
诺崴科技
NUOWEI TEC
隐私计算
赋能医疗健康大数据价值流转
演讲人：诺崴科技合伙人 医疗大健康事业部总经理 吴凡
2023年2月

数据成为生产要素和战略资源，价值亟待释放

诺崴科技 NUOWEI TECH

国家高度重视数据要素及其市场化配置改革，不断完善中国特色数据基础制度体系，促进全体人民共享数字经济发展红利；激活数据潜能、释放数据价值成为数字经济深化发展的主要引擎

法律要求

《中华人民共和国数据安全法》（2021年9月）

提出国家将对数据实行分级分类保护、开展数据活动须履行数据安全保护义务承担社会责任等。

《个人信息保护法》（2021年11月）

保护个人信息权益，规范个人信息处理活动，保障个人信息依法有序自由流动，促进个人信息合理合规使用。

政策驱动

《关于构建数据基础制度更好发挥数据要素作用的意见》（2022年12月）

从数据产权/流通交易/收益分配/安全治理四方面构建数据基础制度体系，充分发挥数据要素作用，赋能实体经济，推动高质量发展。

《建设全国统一大市场》（2022年4月）

加快培育数据要素市场，建立健全数据安全、权利保护、交易流通、开放共享、安全认证等基础制度和标准规范。

“十四五”数字经济发展规划（2022年1月）

把“数字技术与实体经济深度融合”作为我国数字经济发展重要主线。

市场需求

企业进入数据要素时代，数据流通、共享与协同成为数字经济时代企业的刚性需求。企业自身的数据是单一有限的，越来越多的业务场景需要多方数据共享。因此，企业需要进行多方数据共享，以释放数据的应用价值。

目前医疗、金融、通信、政务、互联网、能源等多个领域对隐私计算需求大。据相关数据显示，金融、通信、政务、医疗、互联网、能源领域隐私计算需求分别占比53%、17%、13%、9%、5%、3%。2021年我国隐私计算市场规模达8.6

隐私计算技术赋能数据要素安全流通

诺崴科技 NUOWEI TECH

数据要素安全流通面临的挑战

数据泄露
数据易复制和传播，在流通过程中极易带来数据泄露和数据安全的风险

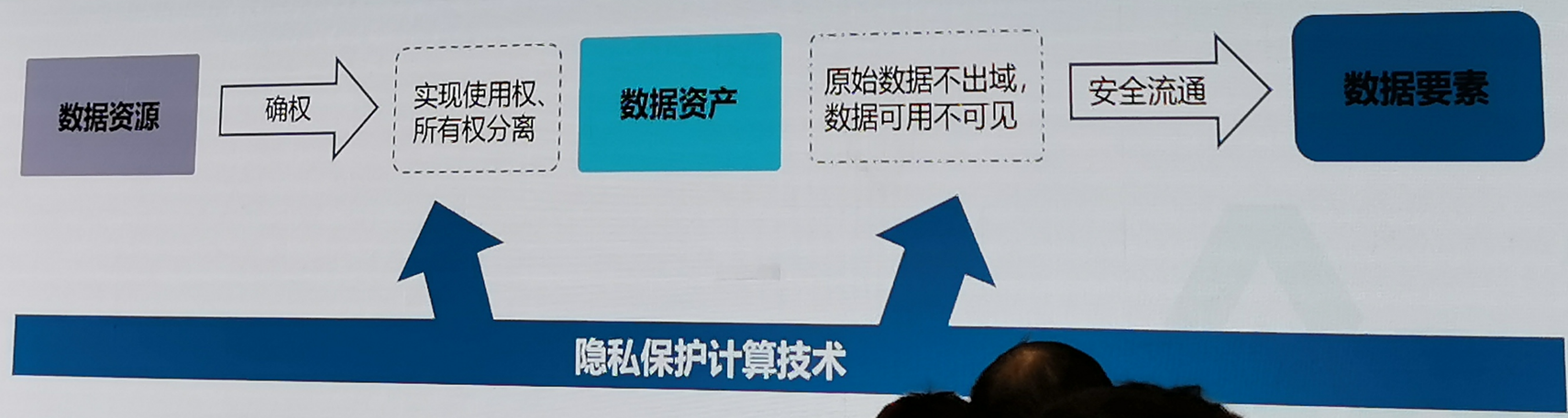
收益分配
权属难确立，重要性难区分，收入难分配

参与意愿
数据所有方参与交易的意愿和积极性问题

权属确立
交易过程的信息不对称，数据的所有权、使用权难以保障

融合难度
数据类型复杂多样等，带来的融合难度

监管难点
一些违规行为“黑箱”化，带来数据交易监管“盲区”



医疗数据安全共享三种模式

诺崴科技 NUOWEI TECH

传统的多中心合作模式

× **需要数据硬拷贝移动到授信第三方**，由于不同机构间的隐私保护政策的不同，给数据分享带来合规性挑战

× 脱敏后数据依然有隐私泄露风险

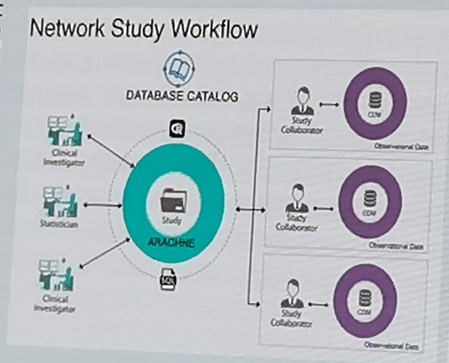
× 数据使用权、管理权、所有权**无法有效分离**

V1.0 (脱敏)

×在数据网络内分发计算任务，但每个数据源独立完成计算，返回独立模型结果，通过meta analysis汇总

×**汇总后的模型精度不高**

×**多中心数据利用有限，不能有效联合数据分析（横向/纵向）**



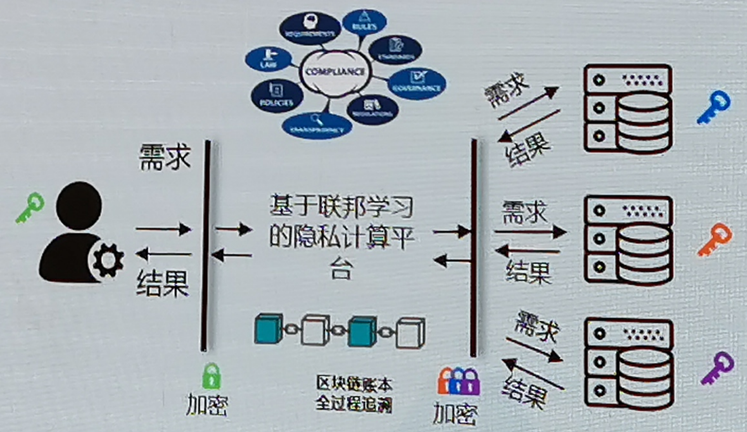
V2.0 （沙箱）

隐私计算模式

✓**通过隐私计算技术（结合联邦学习，可信计算环境、多方安全计算等技术）来保护数据检索、数据建模、模型推理等数据应用**，计算结果精度可保证、数据使用过程可追溯

✓保护各方用户隐私、打破数据孤岛、增加样本量或丰富数据的维度

✓满足数据使用合规性、 实现数据有效利用



V3.0 （隐私计算）

隐私计算技术介绍

诺崴科技 NUOWEI TECH

隐私保护计算(Privacy-Preserving Computing)

简称“隐私计算”指在保护数据本身不对外泄露的前提下实现数据分析计算的一类信息技术。

原始数据不出域，数据可用不可见，数据不动模型动，用途可控可计量。

- 联邦学习
- 多方安全计算
- 同态加密
- 可信计算环境
- 差分隐私
- 区块链
- 零知识证明
- 混淆电路
- 秘密分享
- 不经意传输

联邦学习 — **依赖机器学习**

本质是一种加密的机器学习技术，通过多中心之间交换统计信息，而不是个体信息的方式，实现数据在“可用不可见”的前提下的联合计算

可信执行环境 — **依赖可信硬件**

其核心思想是构建一个硬件安全区域，在处理器中形成飞地，数据仅在该安全区域内进行计算

多方安全计算/同态加密 — **依赖密码学**

基于密码学的技术，可以在密文下完成一定程度的计算，并且保证在密文下的计算结果等价于明文下的计算结果

隐私计算是数据安全共享的技术“最优解”

诺崴科技 NUOWEI TECH

隐私计算被认为是数据安全共享的技术“最优解”，
帮助数据安全存储、可信流转、协调生产，实现数据价值安全共享，释放数据价值潜力

安全存储 → 可信流转 → 协同生产

个人数据隐私、最小采集
用户自主管理数据、安全存储、选择披露、享受机构提供的数据

医疗机构数据隐私、最大协同
可信传输、安全多方计算、同态加密、零知识证明、联邦学习等

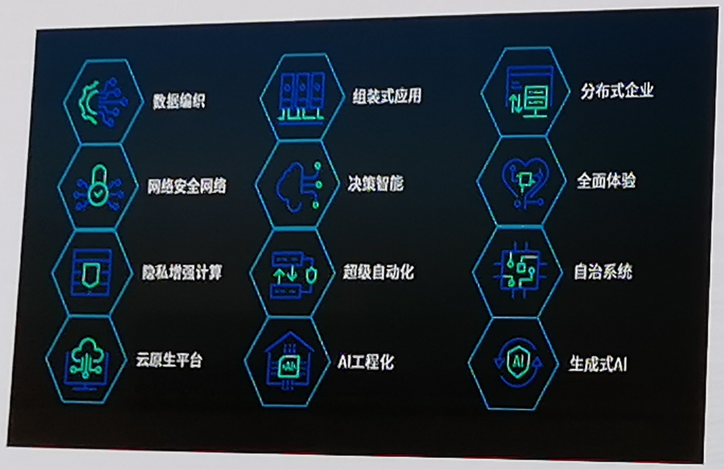
三方数据隐私、最大开放
不敢开放、不能开放、不愿开放的数据，可以对外开放利用

个人 ⟷ 用户授权数据 ⟷ 医疗机构 ⟷ 开放数据 ⟷ 三方

隐私计算是未来的技术趋势

中国数字经济发展高速增长，隐私计算是数字经济底层基础设施
未来只要涉及数据的行业，都可能使用隐私计算，3年后隐私计算市场规模将达近200亿人民币

隐私计算：数字经济新基建



- 隐私计算连续三年入选Gartner重要战略技术趋势。
- Gartner预计，到2025年，60%的大型企业机构将使用一种或多种隐私增强计算技术。

- IDC Directions：2021年中国隐私计算市场规模已突破8.6亿元，未来有望实现110%以上的市场增速。
- 信安世纪：隐私计算产品市场规模约为十亿元，基于隐私计算的数据交易应用模式市场或将达到千亿级。
- 艾瑞咨询：到2025年，中国隐私计算市场规模将达到145.1亿元。
- 毕马威：未来市场规模将快速发展，三年后技术服务营收将达到100-200亿人民币，甚至将撬动千亿级的数据平台运营收入空间。
- 隐私计算研究院：中国隐私计算行业将迎来快速增长，预计至2026年市场规模将达184亿元，年复合增长率103.3%。

隐私计算典型应用场景

诺崴科技 NUOWEI TECH

医疗

多中心专病库
带隐私保护的临床医疗数据专病库，为辅助诊断、新药物研发提供合规数据服务

新药研发
多机构罕见病联合分析，真实世界研究，新药临床研究等

科研服务
跨机构数据联合分析，满足医疗科研对数据的要求，提升医疗服务水平

政企

普惠金融
联合多维企业数据，全面构建企业画像，服务小微企业金融等方面

数据共享
打通各委办局，有效实现政务数据一张网，智慧城市等

安全服务
AI人脸识别，治安巡逻一体防控，边防影像监控等

金融

风控评分
结合政务里企业多维度数据，服务中小微企业信贷风险评估、完善银行风控模型

客户画像
联合画像和产品推荐，在多头借贷等场景下有效降低违约风险

精准营销
联合建模、预测等，构建精准的营销模型

锘崴科技
NUOWEI TECH

隐私计算在医疗领域应用的现状

隐私计算技术在保护医疗数据安全的前提下实现了合规流通和价值，
医疗健康已占据隐私计算的10%+，医疗领域的隐私计算产品已能支持较大规模应用的实施

价值

隐私计算对不同数据源进行横向与纵向联合建模，保证各方医疗数据安全。医疗机构可以在本地先建立模型，再通过隐私保护计算技术联合其他医疗机构更新模型参数，以最安全、最高效的方式提升模型诊断能力与诊断准确率。

医疗机构间的数据共享：
属于医疗机构、制药企业、基因测序机构、科研机构之间的横向场景，以增加样本数量进行建模。

跨机构跨域的医疗数据开放：
纳入了保险公司、运营商、政务系统、互联网等外部数据，以增加样本特征进行建模。

临床医学：
目前，国内外的应用实践案例也基本围绕临床医学研究、基因分析、疫情防控等场景展开。联合风控、联合营销、医保付费预测等场景也在解锁中。

一图看懂隐私保护计算发展

诺崴科技 NUOWEI TECH

隐私保护计算

- 联邦学习 (FL)
  - 2012年 王爽教授团队发表了全球首篇医学在线安全联邦学习文献，提出了"数据可用不可见"问题，解决隐私保护计算的基础性框架联邦学习的工程落地方案[1]
  - 2016年 Google提出了联邦学习在移动互联网上应用的概念，隐私机密计算技术被完全认可并进入快速发展阶段[2]
  - 2019年 杨强教授团队携手腾讯/微众银行，用联邦学习技术发布开源项目FATE [3]
- 可信执行环境 (TEE)
  - 2009年 OMTP提出了TEE标准。2015年Intel发布首款商业化支持TEE方案的CPU, Intel SGX [4]
  - 2016年 王爽教授团队完成基于TEE和安全联邦学习的全球首例支持跨多个国家的罕见病跨国医疗数据隐私保护下互联互通[5]。并获得Intel杰出贡献奖。
  - 2018年 百度发布Mesa TEE解决方案[6]。2020年阿里巴巴发布Occlum TEE 系统[7]，可信计算环境进入高速商用发展阶段。
- 密码学
  - 同态加密 (HE)
    - 1978年 随着非对称式加密算法RSA出现，同态加密的概念被首次提出[8]
    - 2009年 Gentry提出了首个实用的全同态加密算法，标志着全同态计算时代的开始[9]；
    - 2015年 王爽教授完成基于同态加密的罕见病多中心研究[10] 2年后，王爽牵头成立国际同态加密标准委员会，同态加密在全球进入高速发展阶段[11]
  - 多方安全计算 (MPC)
    - 1981年 Rabin首次提出通过oblivious transfe(OT) 协议实现机密信息交互[12]，奠定多方安全计算理论基础
    - 1982年 姚期智教授提出百万富翁问题，引入安全两方计算[13]。1987年GMW拓展到多方安全计算[14]。
    - 2016年 王爽教授拓展MPC/联邦学习服务国家医疗网络联合分析[15]

新基建&应用

- 医疗
  - 临床研究
  - 新药研发
  - 科研协作
- 金融
  - 征信
  - 获客
  - …
- 政务
  - 智慧城市
  - 融资贷款
  - 开放平台
  - …

参考文献

[1] S. Wang et al., "EXPLORER", Biomed. Inform., 2012.
[2] J. Konečný et al.,"FL: Strategies for Improving Communication Efficiency" ArXiv, 2016.
[3] D. Gao et al. "Hierarchical Heterogeneous Horizontal FL for EEG", ArXiv, 2019
[4] Trusted execution environment," Wikipedia, 2021.
[5] F. Chen et al., "PRINCESS" Bioinformatics, Vol. 33, no. 6, p. 871, Jan. 2017.
[6] "MesaTEE开源：隐私保护的高性能通用安全计算终成现实 - 百度安全社区.", 2021.
[7] Y. Shen et al.,"Occlum", The 25th Int Conf on Architectural Support for Programming Languages and Operating Systems, 2020.
[8] R. L. Rivest and L. Adleman, On data banks and privacy homomorphisms," Foundations of secure, 1978.
[9] C. Gentry, "FHE using ideal lattices", ACM symposium on Symposium on theory of computing, 2009.
[10] S. Wang, "HEALER", Bioinformatics, 2015
[11] "Applications of Homomorphic Encryption Standard", https://homomorphicencryption.org/standard/ 2017.
[12] M Rabin, "How to exchange secrets by oblivious transfer",Harvard Aiken Comp, 1981
[13] A. C. Yao, 23rd Annual S… Foundations of Computer Science (sfcs 1982), 1982.
[14] O. Goldreich, et al, "I… ntal game", ACM Symposium on Theory of Computing, 1987
[15] S.Wang. "SMAC-GL…

典型应用场景

锘崴信®隐私保护计算产品部署并实现应用转化：超100个

锘崴科技 NUOWEI TECH

- 医疗多中心研究平台
- 院内科研协作平台
- 生信科研协作平台
- 医疗大数据应用开放平台
- 生信分析数据应用平台
- 生信数据处理平台
- 基因数据隐匿查询服务系统
- 数据安全防护专病库
- 区域医疗影像数据应用开放
- 区域传染病防控研究平台
- 区域健康医疗数据治理开放
- 区域医疗健康数据应用平台
- 多中心基因组关联分析
- 罕见病基因隐私查询服务平台
- 医院内部数据隐私计算平台

- 金融隐私计算平台
- 金融助贷服务
- 跨境电商助贷
- 睡眠卡激活
- 银行新卡营销
- 地方金融服务数据平台
- 金融防赌反诈服务

- 数据要素流通平台
- 政务数据开放应用平台
- 公共数据应用开放平台
- 城市综合治理数据安全服务
- 工业大数据应用开放平台
- 区域数据应用平台
- 政务数据互通管理方案

- AI云隐私保护计算平台
- AI验证平台+数据应用服务
- 智能外呼隐私计算云平台
- 连锁酒店AI一体化防疫平台

- 智慧校园监控方案
- 带隐私保护的人像识别方案
- 矿业企业安全生产监控方案
- 企业数据安全存储

- 精准定位数据的营销平台
- 全链路归因数字营销平台
- 保险业存量客户转化营销
- 银行营销管理服务

- 敏感人群数据管理隐私保护
- 基与TEE的密钥管理系统
- 云SaaS隐私计算服务平台

医疗 金融 政务 AI 营销 其他

2022 CHINC

2022中华医院信息网络大会

暨中国医疗信息技术展览

CHINC

医疗应用案例：全球首创-基于隐私保护的超大规模医学科研网络pSCANNER

诺崴科技 NUOWEI TECH

项目需求

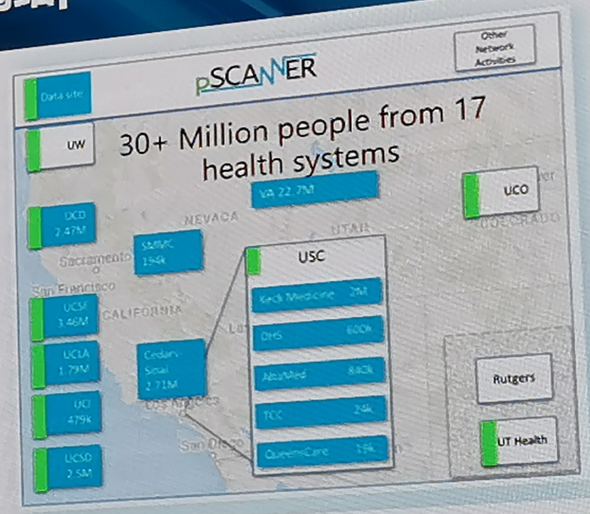
- 在实践中，单一中心样本量和样本维度往往难以支撑一项研究的进行，需要多家机构/中心合作以增加样本量、丰富数据维度。因此，生物医疗数据开放互联尤为重要。而为了保护患者隐私，相关法律法规严格限制生物医学数据的不安全流动，这也意味着需要将明文数据物理聚合的传统集中式计算不再适用于医疗场景。

项目方案

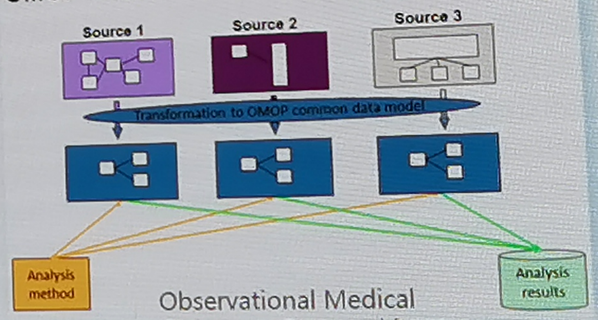
- 依托诺崴的隐私保护计算平台以及安全联邦学习等底层技术，构建了一套名为pSCANNER的医疗数据互联网络，应用在美国国家级医疗网络中。

项目效果

- pSCANNER覆盖了17个医共体下约300家医院，接近三千万患者数据。通过诺崴的技术，实现在带有隐私保护的前提下，跨多个机构，在3000万[illegible]成多中心的联合计算研究。



OMOP Common Data Model



Observational Medical Outcomes Partnership (OMOP) CDM

Epic, VISTA (VA), Sage system (QueensCare), Allscripts-Sunrise (UCI before 2017)

pSCANNER, JAMIA, 2014, Source: http://[illegible]anner.ucsd.edu/data-sites

医疗应用案例：全球首创-隐私保护的跨国医疗大数据分析系统

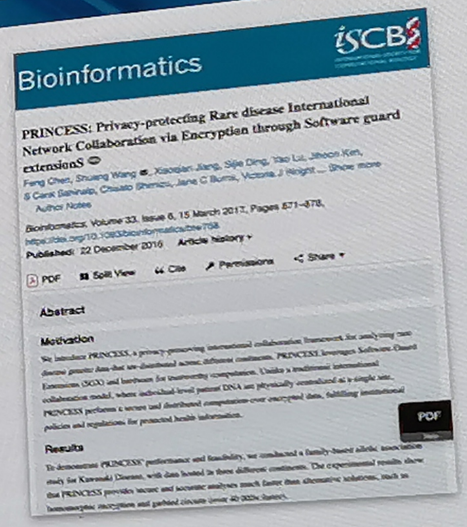
诺崴科技 NUOWEI TECH

项目需求

- 罕见病研究往往受限于单中心数据量不足而无法进行，而生物医疗数据的跨中心流动则又受到法律监管，对安全措施要求高。儿童罕见病联盟为推进儿童川崎病研究，需要联合三国（美国、英国、新加坡）医疗机构的数据。

项目方案

- 依托诺崴信®隐私保护计算平台为其开发了一套带有隐私保护的跨国多中心数据协作系统，用于分析儿童川崎病基因数据。利用该系统，可以对加密数据执行安全的分布式计算，解决医疗数据跨境流动难的问题，保证所有数据共享符合各国数据流动法规监管要求。在这一过程中，不论是有意或无意，都不会泄露个人隐私数据及中间结果，同时，不会引入显著的计算负荷或大的限制，使得安全的大规模跨国遗传数据分析在实践中的可行性大幅提高。


Bioinformatics ISCB

PRINCESS: Privacy-protecting Rare disease International Network Collaboration via Encryption through Software guard extensionS

Feng Chen, Shuang Wang, Xiaoqian Jiang, Sijie Ding, Yao Lu, Jihoon Kim, S Cenk Sahinalp, Chisato Shimizu, Jane C Burns, Victoria J Wright ... Show more

Author Notes

Bioinformatics, Volume 33, Issue 6, 15 March 2017, Pages 871–878, https://doi.org/10.1093/bioinformatics/btw758

Published: 22 December 2016 Article history

PDF Split View Cite Permissions Share

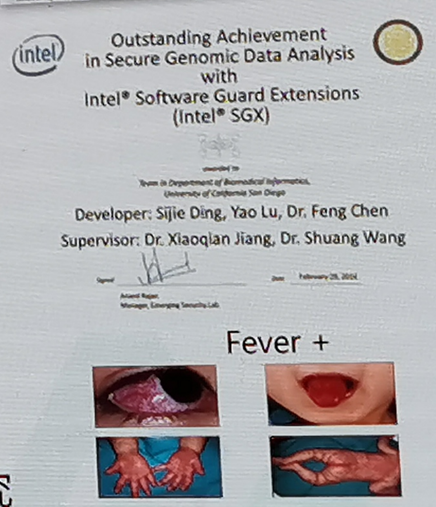
Abstract

Motivation

We introduce PRINCESS, a privacy-preserving international collaboration framework for analyzing rare disease genetic data that are distributed across different continents. PRINCESS leverages Software Guard Extensions (SGX) and hardware for trustworthy computations. Unlike a traditional international collaboration model, where individual-level patient DNA are physically centralized at a single site, PRINCESS performs a secure and distributed computation over encrypted data, fulfilling institutional policies and regulations for protected health information.

Results

To demonstrate PRINCESS' performance and feasibility, we conducted a family-based allelic association study for Kawasaki Disease, with data hosted in three different continents. The experimental results show that PRINCESS provides secure and accurate analyses much faster than alternative solutions, such as homomorphic encryption and garbled circuits (over 40 000× faster).


Outstanding Achievement in Secure Genomic Data Analysis with Intel® Software Guard Extensions (Intel® SGX)

awarded to

Team in Department of Biomedical Informatics, University of California San Diego

Developer: Sijie Ding, Yao Lu, Dr. Feng Chen

Supervisor: Dr. Xiaoqian Jiang, Dr. Shuang Wang

Signed Date February 28, 2016

Ariard Rajan, Manager, Emerging Security Lab

Fever +

通过隐私机密计算技术，实现多方在原始数据不透露、加密安全计算环境下完成川崎病研究

医疗应用案例：全国首套-基于隐私计算的跨省多中心基因分析系统

锘崴科技 NUOWEI TECH

□项目需求

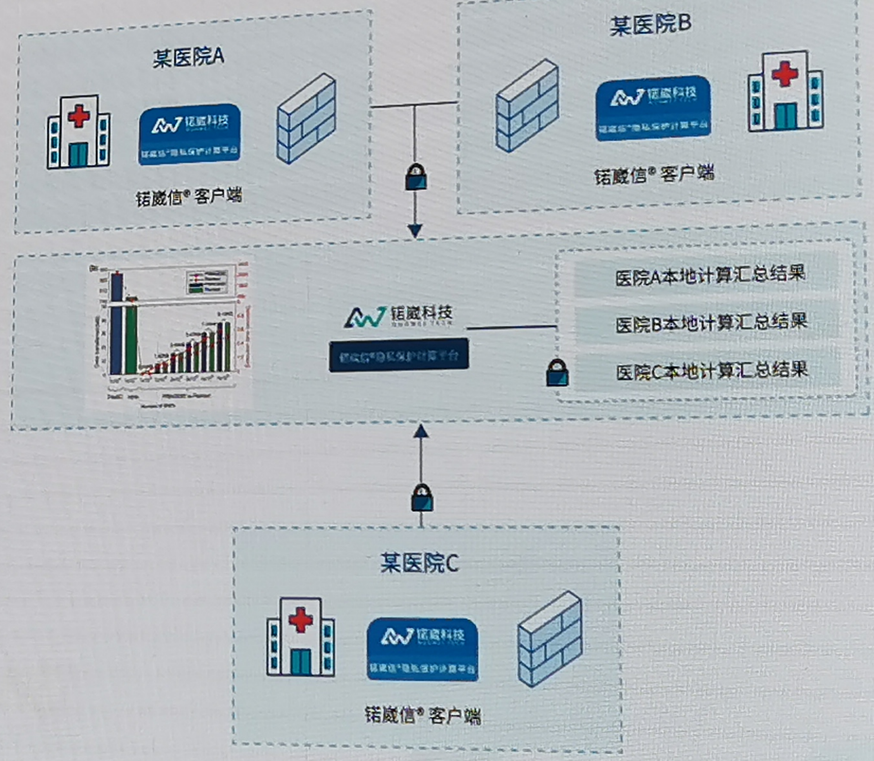
- 某三甲医院要进行有关强直性脊柱炎的全基因组关联分析。基因数据具有高通量、高敏感度特点，分析传输难度大。同时，单家医院/研究机构的数据量不足以支持全基因组关联分析研究，加上法律明确限制医疗数据的不安全共享。

□项目方案

- 依托锘崴信®隐私保护计算平台为该医院构建开发了一套跨省级多中心基因数据分析系统，所有数据物理分散，逻辑集中，可实现多家医疗机构数据虚拟聚合，也满足全基因组关联分析大数据量需求，帮助研究顺利完成。

□方案效果

- 项目纳入上海市科技进步奖一等奖；
- 研究成果刊发在顶级生物医学期刊Briefing in Bioinfomatics上；
- **精度**：等价于数据集中方式；
- **算法时间**：等价于数据集中方式；
- **特征靶点**：与集中计算结果一致；
- **统计意义**：高1个数量级。



医疗应用案例：全国首例-儿童罕见病基因隐私查询网络

诺崴科技 NUOWEI TECH

项目需求

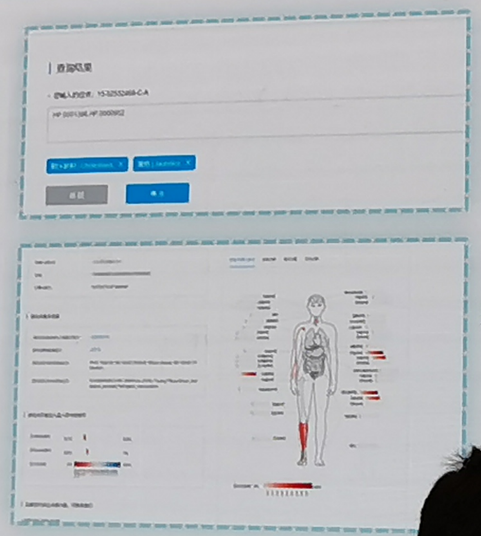
- 外显子组测试作为一种基因测序技术，对于儿科疾病的临床诊断有极大帮助。然而，现在使用的隐私保护手段很难保证在表型数据（phenotype）和基因型数据（genotype）匹配过程中基因数据的安全。某研究机构希望能在保证隐私安全的前提下，进行一项中国儿童致病变异表型的研究。

项目方案

- 依托诺崴信®隐私保护计算平台为其开发了一个用于中国儿童群体变异表型遗传诊断研究的隐私保护在线系统（如图）。该系统利用可信执行环境等技术为单个变体提供表型谱的隐私保护查询（PIR），在数据匹配的过程中能够保护用于查询的数据信息的隐私安全，同时对后端模型和数据以及最终结果的隐私安全进行保护。

项目效果

- 在兼具性能和安全性的前提下实现带有隐私保护的查询，本案例覆盖超大数据量，包含 20,909 例患者和 3,152,508 个变异体。
- 针对荷载的均衡性进行了算法优化，更好地提升计算效率，避免在大样本量和高并发情况下可能导致的性能下降。



医疗应用案例：基于隐私保护计算驱动的癌症多中心CDR大数据分享

诺崴科技 NUOWEI TECH

项目需求

- 医疗临床数据库或专病数据网络的构建有助于提高科研效率，进一步挖掘临床数据价值，可为临床医生提供更多的真实世界证据、辅助临床决策。

项目方案

- 通过隐私计算构建带有隐私保护的医疗临床数据库或专病数据网络能够解决隐私安全隐患。依托诺崴信®隐私保护计算平台，中华医学会消化外科结直肠癌学组将隐私计算技术应用到类似医疗临床数据库的搭建中，实现了全国范围内带有隐私保护的结直肠癌数据共享。该数据库支持标准的eCRF输入，每家医院也可单独管理自己的数据及密钥，不同医院之间又能进行跨院的联合数据统计、分析等，兼顾了隐私保护和数据共享的双重目标。

项目效果

- 在诺崴技术的支持下，横跨24个省/直辖市/自治区，覆盖60+三甲医院、72,650例患者的回顾性数据。为相关领域的科研项目提供了高质量的数据样本支持，也进一步推动了相关防治、干预措施的发展进程，填补了国内相关领域的空白。

中国地图

新疆维吾尔自治区录入量：237
甘肃省录入量：323
青海省录入量：186
重庆市录入量：85
宁夏回族自治区录入量：1,997
四川省录入量：5,210
广西壮族自治区录入量：624
河南省录入量：239

内蒙古自治区录入量：137
陕西省录入量：2,728
吉林省录入量：1,838
辽宁省录入量：2,320
北京市录入量：4,172
河北省录入量：2,206
山东省录入量：4,862
山西省录入量：272
江苏省录入量：4,056
上海市录入量：20,930
安徽省录入量：101
湖北省录入量：4,760
江西省录入量：4,609
福建省录入量：6,005
广东省录入量：4,27
海南省录入量：474

新药研发

2022 CHINC 2022中华医院信息网络大会 中国医疗信息技术展览