

智能创新带动卓越品质 · 基于云计算和人工智能



科学数据基因组

SCIENTIFIC DATA
HIERARCHY

数据 · 快人一步

INTRODUCTION

▶ 产品简介

通过科学数据基因组平台 (SDH)，企业组织可以对来自不同数据源的海量科学数据进行抽取、整合、清洗、索引，建立统一数据规范，灵活构建和管理主题库，并按需抽提所需数据。赋能业务人员的数据使用能力，加速企业的智能化创新。

▶ 传统方式面临的挑战

1. 医药、材料及化工企业传统数据采集与管理方式：基于不同历史时期、为满足不同业务目的而单独建设专项专业科研数据库：

- ▶ 管理种类繁多的专业科研数据库及数据中心对IT依赖大，但既懂业务又懂数据的人才匮乏；
- ▶ 数据标准与数据质量普遍存在问题；
- ▶ 各个专业科研数据库单独建设与维护，造成企业资源和成本大量浪费；

2. 针对业务领域数据的整合与提取存在困难

- ▶ 研发层面：实验数据、结构数据与毒理/活性等分析测试数据分散在各业务环节，无法集中统一视图呈现；
- ▶ 工艺层面：过程工艺参数与质量相关数据种类繁多，缺少过程监控与质量研究的有效工具；

3. 数据的分析与赋能门槛较高

- ▶ 关键工艺参数对于CMC阶段或材料研发项目转产十分重要，但寻找关键工艺参数工作量大且繁琐；
- ▶ AI建模与模拟仿真对于数据质量要求极高，赋能困难；



▲ BI报表-质量工具

SCIENTIFIC DATA HIERARCHY SDH

生命科学 管理



NEO 材料科学 DATABASE 科研数据库

cod x 通过元素、化学式、ID等搜索

元素 化学式

至少包含元素 仅包含元素

H																	He
Li	Be	¹⁴ Si								B	C	N	O	F	Ne		
Na	Mg	Si Silicon								Al	Si	P	S	Cl	Ar		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Lr	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og
		La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb		
		Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No		

▲ NEO科研数据库

SCIENTIFIC DATA HIERARCHY

SDH

Scientific Data Hierarchy

数据源 + 锂电池容量-转换规则

王五_锂电池容量.xlsx
李四_锂电池容量.xlsx
张三_锂电池容量.xlsx

Sheet1

remaining SOH	groupnumber	sample number	cycling current (C-rate)	1st cycle Coulombic efficiency	1st-to-2nd cycle capacity change / initial capacity	2nd-to-10th cycle capacity change / initial capacity	electrode thickness (m)	material loading (mg/cm ²)	ratio of BP in active material	surface roughness (m)	particle size (m)
0.165632754	2	7	0.5	0.82244898	0.280397022	0.017369727	83.2	5.13	0.5	2.224	3.54
0.157310705	3	17	0.75	0.652192422	-0.010443864	-0.360966057	77.5	4.28	0.5	1.646	1.76
0.151648352	7	42	1	0.755813953	0.045054945	-0.7	65.7	3.76	0.45	1.734	2.53
0.151162791	8	47	1	0.169876543	-0.305232558	-0.441860465	61.9	3.42	0.7	2.154	3.2
0.150266972	2	8	1	1.065853659	-0.131960336	-0.150266972	83.2	5.13	0.5	2.224	3.54
0.147783251	6	37	1	0.153904473	-0.413793103	-0.157635468	69.7	3.51	0.45	1.22	1.61
0.13893312	3	15	0.5	0.864121087	0.281449045	-0.112659236	77.5	4.28	0.5	1.646	1.76
0.1375	8	48	1	0.184650894	-0.33125	-0.41875	61.9	3.42	0.7	2.154	3.2
0.128509719	8	45	0.5	0.856614246	0.224982001	-0.10475162	61.9	3.42	0.7	2.154	3.2
0.125858124	15	83	0.25	0.553164557	-0.254004577	-0.409610984	37.7	2.34	0.7	0.583	1.34
0.124743677	3	19	0.5	0.875785693	0.287423103	-0.215994532	77.5	4.28	0.5	1.646	1.76
0.123745819	13	74	1	0.173032407	-0.334448161	-0.424749164	51.3	2.61	0.5	1.036	2.73
0.114782609	15	84	0.2	0.635710337	-0.037391304	-0.56173913	37.7	2.34	0.7	0.583	1.34
0.107692308	1	5	0.5	0.010310914	-0.723076923	-0.107692308	100.6	5.57	0.55	2.64	1.11
0.106060606	6	36	1	0.090349076	-0.121212121	-0.333333333	69.7	3.51	0.45	1.22	1.61
0.100318471	10	58	0.1	0.144600507	-0.275477707	-0.382165605	55.2	3.55	0.4	2.086	5.87
0.086956522	1	6	0.5	0.007529874	-0.652173913	-0.152173913	100.6	5.57	0.55	2.64	1.11
0.07815924	15	82	0.5	0.769533446	0.258582907	-0.889700511	37.7	2.34	0.7	0.583	1.34
0.064923354	5	29	0.5	0.165745031	-0.11091073	-0.503155996	71.6	3.62	0.6	1.88	2.48
0.063410454	5	31	0.5	0.167600172	-0.101113967	-0.513281919	71.6	3.62	0.6	1.88	2.48
0.063327032	4	21	0.2	0.409246126	-0.289224953	-0.52173913	72.5	4.55	0.6	1.215	5.25
0.060909091	10	59	0.15	0.178919974	-0.072727273	-0.500909091	55.2	3.55	0.4	2.086	5.87

数据转换编辑

▶ 核心功能

业务数据融合

工具化数据清洗

科学数据溯源

数据基因组构建与管理

多数据源连接

自定义数据模型

按需抽取数据

FEATURES

▶ 功能特点

1. 星链多数据源

汇聚企业内部不同业务系统与数据应用, 兼顾外源各类专业数据库, 科学数据基因组支持多种异构数据源。

多数据格式支持

- ▣ 关系型数据库, 包括Oracle, MySQL, SQL Server等;
- ▣ NoSQL存储, 包括MongoDB, Redis等;
- ▣ Excel文件等;

科学数据支持

- ▣ 分子结构、生物序列、图谱等;

应用数据支持

- ▣ 业务系统, 包括ELN, LIMS, LES, CIMS, CMS, PM等;

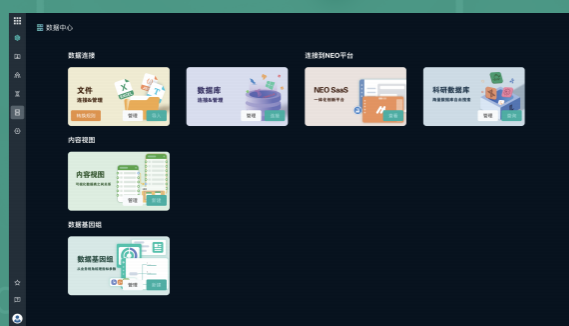
外源数据支持

- ▣ CNPD/MNPD/TCMD数据库等;
- ▣ 小分子药物活性、靶点等相关数据库(千万级)。

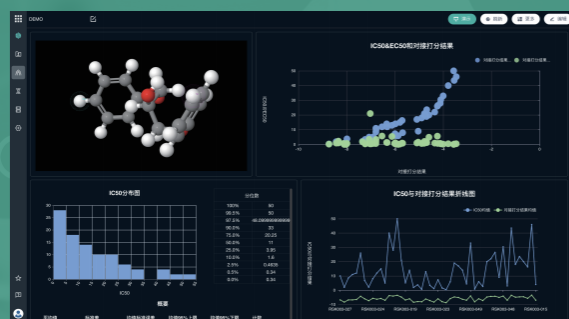
2. 数据赋能

按需抽取数据, SDH输出高质量数据, 助力企业实现数据智能:

- > 自助式数据抽取, 便携搭建BI报表;
- > 引入质量工具, 为用户过程控制和过程能力研究提供图形化分析工具;
- > 与分子模拟与AI智能平台对接, 包括MaXFlow等;
- > 科学数据处理: 结构式绘制与检索, 描述符计算。



▲ 数据中心



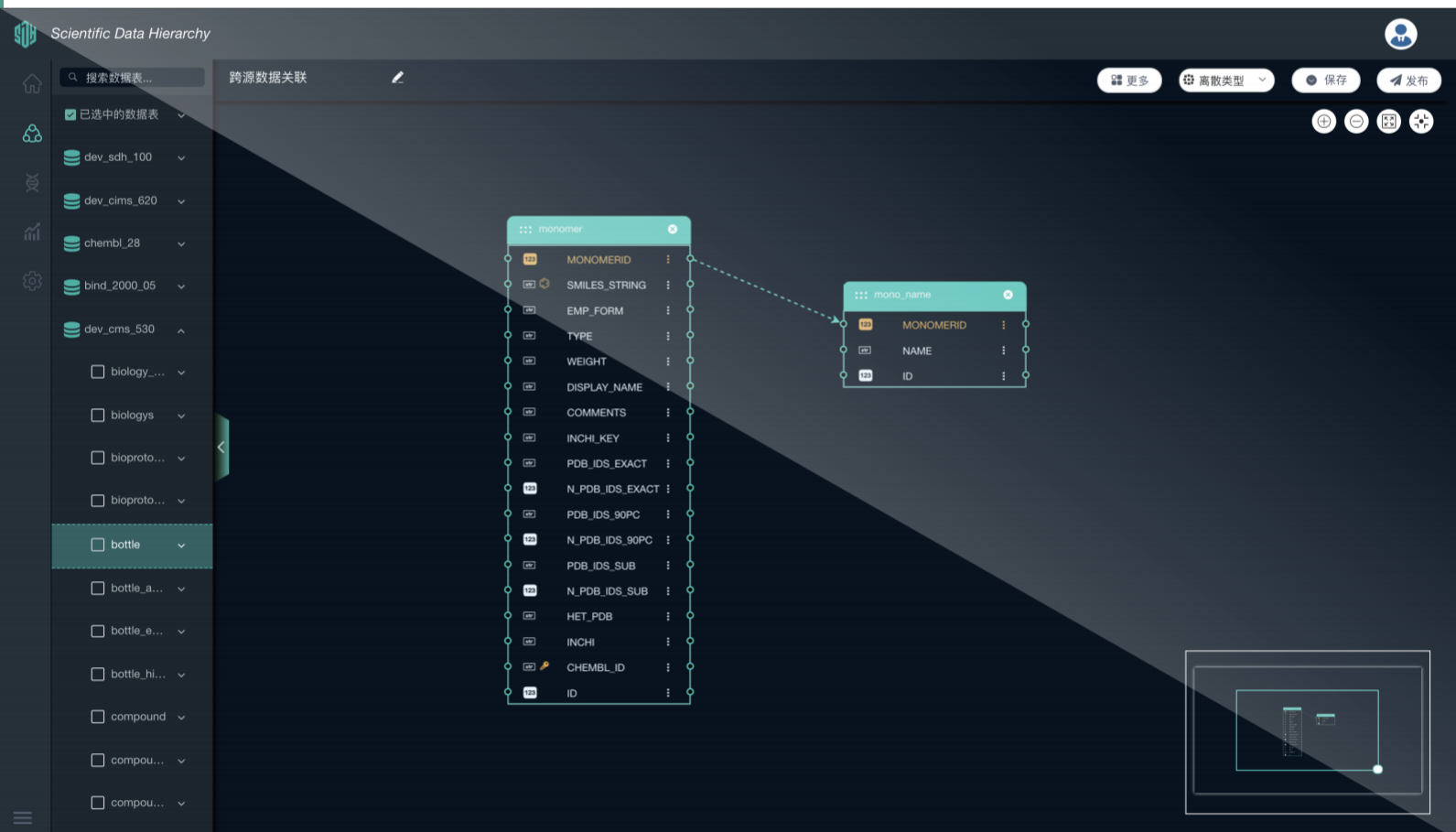
▲ BI报表-图表

SCIENTIFIC DATA HIERARCHY SDH

3.内容视图

内容视图是将来自不同数据源中的数据表之间建立关联关系,根据业务需求,可视化构建与管理数据模型,为数据基因组的搭建作数据准备。

- > 支持跨源数据关联;
- > 自定义数据模型与管理。
- > 锚点连接与拖拽操作,轻松构建数据表关联关系;



▲ 内容视图详情页

SCIENTIFIC DATA HIERARCHY

SDH

FEATURES

► 功能特点

4. 数据基因组

为了破译科研领域中全部科研业务数据信息，将散落在研发进程中的科学数据进行重组、融合，建立数据溯源关系，为企业数据智能积累数据资产。

数据基因组是从业务视角语境化构建数据模型，为研发智能打造业务主题数据库。

- ◆ 树状思维导图与旭日图 - 双视图，从不同视角管理数据基因组；
- ◆ 灵活建立数据模型与跨源字段的绑定关系；
- ◆ 数据基因溯源，实现数据基因全生命周期管理；
- ◆ 自定义可视化数据基因组构建与管理。



Scientific Data Hierarchy

汽油配方数据基因组

切换为旭日图 保存 发布

节点信息设置

节点名称
理化信息 4/50

关注度
1 2 3 4 5 6

描述信息
理化信息是汽油配方研究关键指标参数 0/100

▲ 数据基因组

VALUE & BENEFITS

▶ SDH带来的价值与收益

科学数据基因组 (SDH) 为企业创造数据资产, 以数据价值赋能企业客户;
建立企业数据规范, 打造AI模型和数据资产, 赋能企业新价值。

科学数据基因组 (SDH) 为生命科学和材料领域带来数据接入-数据整合-数据分析/赋能的一体化解决方案, 消除科研数据壁垒, 打造数据可视化, 助力企业聚焦关键参数, 缩短开发与工艺探索时间。



● 数据获取时间



● 数据采集与整理时间



● CMC阶段或材料研发
转产数据整理与优化
时间



● 研发过程中的
知识转移时间



● 质量监控中异常问题
溯源时间



● 基础IT设施成本

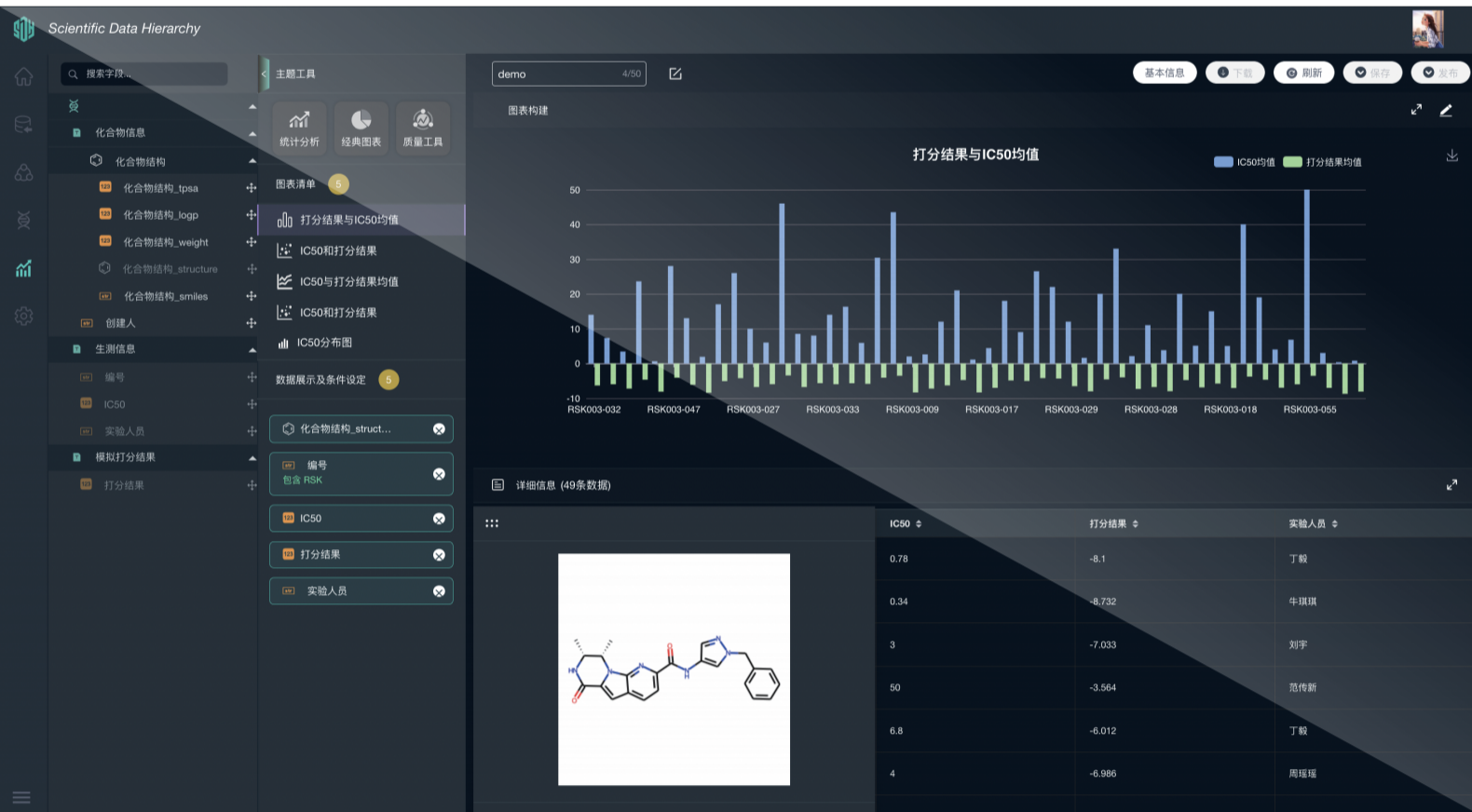


⊕ 加速数字化转型



⊕ AI建模效率

SCIENTIFIC DATA HIERARCHY SDH



▲ BI报表-结构

关于创腾科技

创腾科技专注医药和材料科学研究领域20年,通过对云计算,移动互联和科学人工智能的技术的创新应用,帮助企业和创新科研机构快速进行研发的数字化转型,实现智能创新。

创腾科技拥有业界领先的从数字化研发到智能创新的整体移动互联解决方案,包括iLabPower数字化研发平台(涵盖研发与质量检测),SDH科学数据基因组平台和MaXFlow科学人工智能创新平台。

创腾科技拥有超过千家用户,正在服务数万名中国的科技精英,国内顶级的科研机构,名列前茅的制药企业,CRO/CDMO和石油化工企业,以及新药创新的明星企业都在利用创腾科技的数字化研发及智能创新平台,实现其智能创新的核心战略。



创腾科技 | 苏州·上海·北京·广州·成都



江苏省苏州市工业园区东长路88号2.5产业园A2栋301室
电话: 0512-67509707

广东省广州市天河区黄埔大道西33号三新大厦16-E房
电话: 020-88527961

北京市海淀区花园北路25号, epark (花园路社区) 2号楼3门
2层O-2-261
电话: 010-82676188

创腾官网: www.neotrident.com
创腾邮箱: market@neotrident.com