100 components for sequence analysis, use familiar algorithms like BLAST and ClustalW. You can extend the system yourself by making new components with Java or Perl, or by extending our integration with EMBOSS.

Mass Spectrometry for Proteomics Component Collection

The Mass Spectrometry for Proteomics Collection offers a comprehensive set of components and example protocols to create and automate customized proteomics and metabolomics workflows. The protocols capture your processing steps and flow your experimental data through components to read, write, visualize, manipulate, analyze, compare, and publish results. The data models support quick processing of large files and carry detailed information at multiple levels. Harnessing the power of the Pipeline Pilot graphical protocol-building capabilities, complex data analysis procedures, such as peptide and protein identification and differential profiling, can be easily constructed without the need to develop code.

LIFE SCIENCE MODELING & SIMULATION

Catalyst Component Collection (Pharmacophore)

Access a comprehensive set of tools for pharmacophore modeling and 3D database management. The renowned technologies of Catalyst® are packaged into the Pipeline Pilot platform for creating automated, easy-to-use workflows that streamline your pharmacophore modeling and analyses. Create customized workflows to deploy Catalyst's sophisticated algorithms to perform conformer generation, 3D database creation, pharmacophore hypothesis generation, virtual screening, and more.

CHARMm Component Collection (Simulations)

Perform simulations on biological molecules using a powerful set of components based on the well validated CHARMm engine. This collection of components extends the standard capabilities of Pipeline Pilot to include stable and accurate molecular mechanics and molecular dynamics simulation of proteins, nucleic acids, small molecules, and protein-ligand complexes.

MATERIALS MODELING & SIMULATION

Materials Studio Component Collection

Access Materials Studio's® premier modeling capabilities within the Pipeline Pilot data pipelining environment. The Materials Studio Component Collection allows access to analytical capabilities such as Reflex and Reflex QPA, key functionality of the QSAR Plus package, and a utility to integrate scripting applications developed within Materials Studio's scripting API into Pipeline Pilot protocols.

Polymer Properties Component Collection (Synthia)

The Pipeline Pilot Polymer Properties Collection provides a method for quickly estimating the properties of bulk, amorphous homopolymers and random copolymers based on repeat unit information, molecular weight and temperature. The collection contains published models from J.Bicerano's Prediction of Polymer Properties (Marcel Dekker, NY, 2002.), as well as extensible models for proprietary compounds or focused training sets. To use the Polymer Properties Collection, all you have to do is specify the monomer repeat units representing your polymer and select the properties you wish to predict.

REPORTING & VISUALIZATION

Reporting Component Collection

The Reporting Collection offers a comprehensive set of components to create customized reports and web applications that you use to display the results of your data analysis and mining protocols. With complete control of what content you include and how to lay it out, you can create highly effective communication tools to share knowledge with your colleagues. Reports can include text, tables, charts, and images, and can be created in a variety of formats, including web pages, Microsoft Office applications, and PDF. By adding interactivity to your web reports, you can create a rich environment to explore your data. These web applications can be shared with colleagues to view either your own or others' data. Web applications can be deployed in Web Port, SharePoint, via RSS, or in a simple web browser.

DATABASE & APPLICATION INTEGRATION

Integration Component Collection

Incorporate existing in-house or third party programs as computational services in Pipeline Pilot using Java, Perl, VBScript, or SOAP with the Integration Collection. Also, retrieve data from Oracle using ODBC, and JDBC technologies for analysis or reporting, and store results directly back in your own corporate database.

IMAGING

Imaging Component Collection

The Imaging Collection brings the power of Pipeline Pilot's workflow automation and visual programming capabilities to enterprise image informatics, delivering sophisticated capabilities for enhancing, processing, analyzing, integrating, cataloguing, searching and reporting image data. The Imaging Collection integrates image data with numerical, chemical, graphical and textual data in a unified computing framework. Template protocols allow rapid development of custom drill-down reporting and image link capabilities. Advanced supervised and unsupervised learning components offer best in class processing capabilities. With the Pipeline Pilot open platform, almost any standard image file format can be accommodated.

ANALYSIS & STATISTICS

Data Modeling Component Collection

Carry out powerful data modeling with this comprehensive suite of learning and data modeling tools, statistical filters, and clustering components optimized for large real-world data sets. This collection of components extends Pipeline Pilot's standard capabilities to include statistics and predictive modeling for data mining applications. Access powerful methods, such as fast data clustering, Bayesian learning, principal component analysis, linear regression, and partial least squares regression. Perform structure-activity modeling, compound clustering, and maximal common substructure search.

Advanced Modeling Component Collection

Access advanced data modeling methods for Recursive Partitioning (RP) and Multiobjective Pareto Optimization. A variety of RP methods are available in the collection including both single tree and forest of tree learners. The methods can learn on single or multiple response variables. The Pareto Optimization components include methods for multi-objective optimization problems to provide solutions whose criteria trade off amongst two or more partially conflicting goals.

R Statistics Component Collection

Perform insightful analyses, create informative graphics, and make educated decisions. The R Statistics Collection includes components that implement statistical methods for data manipulation, clustering, learning, classical and exploratory data analysis. The underlying statistical engine is the widely used public domain R statistics package. Apply R statistical analysis and data manipulation methods to Pipeline Pilot data streams. You can then incorporate output results from R directly into your pipeline for further analysis using other components in the Pipeline Pilot framework. A sampling of learning methods includes partial least squares, neural nets, and support vector machines. You can easily extend the collection by adding in your own R scripts into an R Custom Script component.

DOCUMENT SEARCH & ANALYSIS

Text Analytics Component Collection

Use text mining to gain critical insights that complement your experimental results. Through document search, characterization, and analyses, your processing routines can be augmented with new information. Achieve your most challenging text mining objectives by linking together search, extraction, and characterization steps into automated routines. Integrate literature mining with your existing scientific protocols, and run them interactively or automatically every night.

ChemMining Component Collection (Chemical Text Mining)

Survey the IP landscape and avoid duplicating research efforts by using ChemMining to perform chemically-intelligent text mining of public and internal documents. The ChemMining Collection lets you find chemical names in text documents and convert them to live chemical structures ready for further manipulation. You can explore a few documents at once and create a report showing the structures found. Alternatively, you can process a large set of documents and create a searchable database of structures found. Combine your chemical structure searches with textual searches for key terminology, such as terms related to disease processes or biological molecules, to perform truly context-specific, chemically-intelligent searching.

LABORATORY

Plate Data Analytics Component Collection

Plate Access methods to read, write, report, visualize, manipulate and perform calculations on plate data. The Plate Data Analytics collection allows each record on the data pipeline to carry an entire plate and its associated wells, and to perform both plate and well level operations. Harnessing the power of the Pipeline Pilot graphical protocol building capabilities, complex data analysis procedures, such as analyzing the results of a screening experiment, can be easily constructed without the need to develop code. Coupled with the integration collection, this plate-based data can easily be inserted into or retrieved from a database.

Analytical Instrumentation Component Collection

The Analytical Instrumentation Collection enables Pipeline Pilot to access, process and share data generated in your analytical lab. With this collection you can streamline the processing of analytical data, generate reports and share results more easily with other departments. The collection supports common data processing operations such as, peak identification, peak integration, line width analysis background detection and removal, interpolation, truncation, scaling and smoothing of spectra, subtraction of spectra, general purpose Fourier transform as well as NMR specific functionality.

To learn more about Pipeline Pilot, go to
accelrys.com/pipeline-pilot