

Pipeline Pilot™

强大的信息整合和流程定制的科学平台

Solutions for Innovation.

信息化建设领导者

目录 / CONTENTS

强大的数据处理能力	03
灵活的数据与应用程序整合能力	04
最佳流程的共享和部署能力	04
可定制的报告和网络应用的开发能力	04
系统架构和客户端类型	04
拓展和提升 Biovia 应用程序的能力	05
Pipeline Pilot 的主要科学模块和说明	06
■ Chemistry 化学	06
■ Biology 生物	07
■ Life Science Modeling & Simulation 生命科学建模和模拟 *	09
■ Materials Modeling & Simulation 材料模拟	10
■ Reporting & Visualization 报告和可视化	11
■ Database & Application Integration 数据库和应用程序整合	11
■ Imaging 图像处理	12
■ Analysis & Statistics 分析与统计	12
■ Document Search & Analysis 文档挖掘和分析	13
■ Laboratory 实验室数据分析	15
系统要求	17

BIOVIA Pipeline Pilot 是一个图形化的科学创新应用平台，通过实现科学数据的自动化分析，让用户能够更好更快地获取企业内部资源，以及可视化、可定制的研究结果报告，从而优化科研创新流程，提高工作效率，同时降低研究和 IT 的成本。通过 Pipeline Pilot，用户不仅能够整合和挖掘杂乱无章的海量研究数据，自动化数据的分析流程，而且还可以实现企业级的研究成果快速分析、可视化与共享，提升大范围的协作能力。

Pipeline Pilot 既支持简单的数据格式(文本和数字)，又支持复杂的科学数据类型(图像、化学结构和生物序列)。科学的广度与深度，再加上 Biovia 公司广泛的合作伙伴，使得 Pipeline Pilot 能够满足众多研究性工业的各种需求，包括制药、生物技术、材料科学、个人消费品、石油、天然气、汽车、航空航天、能源和国防。



Pipeline Pilot 优化研究创新周期：为科学数据分析（绿色）提供专业的分析工具；实现数据准备和数据采集工作（蓝色）的自动化和标准化。让科学家和工程师能够将更多的时间和精力投入到真正研究工作中，从而更多地产生知识创新。

Pipeline Pilot™

强大的信息整合和 流程定制的科学平台

■ 通过 Pipeline Pilot，用户可以

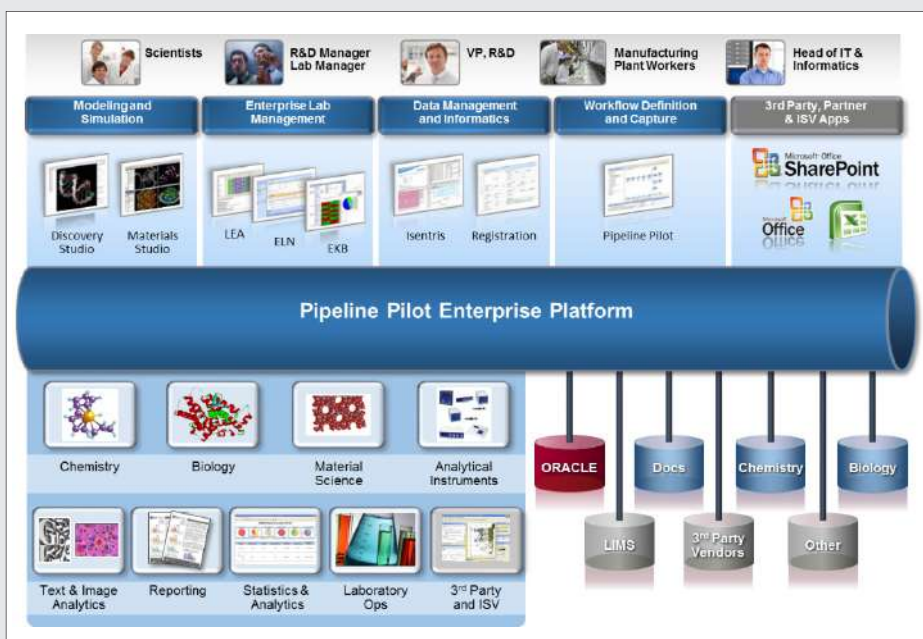
- 通过利用已有的研究成果和自动化日常的数据采集和分析过程，更快地完成研发项目，速度通常可以提高 10 倍，甚至更多。
- 通过在单一环境中快速整合和处理来自多重研究领域的海量结构性和非结构性数据，发现隐藏在数据里面的知识。
- 把最佳的方法和流程封装并部署在团队内部，以确保在整个研发团队内的兼容性和协作性。
- 根据建立用于发现的模型进行预测，减少直接的研发费用。
- 基于标准技术和广泛接受的科学方法，快速建立和部署高质量的科学解决方案。
- 根据实时报告得到更好更快的实验结果，从而提高决策能力。
- 通过整合不同软件工具和数据库来提高软件的应用能力。

强大的数据处理能力

Pipeline Pilot 能够图形化地定制符合您的要求的数据管理和信息挖掘等科研工作流程。Pipeline Pilot 中包含了上千个不同功能的组件 (Components)，您可以根据自己的需求，将其组合成不同的数据处理流程 (Protocols)，实现各种功能，如数据检索和处理，化合物各种性质的计算、筛选与显示等。Pipeline Pilot 平台还提供了 Web 服务层接口，允许您将建立的数据处理流程部署在服务器上，通过 Web 向外界提供服务。

Pipeline Pilot 拥有强大的数据处理引擎和灵活的架构，这使得它能完美地解决研发机构面临的众多难题。Pipeline Pilot 具有以下功能：

- 对文本、数字和包括化学结构、生物序列和科学图像在内的复杂数据结构进行数据管理、分析和生成报告。
- 对工程师、开发人员和领域专家来说，Pipeline Pilot 是一个快速的应用开发环境，它帮助收集用户需求并将之模型化，从而直接形成一个最终的产品系统。
- 提供可用于数据检索、处理、过滤和显示的可配置组件。
- 使用标准技术来构建自定义组件，这些标准技术包括 SOAP/Web Services、Perl、Java 和命令行方式。
- 提供一个灵活的、可配置的、面向服务的架构。
- 提供大量已由实践检验的科学组件和最佳工作流，这些组件和工作流涵盖了众多的科学领域。
- 能够整合第三方应用程序、数据库和已有的常用科学工具，包括微软 SharePoint、微软 Office 应用程序、化学制品注册数据库、图像数据管理系统、在线数据库（如 PubMed 和 PubChem）、专利数据库、RCSB（生物序列数据和蛋白质结构）和 BLAST（序列数据库）。Pipeline Pilot 的开放式架构使得它能够整合大量其它的第三方数据资源，包括 ODBC、JDBC 或网络服务兼容的那些资源。



灵活的数据与应用程序整合能力

Pipeline Pilot 通过一系列的标准化技术，能帮助您将各种不同格式的数据，如内部数据库，文档或任何从企业中获得的数据组织成统一有序的信息，为您的决策提供有力依据。对于科研工作者，Pipeline Pilot 还可以读取各种通用格式的化学、生物序列、文本、图像和数字信息并对它们进行实时的分析。同时，通过 SOAP、Perl、Java、VB Script 和命令行等方式，研究人员可以把工作相关的第三方软件整合成为新的组件，建立更符合自己科研习惯和要求的数据处理流程，这些新组件将被自动整合为系统模块组件的一部分，能以系统原有组件相同的方式与其他用户共享使用。

最佳流程的共享和部署能力

在 Pipeline Pilot 中，可以对数据处理流程实现自动获取，注释和版本编号，还可以方便地将建立的数据处理流程发布给其他研究人员，实现流程共享，使得团队合作开发变得更加方便。为了进行更广泛的交流，您甚至可以将您的数据处理流程通过 Pipeline Pilot 的 Web 接口发布到网络上，提供基于浏览器的使用方式，供更多人的交流和使用。

可定制的报告和网络应用的开发能力

在 Pipeline Pilot 中，可以使用表格、图表、图片和文字等定制数据报告，对数据分析和数据挖掘过程进行总结和展示，并按照您的意愿设计合适的报告布局和报告内容，从而更深入地解释所获得的结果。您在同一张报告中使用时表格、图表和图像，可以从不同的视角上观察同一份数据，或使用不同的方式对不同来源的数据进行相互比较，从而得出合理的结论，为最终决策提供依据。您还可以使用多种格式输出报告，包括 HTML、PDF、Word 和 PowerPoint。

为了让您的报告更加生动，您可以在其中加入 Web 交互元素，如报告内加入超链接，各种信息提示等。您甚至可以将几个数据处理流程整合成为一个网络应用程序，由最终用户来决定显示哪些数据和用哪种方式来显示数据，而不需要最终用户设计任何数据处理流程。

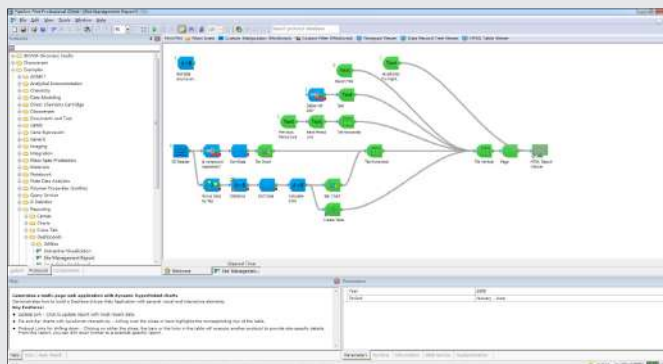
系统架构和客户端类型

Pipeline Pilot 采用灵活的系统架构，其服务端支持 Windows 和 Linux 平台，并可以按 Client / Server（客户端 / 服务器）和 Browse / Server（浏览器 / 服务器）两种模式来部署，使得各种流程和任务能够方便地跨平台（Windows/Linux）运行。

Pipeline Pilot 拥有两种不同的客户端使用模式和界面：

- Pipeline Pilot Professional Client

Pipeline Pilot 专业版客户端能够使用 Pipeline Pilot 所有功能。在该客户端模式中，您可以创建和修改工作流来满足特定的要求，并可以将开发出来的流程发布给其他人共享使用。基于专业版客户端，用户还可以用来创建全新的组件，或者在已有的组件上进行修改，来满足科研及数据分析的需求。



Pipeline Pilot™

强大的信息整合和 流程定制的科学平台

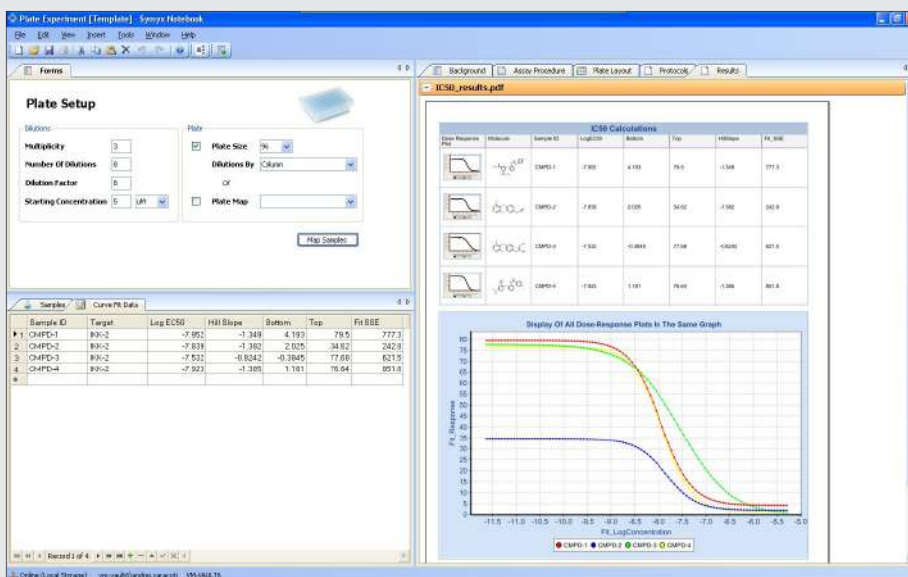
- Pipeline Pilot Webport Client

Pipeline Pilot 网页客户端允许您通过浏览器即可轻松地浏览和使用 Pipeline Pilot 中已发布的数据处理流程，并根据自己的需求设置合理的参数，从而完成相应的数据分析。Pipeline Pilot 的网页客户端使您可以在多种 Web 环境下使用 Pipeline Pilot，如常用的浏览器，Microsoft SharePoint，RSS feeds，网页链接，以及用 SDKs 所开发的任何定制化客户端。



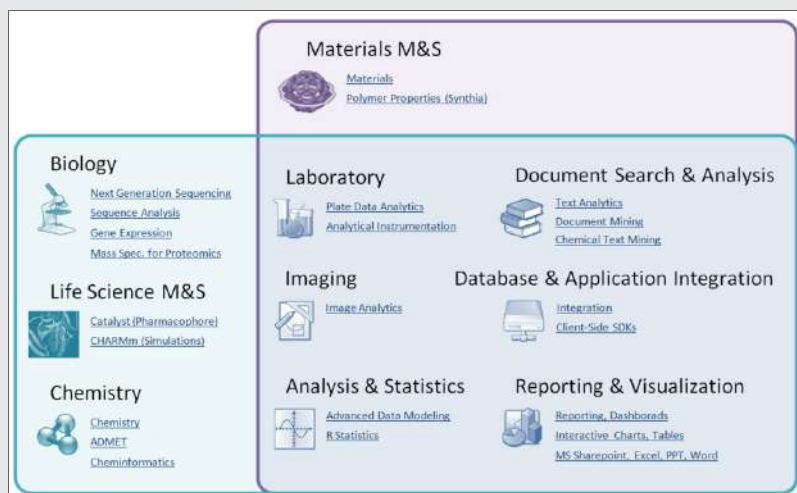
拓展和提升 Biovia 应用程序的能力

Pipeline Pilot 不仅是提升个人和团队工作效率的有力工具，同时也可以用于扩展和提升各种 Biovia 应用程序的应用能力，如 Insight, Insight for Excel, Workbook (ELN), Notebook 和 EKB 等。大部分 Biovia 应用程序都提供有相应的接口，使得 Pipeline Pilot 可以为其提供数据分析服务，如生成分析图表或者是完整的报告，从外部文件、数据库和网络资源中加载数据，与其他应用程序和工具的整合等等。Pipeline Pilot 的强大的科学数据分析处理能力为这些程序的应用能力拓展提供了无限可能。而且这种集成往往是无缝的整合，最终用户通常感觉不到他们正在运行的 Pipeline Pilot 流程。



Pipeline Pilot 的主要科学模块和说明如下

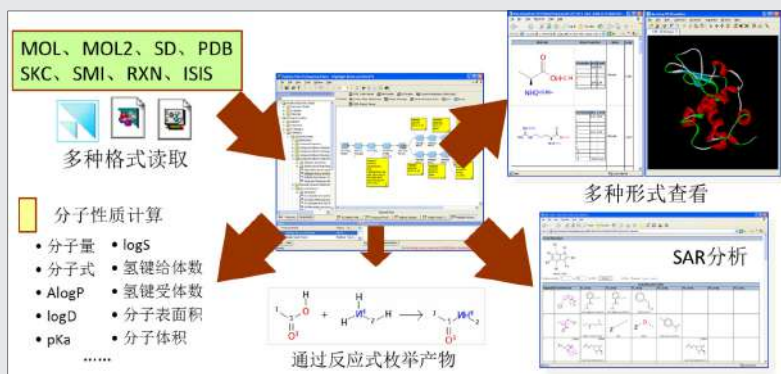
Pipeline Pilot 组件是构成科学信息平台的一个一个“积木”，它们被按照科学领域或功能进行分类。研究人员、开发人员和 IT 专家通过这些为数众多的组件既能执行特定的科学分析，也可以进行常见的数据处理。他们以可视化的方式搭建分析流程，从而实现数据检索、筛选、分析和生成报告等功能。



■ Chemistry 化学

• Chemistry Component Collection

Chemistry 模块提供了多种对化合物的筛选和学习方法，实现化合物的性质计算、分析和处理等功能，并能将您的化合物与其他数据库中的化合物进行对比，发现其中性质独特、结构新颖的分子。另外还能快速枚举和分析组合化学的反应产物，识别一系列化合物中的公共子结构，实现第三方处理程序与 Pipeline Pilot 的无缝连接，自动执行对化合物的全分析过程。该模块支持绝大多数的工业标准化学格式，包括来自 MDL 的 MOL、SD、RG、RXN、ISIS，来自 Daylight 的 SMILES、SMIRKS、MARTS、TDT，以及常用的 MOL2 等各种表达化学结构的文件格式，并可实现各种格式之间的转化。



谁将从 Pipeline Pilot 中受益？

化学家

提高关键研究和开发过程的效率，包括化学制品的注册、筛选和数据分析。更容易和有效地访问、分析和展示化学制品信息。

在一个灵活的环境下进行强大的特定化学分析和生成报告。

管理化合物列表，收集相关信息，通过我们的 MedChem 应用案例进行分析和纯化

生物学家

解决从生物疗法到转化型研究多个生物科学领域的问题。

提高了知识共享、决策和工作能力。

可以对所有和图像相关的数据进行加强、整合、分析、分类和生成报告。

通过 Accelrys 的 Biologics

Registration system 进行注册、鉴定、更新、追踪、关联和查询已有的生物制品。

信息学家

数据流的拖放操作方式使得直接地调试和迅速部署流程成为可能。

与统计工具的紧密结合提高了预测结果的可靠性。

提高用户数据模型的质量。

IT 和开发人员

在一个科学信息平台上整合自己开发的和商业的应用程序。

能读取和分析所有的科学数据格式。

快速建立和部署网络应用服务。

Pipeline Pilot™

强大的信息整合和 流程定制的科学平台

通过 Chemistry 模块，用户可以：

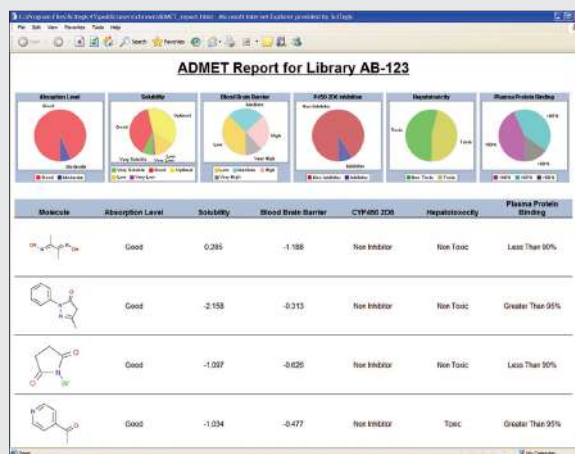
- 兼容大部分工业标准化学数据格式
- 计算众多的分子属性和指纹
- 标准化和处理分子结构
- 通过反应和 RGroup 枚举组合化合物库
- 枚举产生生物等电体库
- 使用 Perl 和 Java 工具包来创建新的化学组件
- 多个化合物库的标准化和比较
- 子结构和相似结构检索
- 化合物的聚类分析和多样性分析
- SAR 分析和最大公共子结构检索

• ADMET Component Collection

ADMET 模块能够计算化合物在生物体内的吸收、分布、代谢、排泄和毒性特征，包含的模型有：人体肠内吸收预测模型、水溶性预测模型、血脑障碍渗透性预测模型、血浆蛋白结合率预测模型、细胞色素 P450 2D6 抑制率预测模型和肝毒性预测模型，此外，本模块还能够预测已知药物的代谢产物。

通过 ADMET 模块，用户可以：

- 快速计算化合物的 ADMET 性质
- 排除 ADMET 特性不佳的化合物
- 在合成之前，评估用于提高化合物 ADMET 属性的纯化方法
- 在化合物分析报告里整合 ADMET 属性、实验和其它计算属性



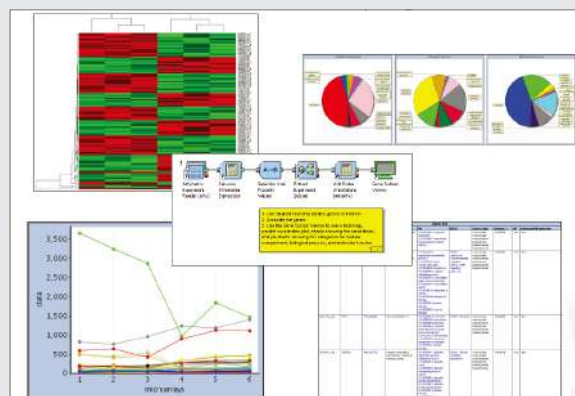
■ Biology 生物

• Gene Expression Component Collection

Gene Expression 模块在开源软件 BioConductor 的基础上，实现了基因表达实验的数据处理、可视化、分析、标记和报告生成等一系列功能。利用 Pipeline Pilot 的图形化数据处理流程设计能力，与 BioConductor 引擎相结合，可建立复杂的工作流程，同时也可将基因表达分析与 Pipeline Pilot 的其他功能的联合，如序列分析、文本分析和报告生成等，以完成更复杂的工作，而这一切都无需任何编程工作。

通过 Gene Expression 模块，用户可以：

- 分析和注释基因表达实验
- 无需编程和熟悉 R 语言，就能够使用 BioConductor 工具
- 很容易地创建用于比较不同分析方法的流程
- 整合 R/BioConductor 和其它的 Pipeline Pilot 模块的分析，包括 R Statistics 和 Sequence Analysis 模块
- 创建包含 R 产生的图表和 Reporting 模块分析结果在内的综合分析报告
- 下载和分析 GEO 数据集

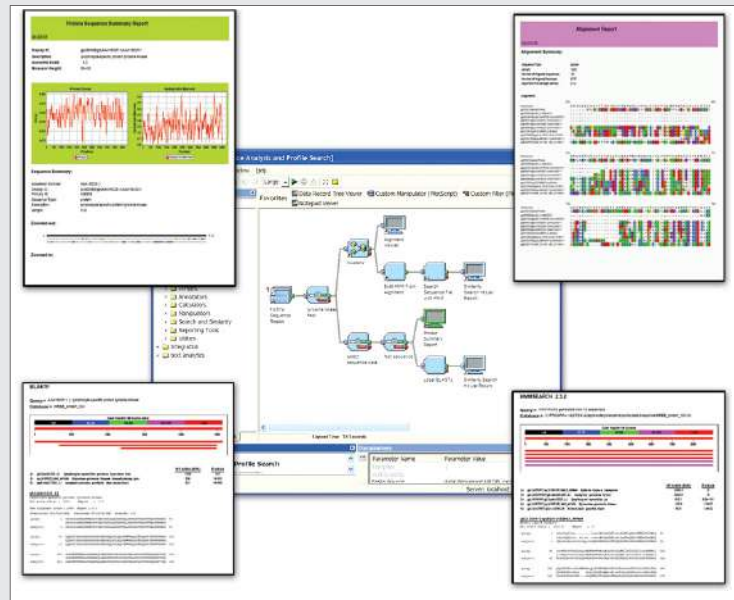


• Sequence Analysis Component Collection

Sequence Analysis 模块通过 180 多种的序列分析组件提供了一整套的生物信息学工具，将模块化工具与图形化的设计方式结合，实现对蛋白质和核酸的序列的注释和特征分析，序列搜索和比对，产生引物和扩增子，siRNA 设计，发现限制性酶切位点等功能。同样，本模块可以通过 Java 和 Perl 建立新的组件来扩展系统功能，或通过整合而实现功能的扩展。

通过 Sequence Analysis 模块，用户可以：

- 自动化新序列数据的常规分析
- 跨物种比较基因和基因组
- 在分析流程中整合用户的算法
- 执行 ClustalW、Muscle、HMM 和 BLAST 分析
- 设计 siRNA 分析流程
- 在分析流程中整合 E-Utilities 和 Entrez 查询功能
- 和其它模块结合进行更复杂的分析

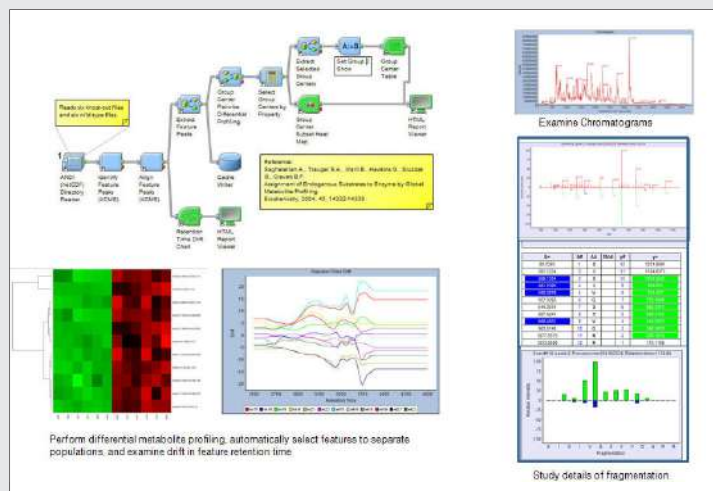


• Mass Spec for Proteomics Component Collection

Mass Spec for Proteomics 模块提供了一系列对基于生物质谱的蛋白质组数据处理工具，可以对来自质谱实验的数据进行读入、输出、显示、处理、分析、比较和发布。可以和 Sequence Analysis 和 Gene Expression 模块进行连用做其他实验数据的分析以及生物标记的识别和验证。

通过 Mass Spec for Proteomics 模块，用户可以：

- 读取多种数据格式，.RAW*、.wiff、SEQUEST DTA、ANDI (netCDF)、Mascot MGF、或 mzXML 文件
- 以多种形式查看用户数据，比如交互式的色谱图、2D 质谱运行图、扫描图、峰值图、保留时间草图和热图
- 使用 X!Tandem 进行蛋白质成分鉴定
- 使用 XCMS 提取、鉴定和比对不同图谱间的峰值
- 以 web port 交互式的方式发布分析结果
- 计算 2 个样本中蛋白丰度的 ASAPRatio 值



Pipeline Pilot™

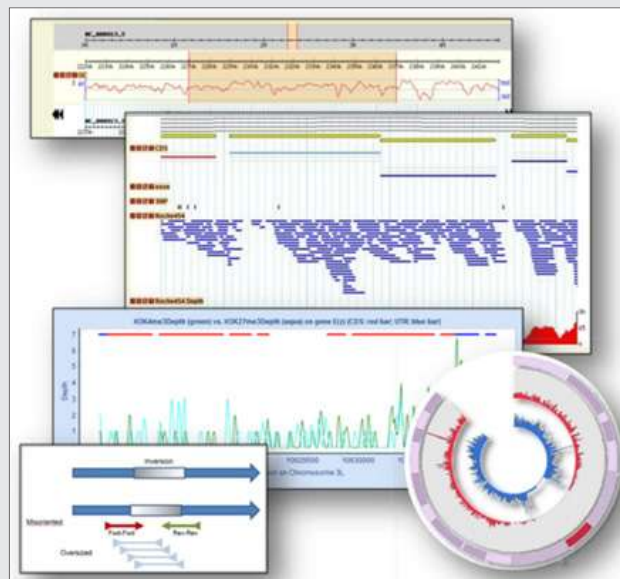
强大的信息整合和 流程定制的科学平台

• Next Generation Sequencing Component Collection

Next Generation Sequencing (NGS) 模块让用户能够分析和解析目前大多数高通量测序仪产生的大量数据集。基于强大的 Pipeline Pilot 平台，NGS 模块带来了许多专门分析 NGS 数据的组件和流程，功能涉及序列拼接、reads 定位、RNA-seq 数据分析、多态性与结构差异查找以及 reads 筛选等多个方面，还可以整合其他更多的 NGS 程序和算法，一起构建 NGS 信息化平台。

通过 Next Generation Sequencing 模块，用户可以：

- 通过简单的鼠标拖放操作创建和执行整套的新数据处理流程
- 通过定制交互式网页报告来评估分析质量和覆盖率
- 为任何知识库添加属性，比如基因和 dbSNP 入口号
- 通过 BWA、Bowtie 或 mapreads 比对 reads
- 通过 Velvet 执行 de novo 拼接
- 查找 SNP、插入、删除和拷贝数变化等信息
- 浏览用户实验的比对分析统计结果



■ Life Science Modeling & Simulation 生命科学建模和模拟 *

• Catalyst Collection (药效团)

Catalyst 模块提供了一整套药效团建模和 3D 分子数据库筛选工具。Pipeline Pilot 整合了著名的药效团软件 Catalyst，从而简化了药效团的建模和分析过程。通过该模块，用户可以轻松实现基于 Catalyst 的化合物 3D 构象的生成、创建 3D 数据库、产生药效团模型和进行虚拟筛选等功能。

• CHARMM Collection (分子模拟)

CHARMM 模块整合了著名的分子动力学引擎 CHARMM，CHARMM 的开发者 Martin Karplus 获得了 2013 年的诺贝尔化学奖。CHARMM 模块使得 Pipeline Pilot 的应用能力得到进一步拓展，使其可以用来模拟生物分子，能为蛋白质、核酸、小分子和蛋白质配体复合物进行稳定而又准确的分子力学和分子动力学模拟。

* Life Science Modeling & Simulation 的模块通过 Discovery Studio 中提供，更多模块功能和细节信息请参考 Discovery Studio 的相关资料。

Materials Modeling & Simulation 材料模拟

Materials Studio Component Collection

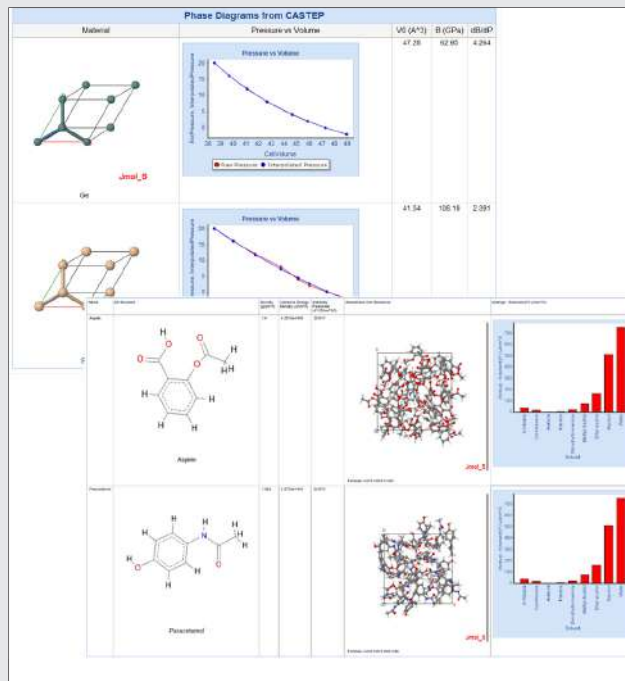
Pipeline Pilot 中的 Materials Studio (MSC) 模块, 集成并支持 Materials Studio (MS) 中主要的建模和模拟工具, 可以为材料研究的预测分析和自动化工作流程提供完整的软件解决方案。与传统的材料模拟不同的是, 通过 Pipeline Pilot, 复杂的工作流程能够被轻松迅速的开发、记录和共享, 同时在与 Pipeline Pilot 的其它 Collection 模块相结合后, 工作流能够在不同的部门内部更加深化和部署。

MSC 模块使用户可以通过 Pipeline Pilot 构建多种 MS 应用程序, 包括:

- 量化工具 (CASTEP、DMol3、VAMP)
- 分子力学模拟工具 (Forcite Plus、Amorphous Cell)
- 晶体结构分析工具 (Polymorph Predictor、Reflex、Reflex Plus)
- 聚合物性质计算工具 (Synthia)
- 分子性质建模 (QSAR)

通过 Materials Studio 模块, 用户可以:

- 采用属性计算组件来构建流程, 从而简便地计算材料分子的复杂属性
- 可以直接读入和输出 Materials Studio 支持的文件格式, 与 Materials Studio 良好互动
- 整合来自 Materials Studio 的脚本 (MS Perl Script) 应用程序
- 拥有对周期性结构可视化 and 生成报告的工具

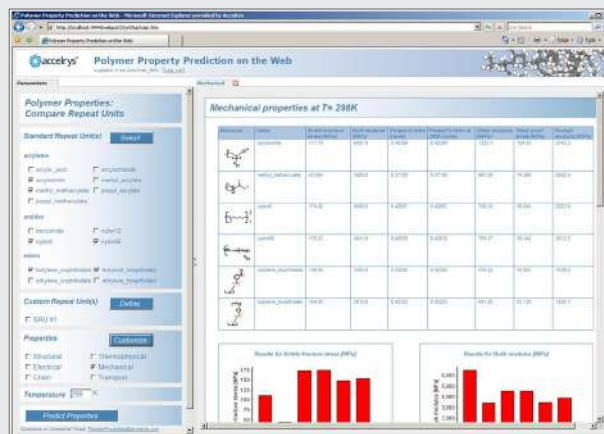


Polymer Properties Component Collection (Synthia)

高分子性质 (Synthia) 模块提供了一系列方法, 可以基于重复单元信息、分子量和温度, 对无定型聚合物和随机共聚物的性质进行快速预测。Synthia 是以美国陶氏公司 Jozef Bicerano 博士的工作为基础, 使用先进的定量结构 - 性能关系方法 (QSPR, Quantitative Structure Property Relationship) 预测聚合物性质的程序。Synthia 的预测结果在实际工作中得到了广泛的验证。

通过 Polymer Properties 模块, 用户可以:

- 预测聚合物的众多属性, 包括界面张力和交联影响等
- 建立基于网络的聚合物属性预测应用
- 可以在已有模型基础上, 添加新实验数据来更新拓展模型, 提高预测的准确性



Pipeline Pilot™

强大的信息整合和 流程定制的科学平台

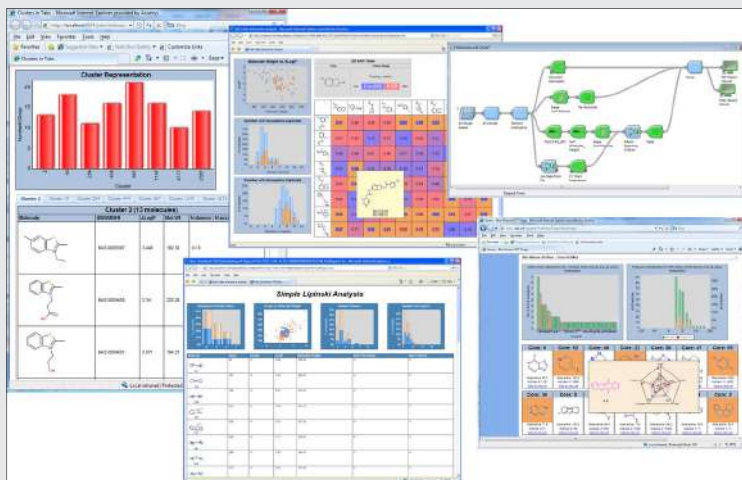
■ Reporting & Visualization 报告和可视化

• Reporting Collection 报表定制

Reporting 模块为数据分析和挖掘工作定制总结报告。报告中可灵活使用表格、图表、图像和文字等各种元素，并通过对报告布局和报告内容的调整，能更清晰地解释和阐述所获得的结果。通过在一张报告内展示多种不同的表格、图表和图像，可以在不同视角上观察同一份数据，或将不同来源的数据用不同的方式展示，从而方便对它们进行对比和分析。此外，Reporting 模块还可以用于设计 Web 网页，加入各种 Web 页面中常用的元素与用户交互，从而满足人们不同的要求。

通过 Reporting 模块，用户可以：

- 总结和显示来自多个数据源的数据
- 自动创建标准化的报告
- 创建始终处于最新状态的报告和网路应用
- 创建交互性极好的网路应用来挖掘数据的深层信息
- 在网络平台上创建合作项目组的页面
- 在文档知识库里创建文档，例如 SharePoint
- 提供对多来源数据的显示界面
- 对任何类型的数据进行定制化的可视化



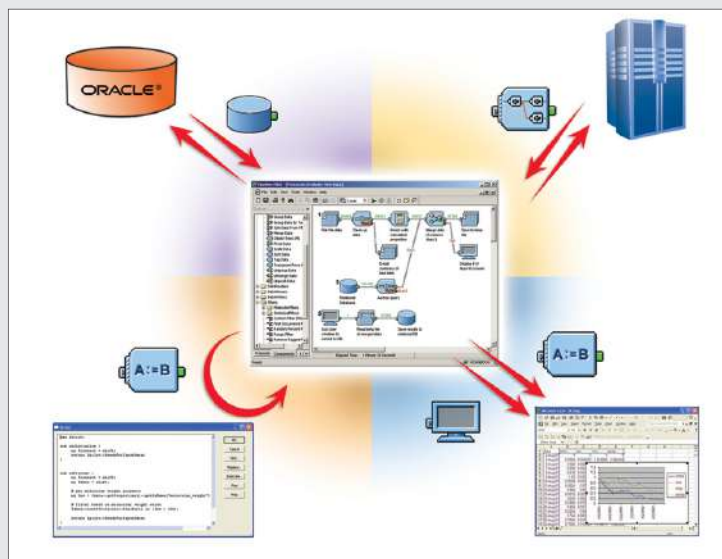
■ Database & Application Integration 数据库和应用程序整合

• Integration Component Collection

Integration 模块提供了强大的集成功能，能够把现有的内部或第三程序以及数据库无缝集成到 Pipeline Pilot 中，从而能充分利用现有的计算资源。被整合入 Pipeline Pilot 的应用工具可以像系统原有的模块组件一样被方便的使用。本模块支持对 Java、Perl、.Net、VBScript 等的脚本和程序进行直接整合，也支持通过 JDBC、ODBC 整合 Oracle、SQLServer 等数据库，实现数据的获取、分析处理和生成报告，并能将结果直接存入数据库。

通过 Integration 模块，用户可以：

- 在自己公司的数据库里直接更新数据和存储结果
- 整合已有的内部和第三方程序
- 采用标准脚本环境来快速开发新组件
- 整合交互式的桌面应用工具



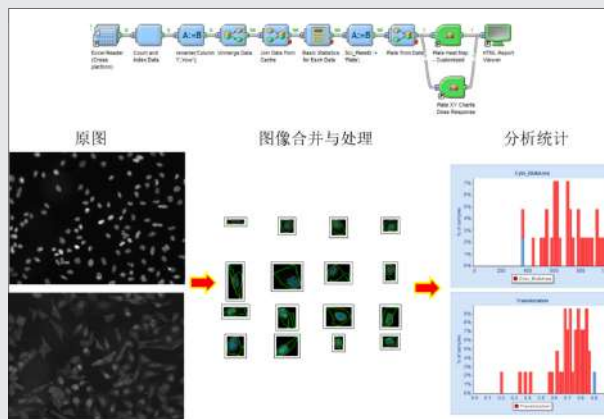
■ Imaging 图像处理

• Imaging Component Collection

图像处理模块大大改变了企业处理图像信息的方式，它不仅应用范围广泛，且简单易用。它能够对图像进行全面、准确的分析，且能够与其他相关数据关联起来，从而为研究人员和决策者提供更及时准确的决策依据。该模块提供了图像数据的增强、处理、分析、数据整合、数据管理、检索和报告输出的全面的综合科学图像分析工具，能够将图像数据与数值、文本、化学及其他科学数据整合在一个统一的工作平台进行分析。

通过 Imaging 模块，用户可以：

- 创建自动化的流程用来分割、标注和注释图像和目标
- 测量多通道图像中目标的数字和统计属性
- 对图像的像素值进行各种各样的数学操作
- 在一个开放的平台上兼容几乎所有的标准图像文件格式
- 能处理图像和非图像数据
- 扩大了 Pipeline Pilot 服务器架构的稳定性、部署和质量控制能力
- 在一个可视化和互动性良好的框架里，能很方便地处理和管理图像



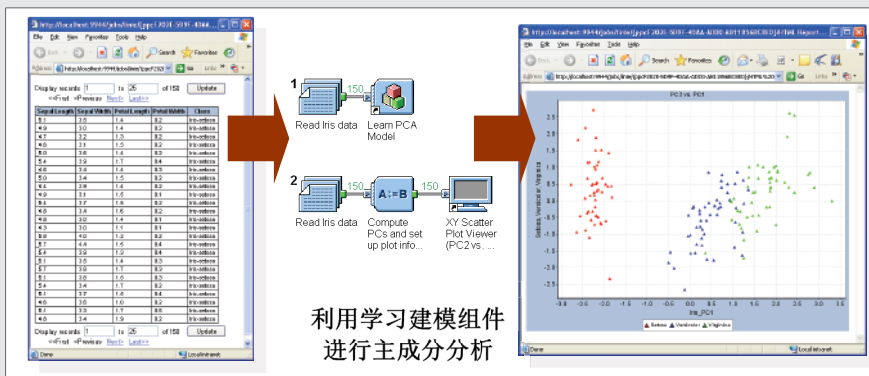
■ Analysis & Statistics 分析与统计

• Data Modeling Component Collection

Data Modeling 模块提供了一整套统计分析工具，包括机器学习和数据建模，统计模型的预测和数据的筛选，以及适用于大数据集的聚类分析工具。通过对数据进行特征分析和建立合理的具有预测能力的模型，能够帮助用户挖掘出隐藏在数据中的潜在规律，为企业和研究人员提供更好的决策。利用该模块的组件，用户可以很容易地对数据进行聚类分析、构建贝叶斯模型、递归分割模型、主成份分析和偏最小二乘回归等分析。结合 Chemistry 模块，还能够基于化学结构进行建模和聚类分析，构建出各种结构-活性关系模型的、进行化合物分类以及自动寻找化合物的最大公共子结构等。

在 Data Modeling 模块里，用户可以通过以下方法对数据进行分析和建模：

- 相似性和聚类分析方法
- Categorical learning
- 主成分分析 (PCA)
- Linear regression、partial least squares (PLS) regression 和 k-nearest neighbor (kNN) regression
- ROC 图、富集分析图和其它用于评估建模质量的可视化技术



Pipeline Pilot™

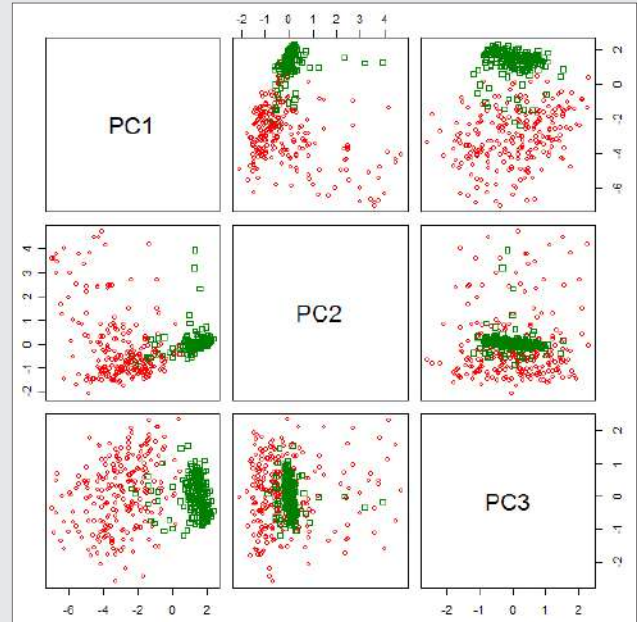
强大的信息整合和流程定制的科学平台

• R Statistics Component Collection

R 语言是目前最主流的统计分析编程语言之一，其具有丰富且强大的数据处理、统计分析、绘图功能。Pipeline Pilot 的 R 统计模块将 R 的一些数据统计和分析方法被无缝整合到 Pipeline Pilot 中，用户无需编写 R 脚本语言，仅需通过调用相关组件，就能够进行基于 R 语言的数据分析和图表生成。R 统计模块能够实现数据处理、聚类分析、机器学习和建模，以及一些传统和探索性的数据分析方法，如偏最小二乘法，神经网络和支持向量机，方差统计，因子分析等。该模块还提供 R 语言的结果，用户可以通过编写 R 脚本语言来实现新的数据分析功能，并扩展 R 模块中的组件功能。

通过 R Statistics 模块，用户可以：

- 在一个热图里显示多个属性的相关性，发现最相关者
- 使用箱线图观察数据在各个子集的分佈情况
- 执行方差分析以决定多个数据集平均值的差异性
- 通过多种方法进行数据建模，比如支持向量机和人工神经网络
- 应用自定义的模型对新数据集进行预测
- 允许随着实验数据的不断获得来扩展模型
- 采用多种不同的聚类方法
- 允许用户对数据应用自有的 R 脚本



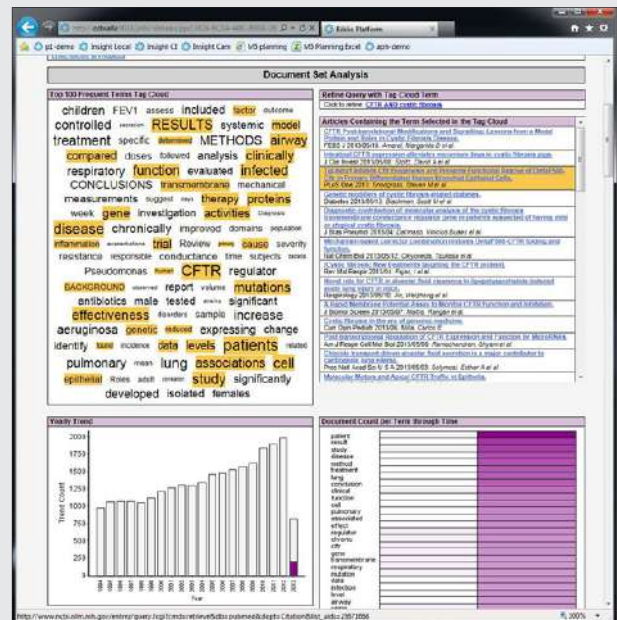
■ Document Search & Analysis 文档挖掘和分析

• Text Analytics Component Collection

Text Analytics 模块提供了很强的文献检索和文本数据挖掘的功能，将文献检索、特征数据提取和表征等步骤整合为一个自动化的流程，可以实现最具挑战性的文本数据挖掘过程。通过整合文献检索和文献描述到自己的科研流程中，达到交互或自动完成包括内部（本地）和外部（互联网）文本数据资料挖掘的目的。

通过 Text Analytics 模块，用户可以：

- 搜索 PubMed、美国和欧洲专利、雅虎新闻、本地文件，还可以整合第三方的搜索引擎
- 使用一条查询语句去搜索多个数据源
- 在本地文档和在文献中提取关键概念或查找关联性
- 根据内部和在线文献，扩展日常生物和化学任务的注释信息和相关性
- 发现新兴的趋势和关联性，分析用户领域的专利情况
- 创建丰富的、交互式的报告来挖掘用户文档中的深层信息

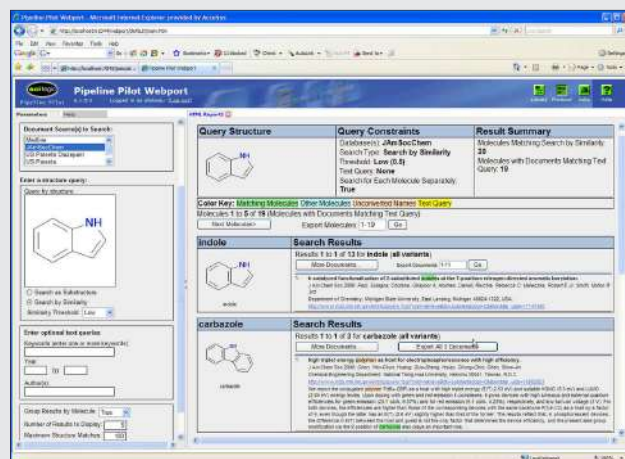


• ChemMining Component Collection (Chemical Text Mining)

ChemMining 模块使 Pipeline Pilot 具备“化学智能”的文本数据分析功能,它能够自动识别出文献中的化学词汇如化学名、药名、IUPAC 名、CAS 号、分子式等,并能够将化学名自动转换为化合物结构。同时,它还能够把识别到的化学结构和文献存储到专门的数据库,用户可以通过化学结构检索或者文本检索相关文献,从而实现更具针对性的,更智能的文献检索和管理。

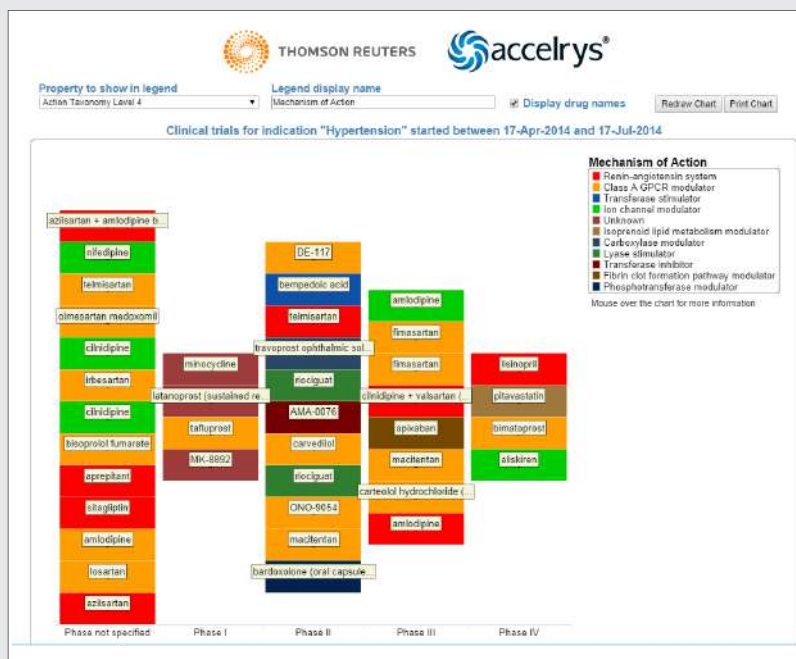
通过 ChemMining 模块,用户可以:

- 根据在摘要、专利和内部文档中发现的结构来创建可搜索的数据库,用户可以在其中进行智能查询
- 搜索 PubMed 的摘要和专利来发现化学结构
- 从内部文档和其它文件中查找化学制品名称,包括自定义和公共的名称
- 创建包含了已发现的结构和感兴趣的术语的报告
- 建立整合了文献和化学分析的灵活分析流程



• Thomson Reuters

通过 Thomson Reuters (汤森路透) Cortellis™ 模块,可以让您使用 Pipeline Pilot 流程,以更高效简单的方式来访问 Cortellis 数据库所提供的海量的、广泛的生命科学数据,并且能够将内部数据与之无缝整合。该模块是 Biovia 与汤森路透社之间持续合作的重要组成部分,可以访问的 Cortellis 数据库包括药物靶点、化合物、公司、临床试验及结果,专利等。



Pipeline Pilot™

强大的信息整合和 流程定制的科学平台

■ Laboratory 实验室数据分析

• Plate Data Analytics Collection

Plate Data Analytics 模块提供了对孔板数据的读、写、报告、可视化、操作和计算等功能，数据处理流程中的每一条数据记录都能够对孔板上的单孔数据或整版数据进行处理。利用 Pipeline Pilot 的图形化数据处理流程设计能力，无需编写代码就能实现如分析筛选结果这样的复杂的数据分析过程，结合 Integration 模块，还可实现孔板数据在数据库中的存储和读取功能。

通过 Plate Data Analytics 模块，用户可以：

- 定义了分析各种筛选试验结果的自动化程序
- 在孔板内部和孔板之间执行统计计算，比如自动计算 IC50 曲线
- 用 Pipeline Pilot 自身的工具、R 统计或 Graph Pad Prism 方法进行剂量反应计算
- 可视化和数据挖掘孔板数据信息
- 通过与 Integration 模块结合，直接从数据库插入或获得孔板数据



• Analytical Instrumentation Collection

Analytical Instrumentation 模块提供对实验室中的分析谱图数据的获取和处理。使用该模块，可以对分析谱图数据自动处理、产生报告，并且与其他部门共享。常用的谱图处理功能包括：谱图峰识别、峰面积积分、背景去除、平滑、傅里叶转换等。而专门的 NMR 数据处理模块，可以实现 FID 到化学位移的一步自动转化，并可以从化学结构产生（同时支持 C13 和质子位移）NMR 谱图。可以读入的谱图格式包括 JCAMP-DX、SPC、RheoML、AnIML，也支持 Bruker 和 Varian 的 NMR 格式。

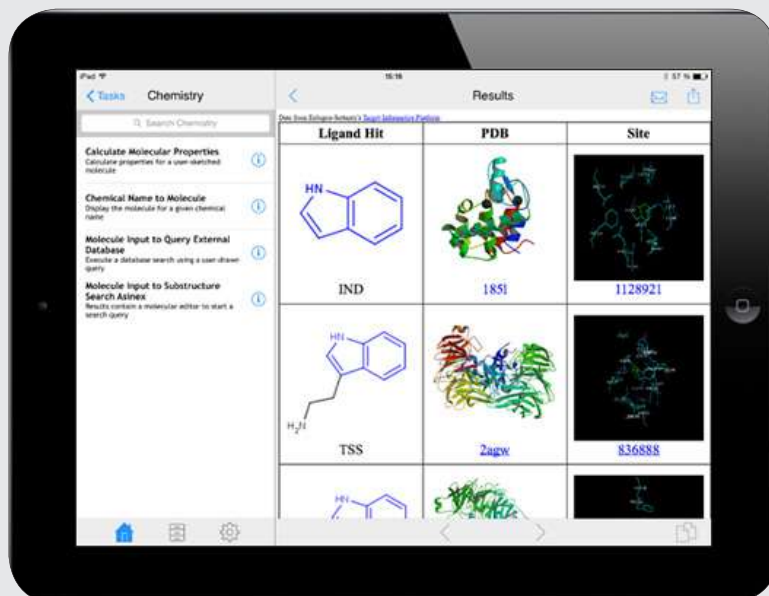
通过 Analytical Instrumentation 模块，用户可以：

- 更有效地在部门内部分享谱图分析结果
- 通过固定重复的分析流程，整合分析算法和分享最佳的数据分析实践来提高分析能力
- 创建能满足用户不断变化需求的报告，从而降低谱图分析的费用
- 整合孤立的数据，建立相互协作的研究环境



• Mobile 移动应用

移动应用模块使 Pipeline Pilot 具备了通过移动终端即时访问数据和信息的能力。通过 ScienceCloud Task App，终端用户可以通过 iPhone 或者 iPad 以简单易懂的方式来调用功能强大的 Pipeline Pilot protocols。用户可以将已有的 Protocols，几乎不用做修改，就能够部署到 App 中，使其成为新的“任务”（tasks）。同时，移动应用模块能够充分利用触摸屏设备的优势，并提供了相应的组件来处理由移动终端传递过来的音频、GPS 定位、照片等数据信息。移动模块还提供的基于 Html5（已做移动优化）的图表工具，从而给终端用户带来更好的使用体验。



Pipeline Pilot™

强大的信息整合和 流程定制的科学平台

系统要求

■ 客户端系统要求

系统	操作系统和软件	最低硬件配置
Windows	<ul style="list-style-type: none"> • Windows 7、Windows XP SP3 或 Windows Vista SP2 • 需要 Windows 的“Power User”级别用户 (需要在客户端机器上注册软件信息) • Internet Explorer 6、7、8 或 Mozilla Firefox 2.x or 3.x • Microsoft Office 2007、2003、XP or 2000 (可选) 	<ul style="list-style-type: none"> • 奔腾级别的处理器 (1 GHz 或更快) • 512 MB 内存 • 150 MB 硬盘空间

■ 服务器端系统要求

系统	操作系统和软件	最低硬件配置
Windows	<ul style="list-style-type: none"> • Windows Server 2008 SP2、Windows Server 2008 R2 (64-bit)、Windows Server 2003 SP2 (32-bit) • 系统管理员帐号 • Internet Explorer 6、7、8 或 Mozilla Firefox 3.5 或 3.6 	<ul style="list-style-type: none"> • Intel 奔腾或至强 (Xeon) 级别的处理器 2.0 GHz (双核) 或 3.0 GHz。 一般来说, 一颗 CPU (核) 可同时运行 2 个作业, 可保证正常的运行速度。 • 2-4GB 内存 (对于 64bit Linux 系统, 需要 3-6GB 内存), 每个 CPU 至少需要 1 GB 内存。 • 软件安装需要 3-5 GB 硬盘空间。软件运行、用户数据需要 30GB 硬盘空间。
Linux	<ul style="list-style-type: none"> • Red Hat Enterprise Linux 4 WS 或 ES (升级到 5 或更高版本), 32-bit x86 和 64-bit • Red Hat Enterprise Linux 5 WS 或 ES, 64-bit • SuSE 10 64-bit (支持 32 位) • SuSE 11 64-bit (R 统计组件不支持图形化输出) • 需要 root 帐号 • Mozilla Firefox 3.5 或 3.6 	

■ 格点系统 (Cluster/Grid System) 软件要求

用户可以在格点系统上部署 Pipeline Pilot , 当前支持的作业管理软件包括 :

格点系统 作业管理软件	<ul style="list-style-type: none">• Sun Grid Engine (SGE) 6.1• PBS Professional 9.2• Platform LSF 7.00 Update 3
----------------	---



创腾科技有限公司成立于2000年初，是业界领先的面向生命科学和材料科学领域提供综合研发、过程、检测、生产信息化平台及咨询服务的高新技术企业。创腾科技的业务总部设在北京，在上海设有研发中心。

作为一家业界领先的信息技术公司，创腾科技可以为生命科学和材料科学的用户，提供包括计算模拟与数据建模、科技创新信息化两大平台的综合解决方案和服务，提升这些用户在研发、过程、检测、生产等领域的管理、决策和创新能力。创腾科技与国际上在业界领先的著名信息技术公司拥有长期而紧密的合作关系，并拥有业界最先进的开发平台和较强的研发能力，能够为中国的企业和科研机构提供当前世界上最先进的信息技术解决方案和服务。

作为一家高科技企业，创腾科技拥有一支具有专业背景和IT背景的高素质复合型人才队伍，以及经验丰富的管理团队，能够保证向用户提供一流的技术支持和服务。现在，在中国已有超过500家的单位得到了创腾科技所提供的产品和服务，其中包括国内最大的制药企业，最大的新药研发外包企业、最大的石化企业以及主要的高校和科研单位等。

为客户提供持续一流的产品和服务，不断为客户创造价值，是创腾科技努力的目标。创腾科技愿意为提高中国科研机构和企业的创新能力与核心竞争能力，做出自己的贡献。

创腾科技有限公司

NeoTrident Technology LTD.

北京办公室

北京市中关村科学院南路2号
融科资讯中心C座南楼1512室
邮政编码：100190
电话：+86 10 82676188
传真：+86 10 82677178

上海办公室

上海市浦东新区张江达尔文路88号
半岛科技园11栋4楼
邮政编码：201203
电话：+86 21 51821768
传真：+86 21 51821758

苏州办公室

江苏省苏州市工业园区东长路88号
2.5产业园A2栋301室
邮政编码：215028
电话：+86 512 67509707
传真：+86 512 67509705

广州办公室

广州市天河区黄埔大道西33号
三新大厦16-E房
邮政编码：510620
电话：+86 020-88527961

创腾官网：www.neotrident.com

创腾学院：training.neotrident.com



官方微信