ETH-RAT

SDSC

SWISS DATA SCIENCE CENTER
A DIGITALIZATION JOURNEY

Prof. Edouard Bugnion

Co-Academic Director, Swiss Data Science Center

Professor & Vice-President, EPFL

Open Data in Sciences Workshop, Brussels 31.01.2018

About me

18y

vmware Industry NuOVA CISCO

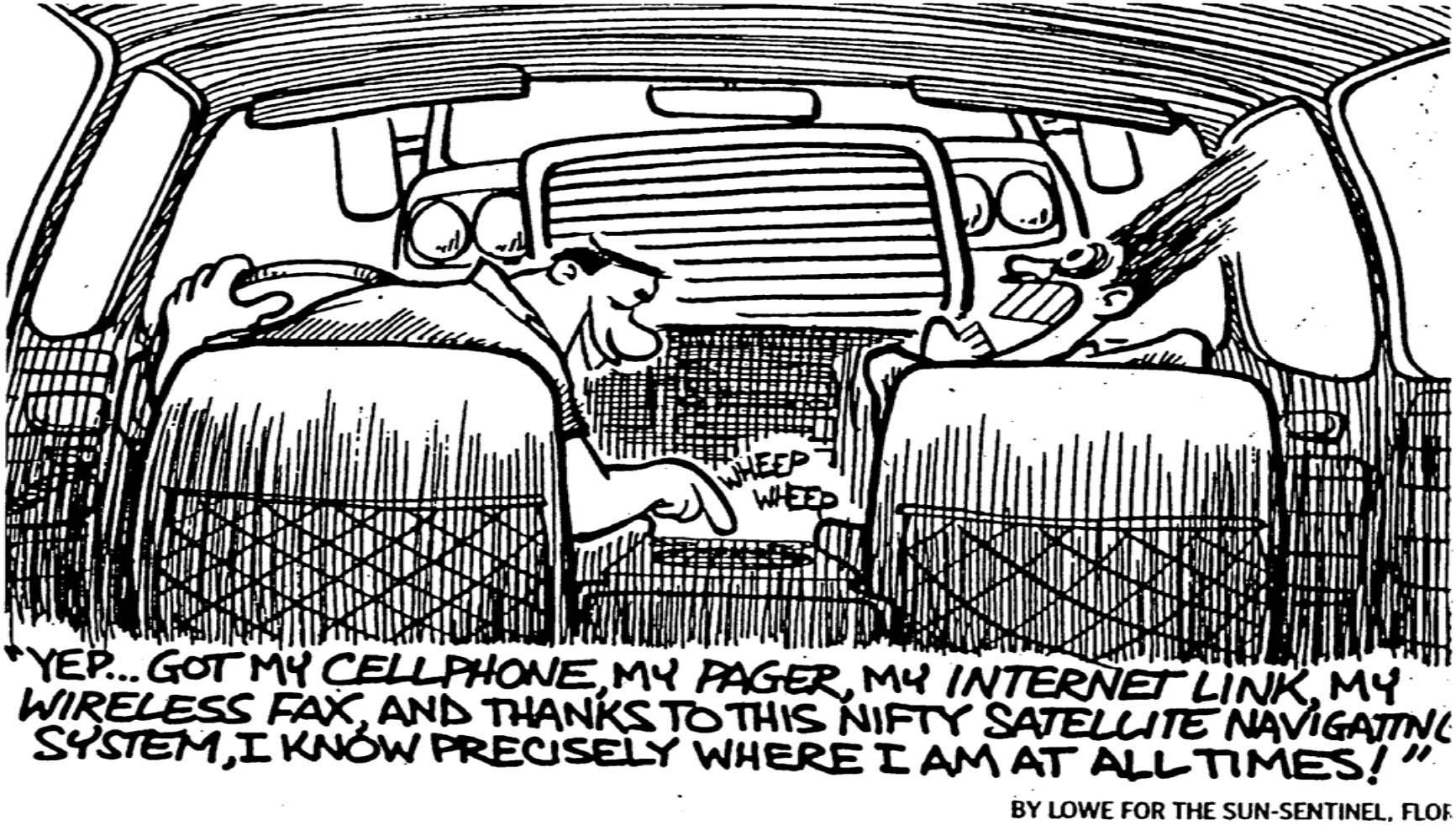
14y

Academia

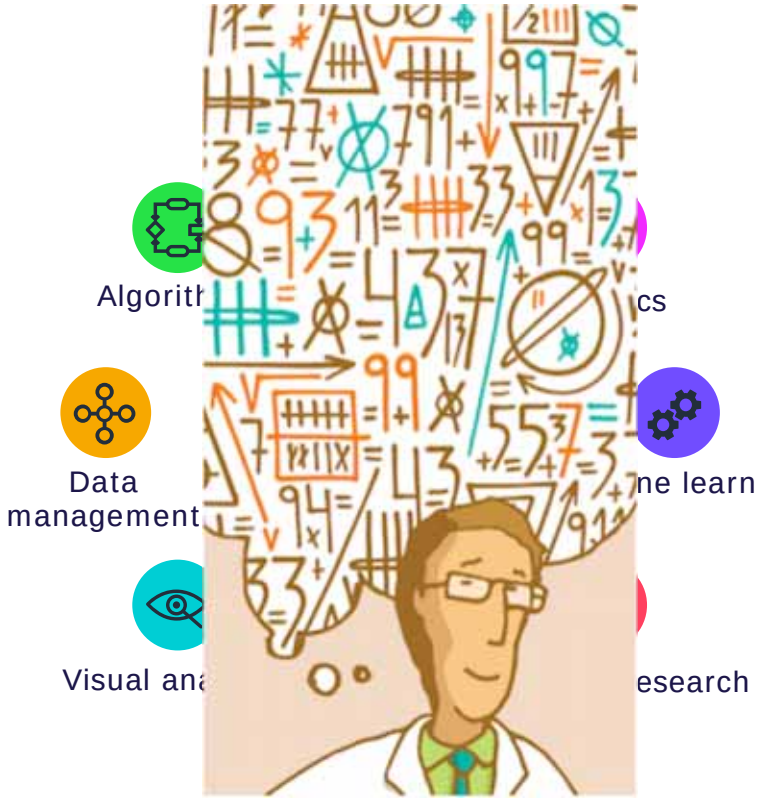
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

SDSC

Big Data, Bad Data



A fragmented ecosystem



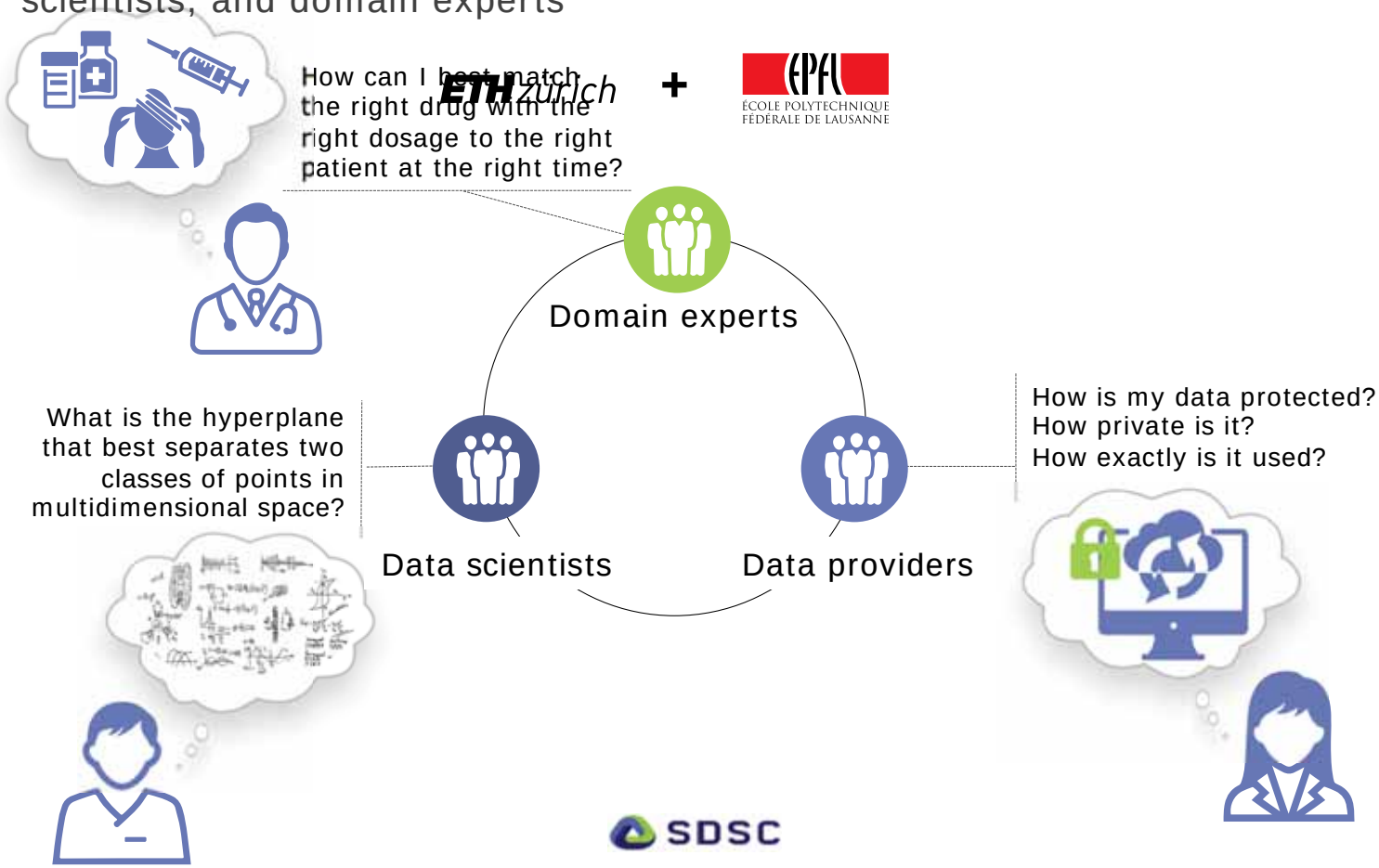
What is the hyperplane that best separates two classes of points in multidimensional space?

GAP

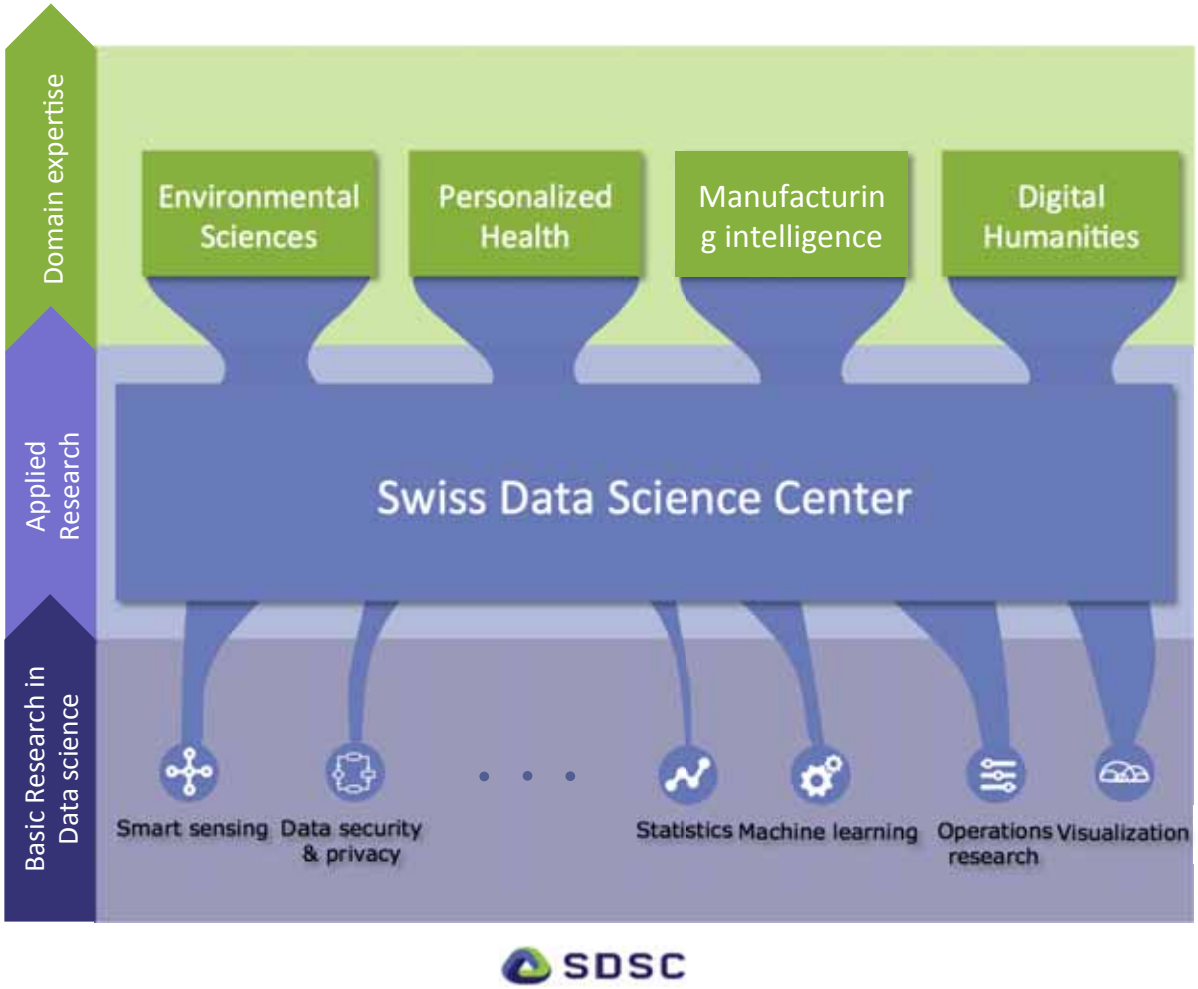
How can I best match the right drug with the right dosage to the right patient at the right time?

Swiss Data Science Center (SDSC)

Multi-disciplinary team of ~40 people from academia and industry
scientists, and domain experts



Where does SDSC fit?



What will the SDSC offer?

Embedded R&D collaboration

We engage in academic and industrial collaborations requiring large-scale distributed data processing (Big & Fast Data) and/or advanced analytics (machine learning & statistics) combined with an in-depth knowledge in select domains

Domain-specific Insights as a Service

We provide secure access to our cloud-hosted analytics platform - **RENGA**, a highly scalable open software platform offering a one-stop-shop for hosting and exploring curated, calibrated and possibly anonymized data at scale, at-rest or in-motion.

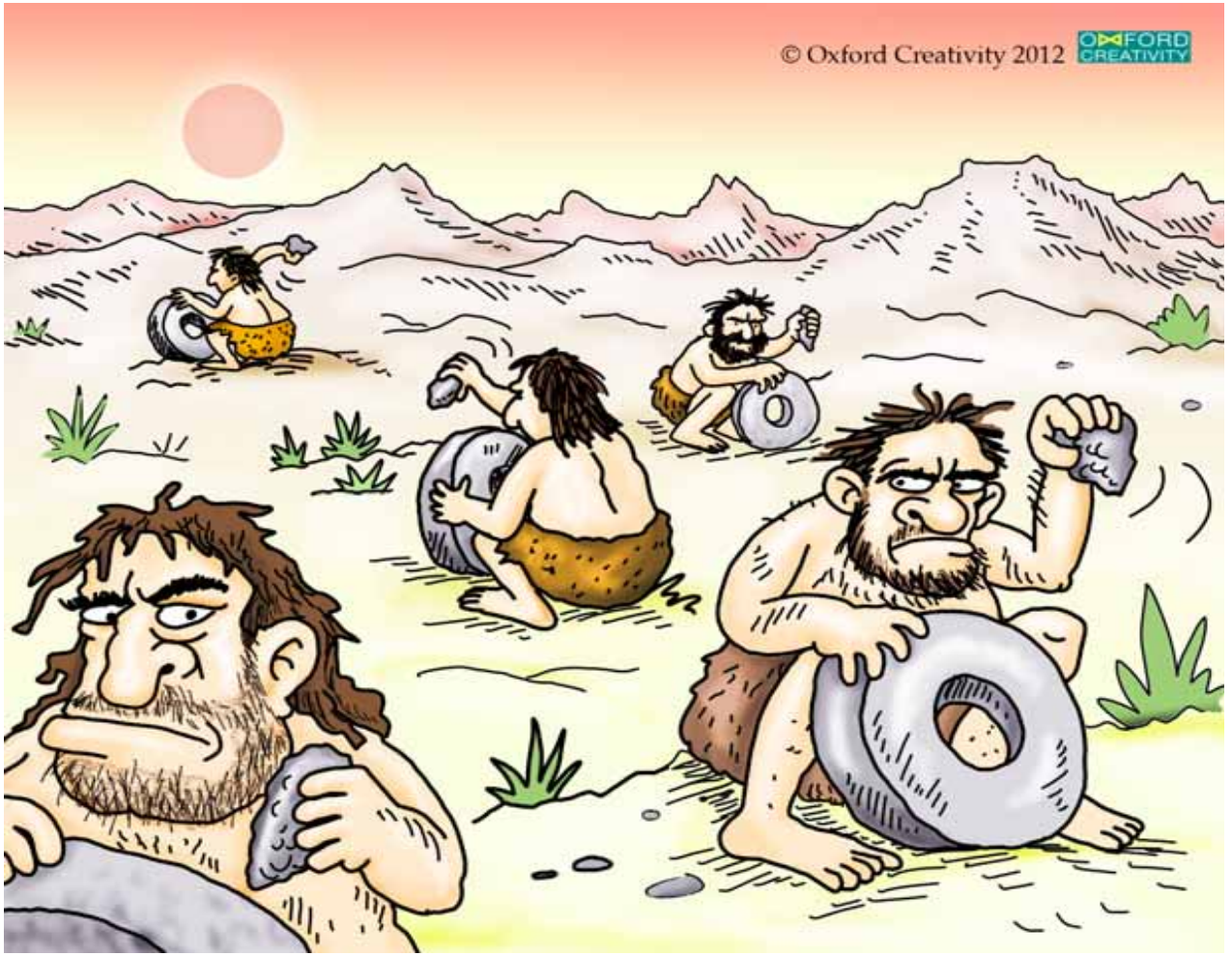
Open (Data) Science

RENGA offers user-friendly tooling and services to help with the adoption of Open Science, fostering research productivity and excellence.

SDSC Analytics Platform

SDSC

Status quo in Data Science

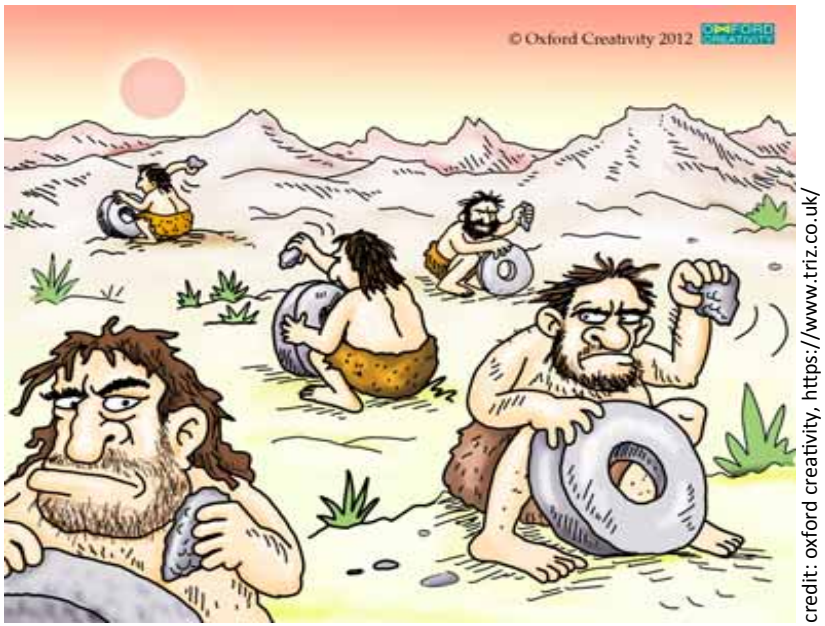


credit: oxford creativity, https://www.triz.co.uk/

Status quo in Data Science

1. Growing disconnect between actors in data science ventures
   - Work in silos, unaware of insights generated by others
   - Communication barriers, among humans & machines
2. R&D not always traceable / verifiable
   - Data and methods often not reusable

Scientific progress hampered by frequent re-inventions of the wheel

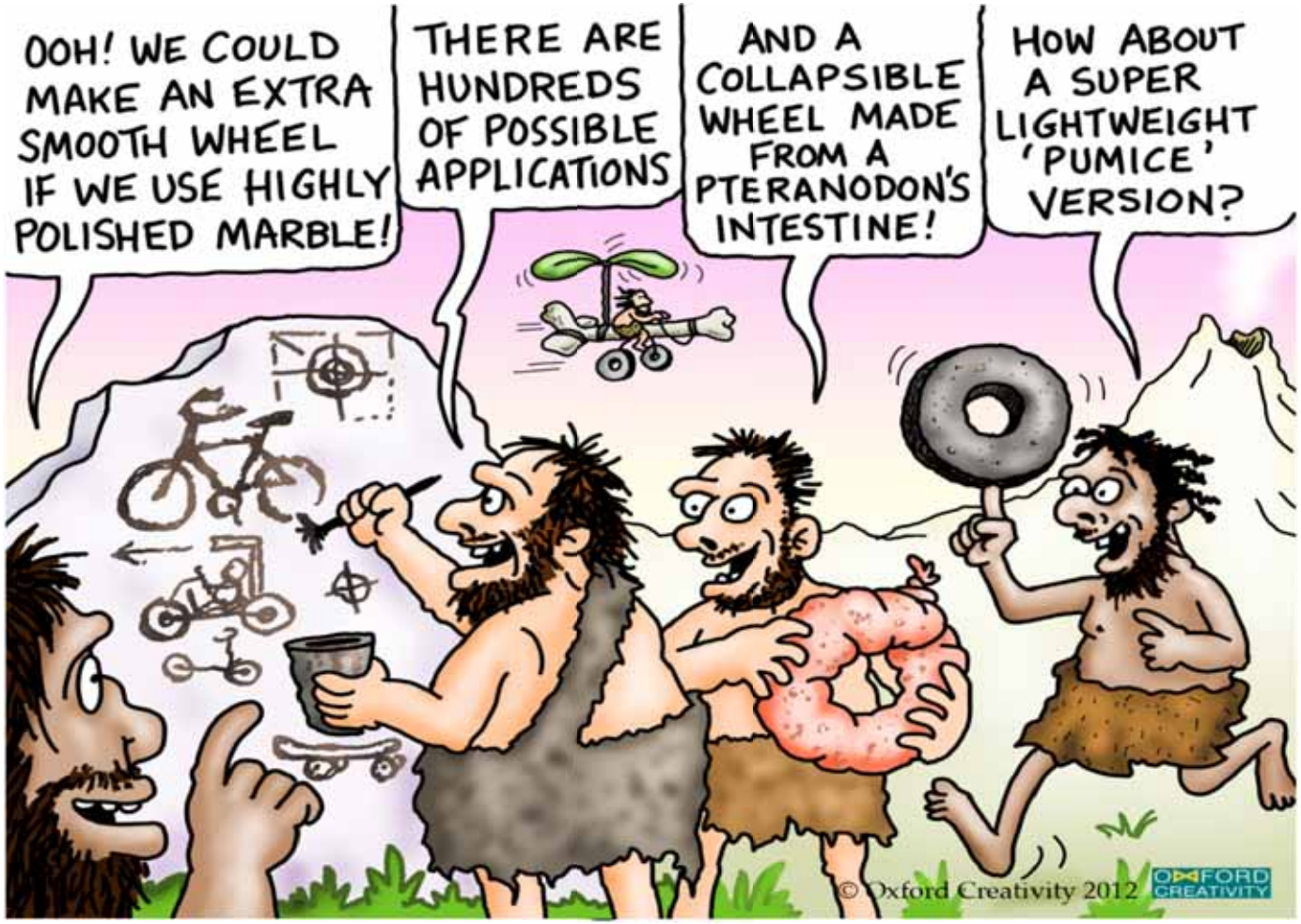


credit: oxford creativity, https://www.triz.co.uk/

Facilitate communication to foster innovation

credit: oxford creativity, https://www.triz.co.uk/

Foster multidisciplinary collaborations



credit: oxford creativity, https://www.triz.co.uk/

RENGA – SDSC Analytics Platform

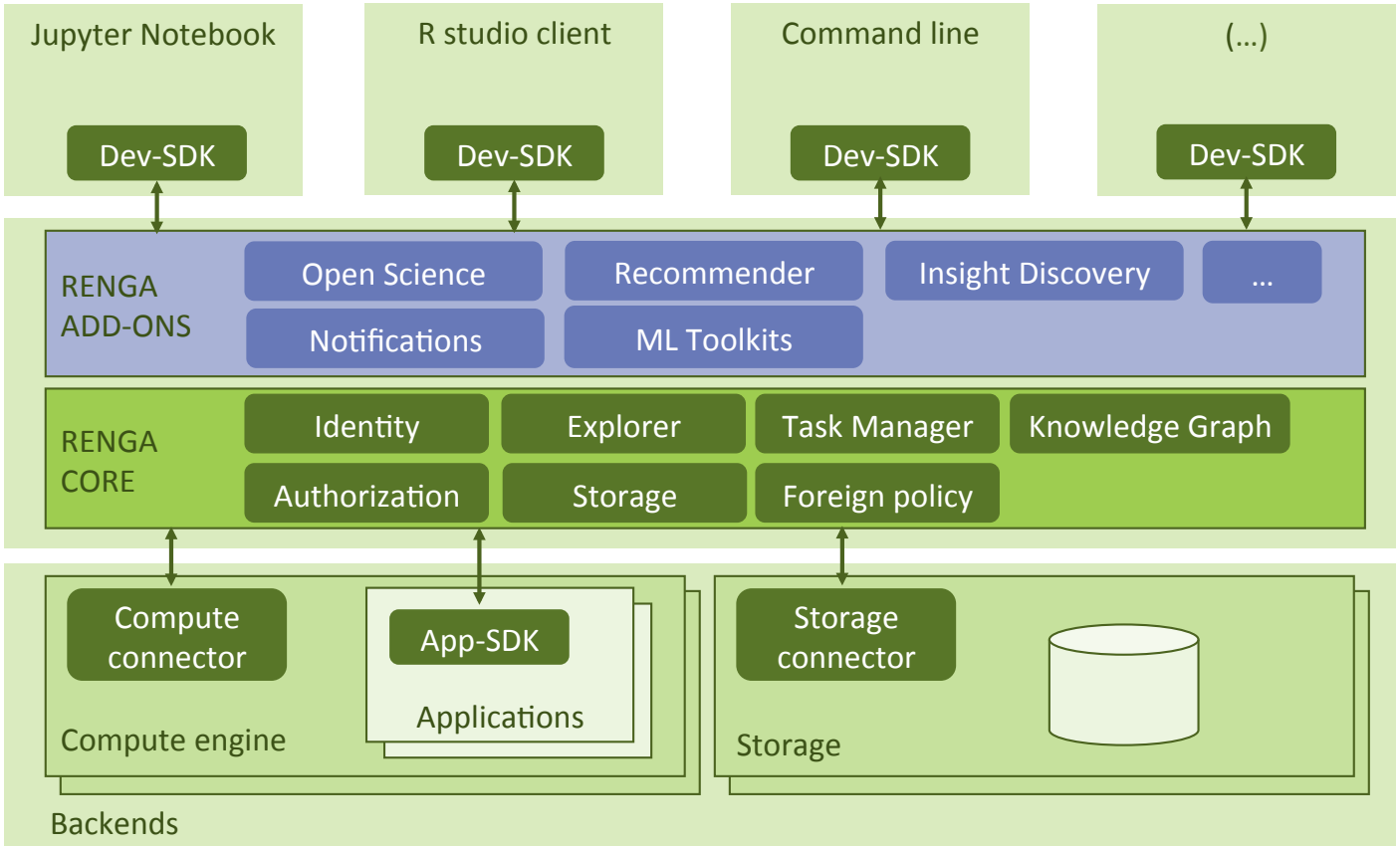
Renga (連歌), *plural* **renga**, *a genre of Japanese linked-verse poetry in which two or more poets supply alternating sections of a poem linked by verbal and thematic associations.*

—Encyclopædia Britannica

- RENGA is a highly-scalable & secure open software platform designed to foster multidisciplinary (data) science collaborations across mutually untrusted academic / industrial institutions
- The platform allows practitioners to:
  - Securely manage, share, find and process large-scale data across untrusted parties operating in a federated environment
  - Collaborate, recommend, share and discover new insights
  - Capture complete lineage automatically up to original raw data for detailed traceability and reproducible research
  - Adhere to FAIR principles and DMP mandate

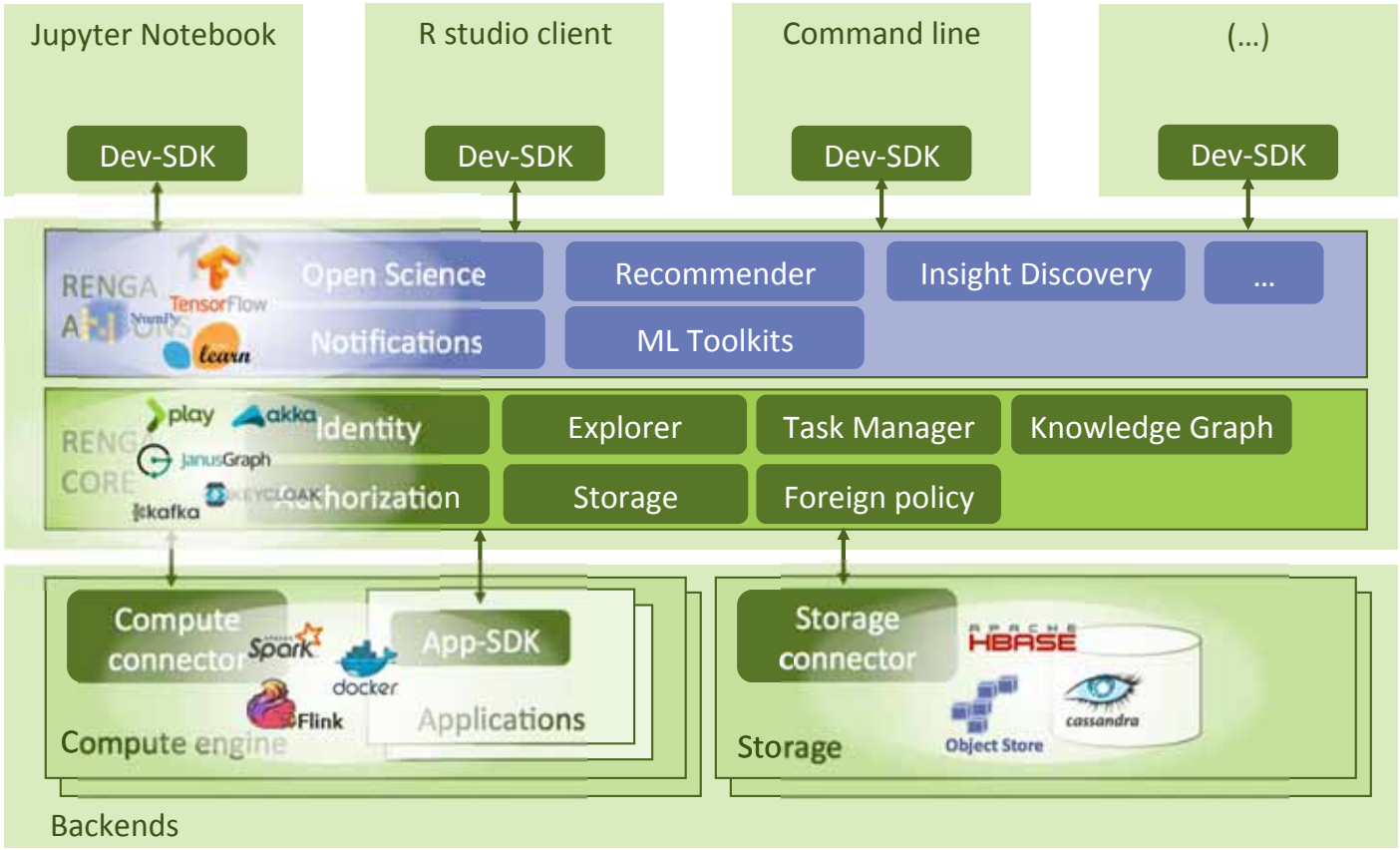
SDSC

RENGA – Highly Modular Architecture



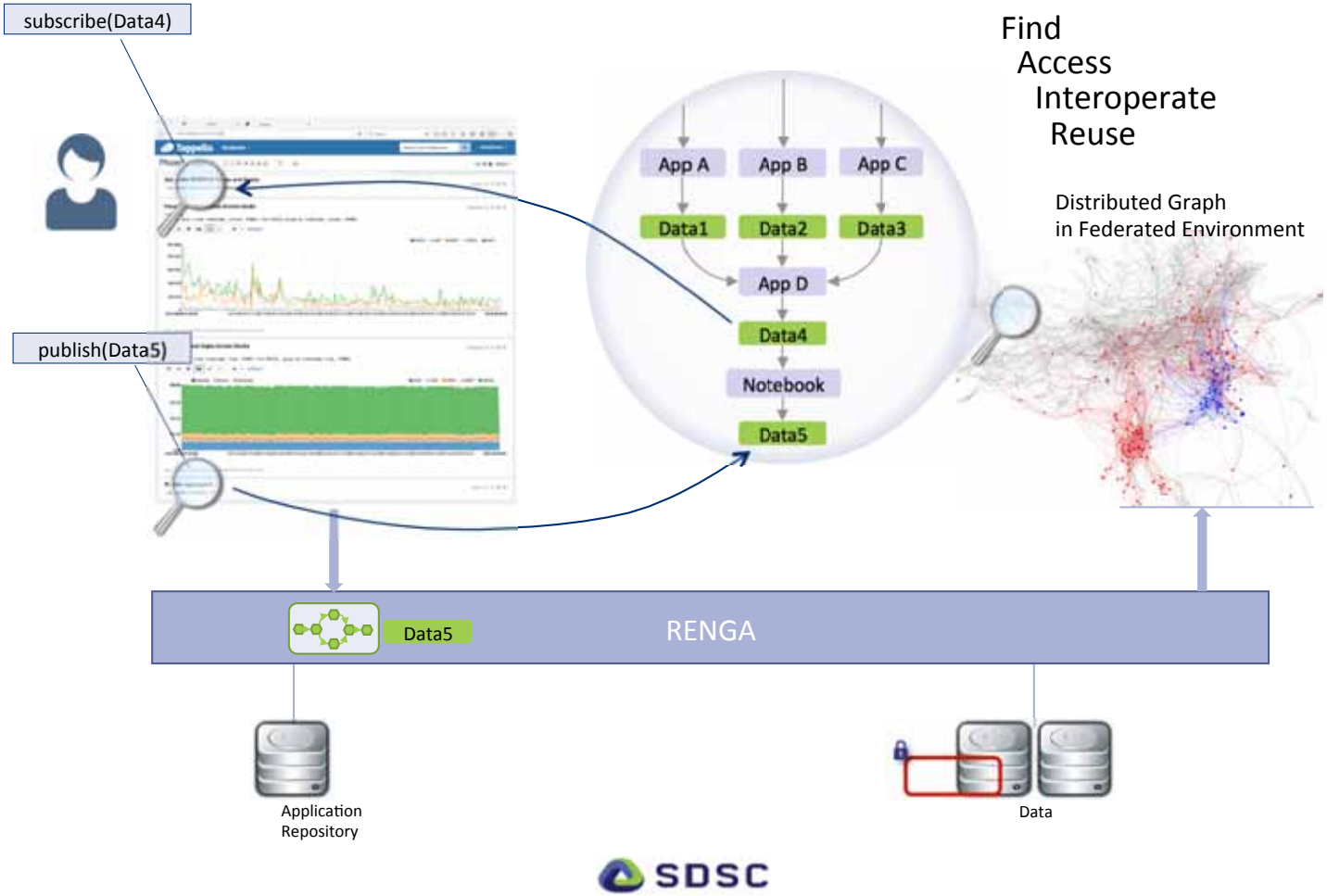
SDSC

RENGA – Built on Proven Technology

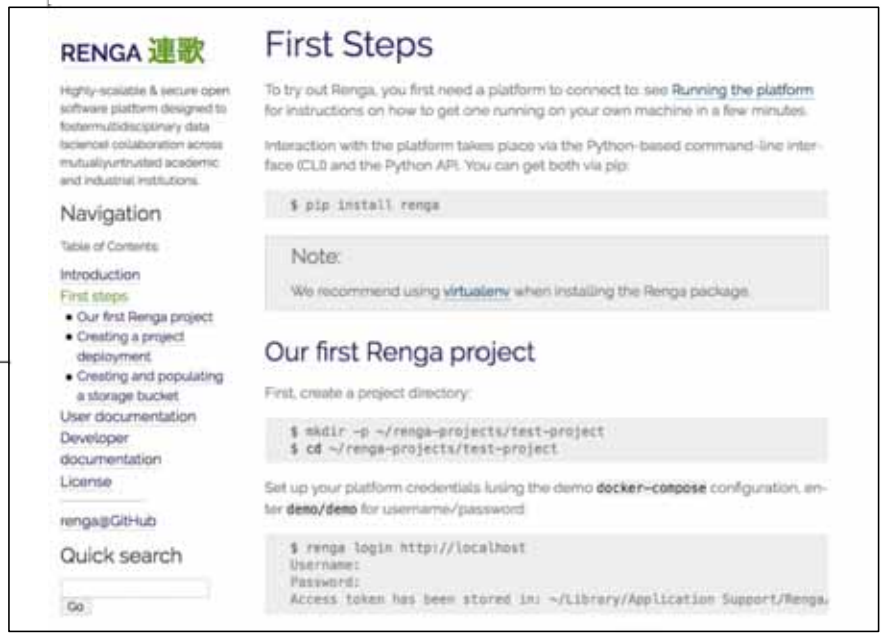


SDSC

Lineage in Data Science



Available as Open Source (Apache v2)



http://get-renga.io

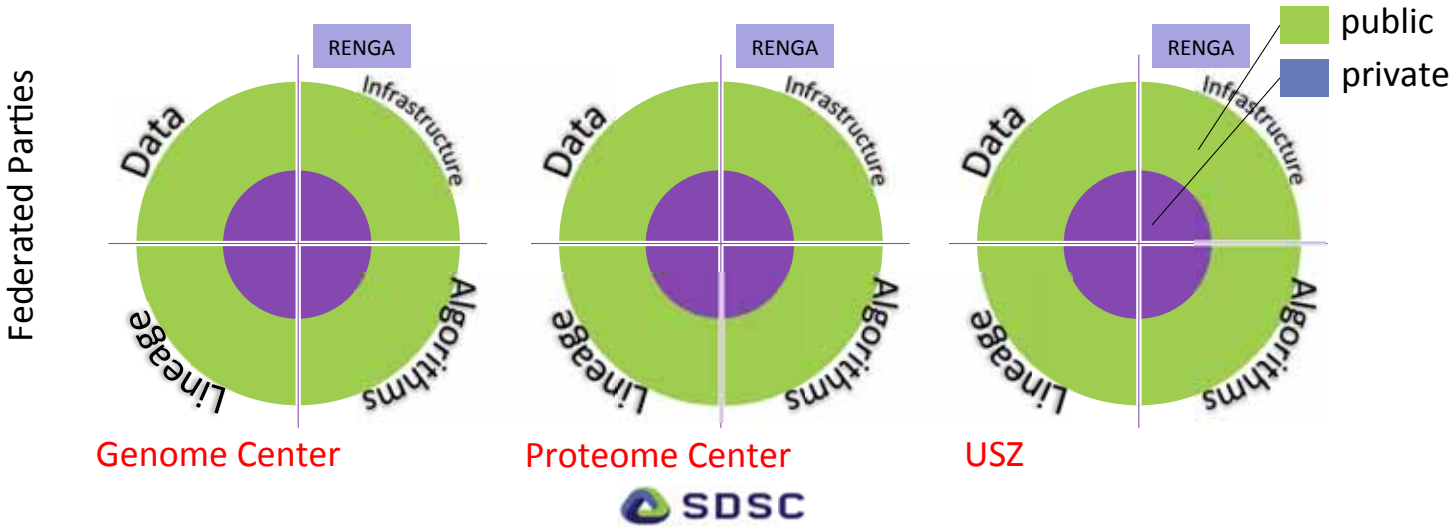
SDSC

Roamap – FEDERATED RENGA

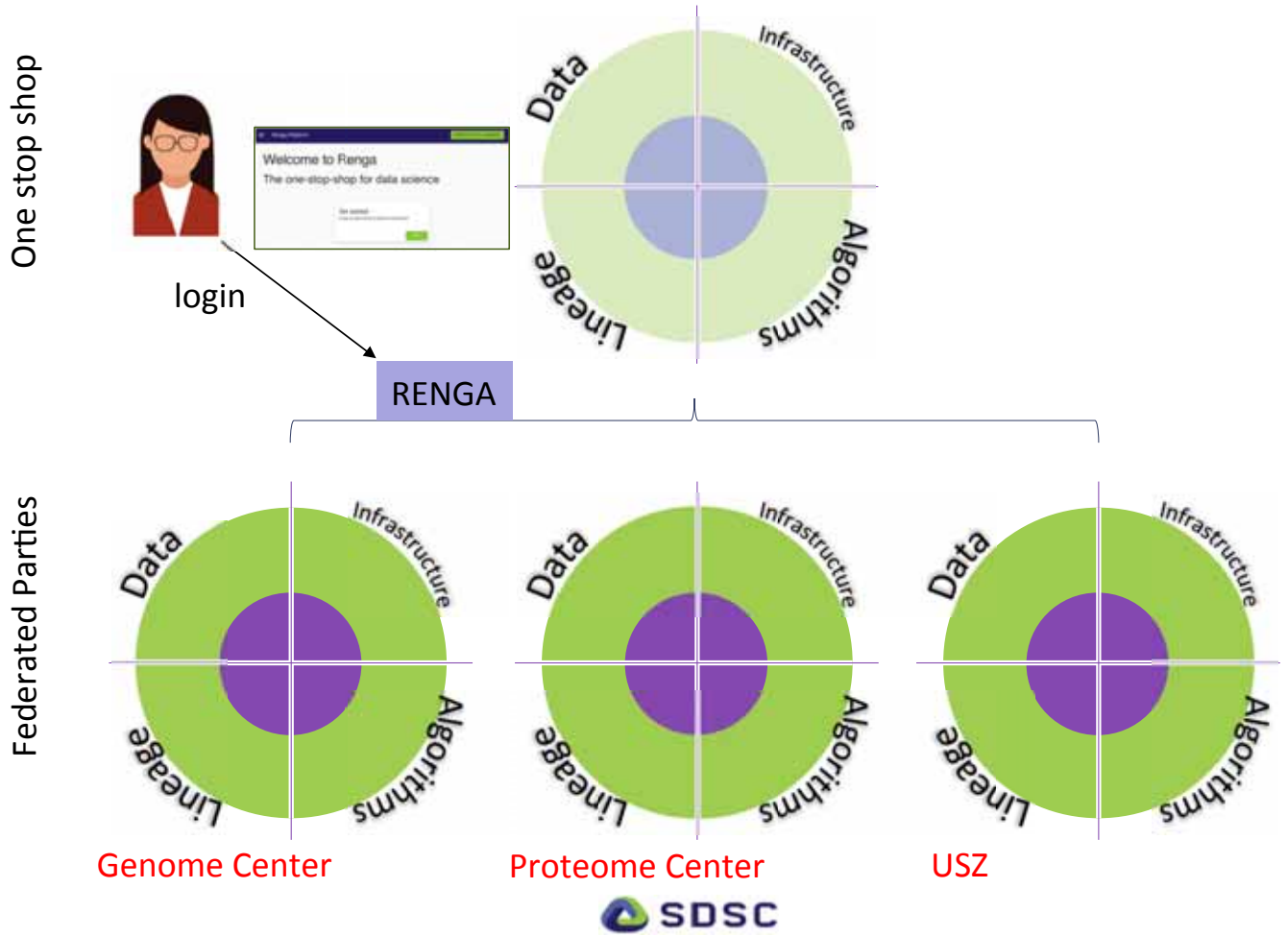
- Open (Data) Science in an privacy-conscious world
- Controlled data sharing

SDSC

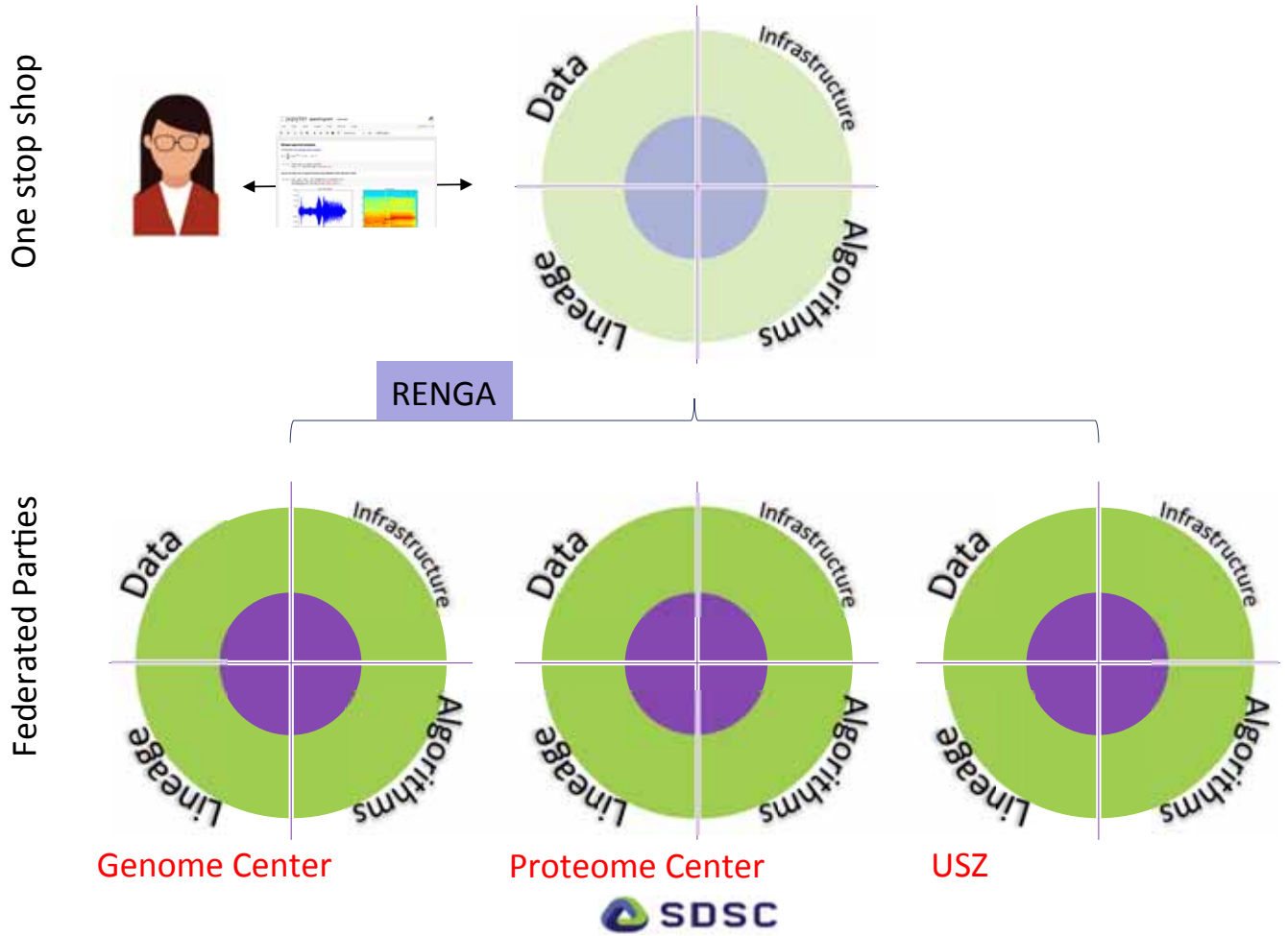
Use Case in Biomedical Research



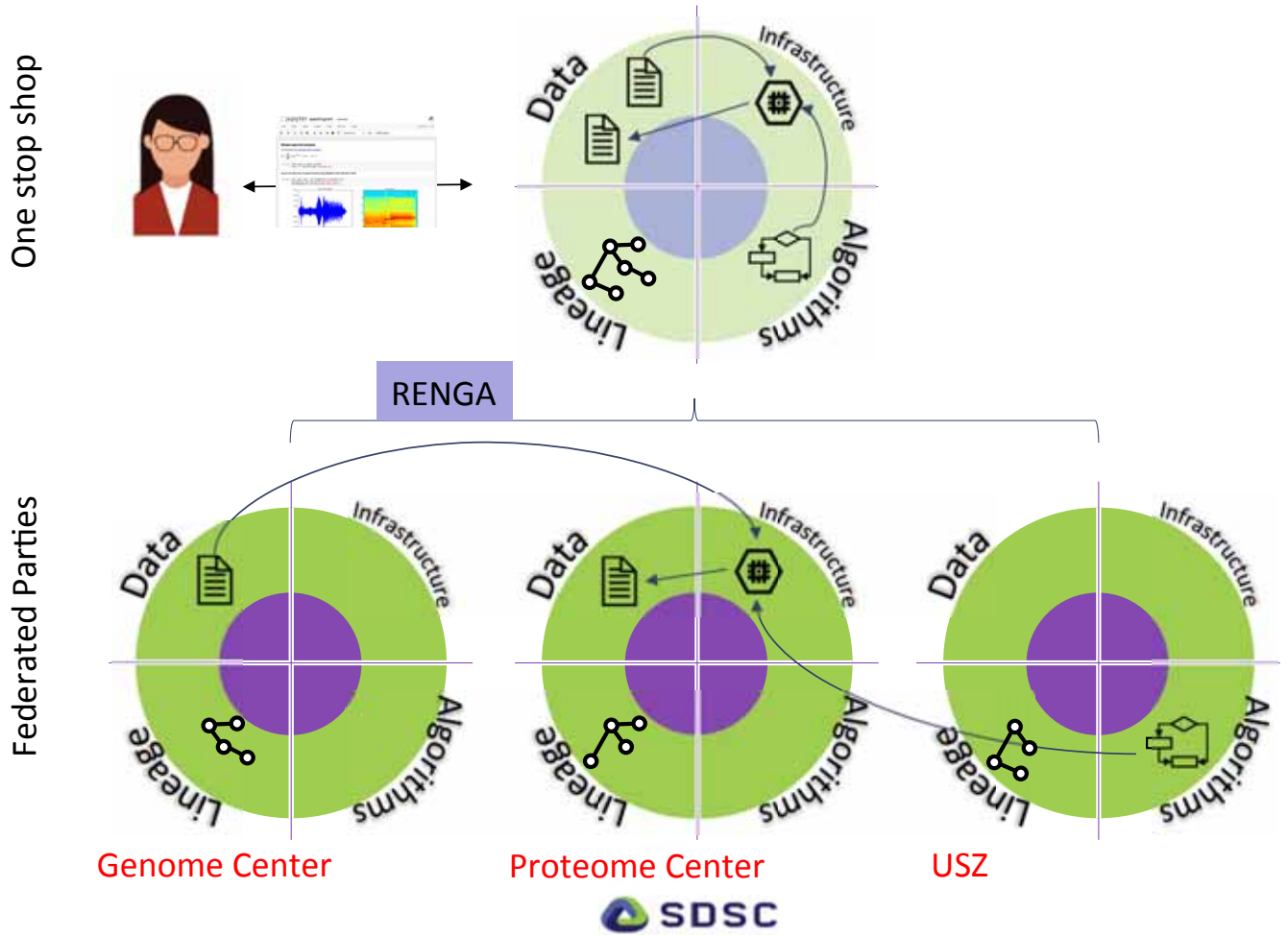
Use Case in Biomedical Research



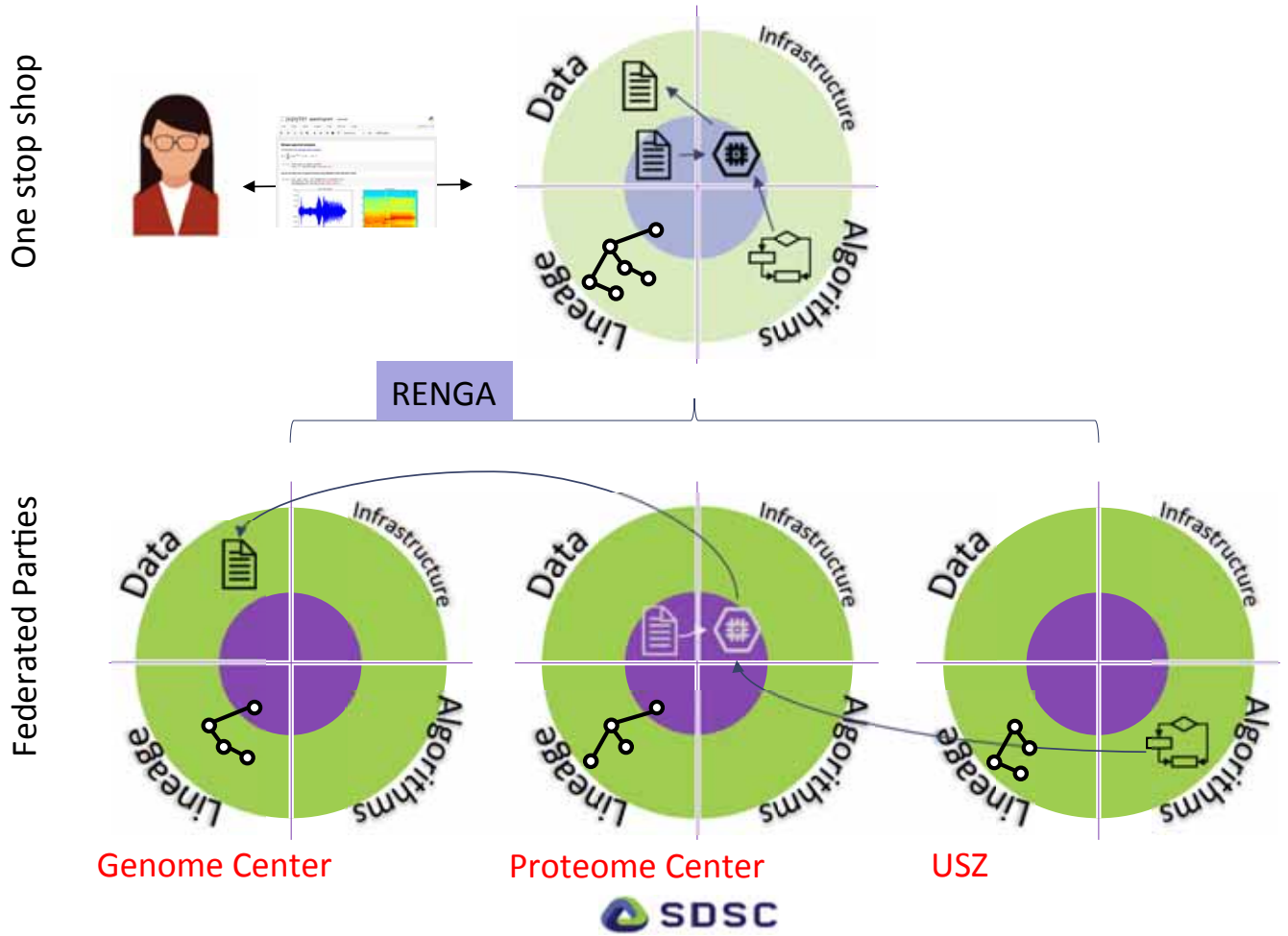
Use Case in Biomedical Research



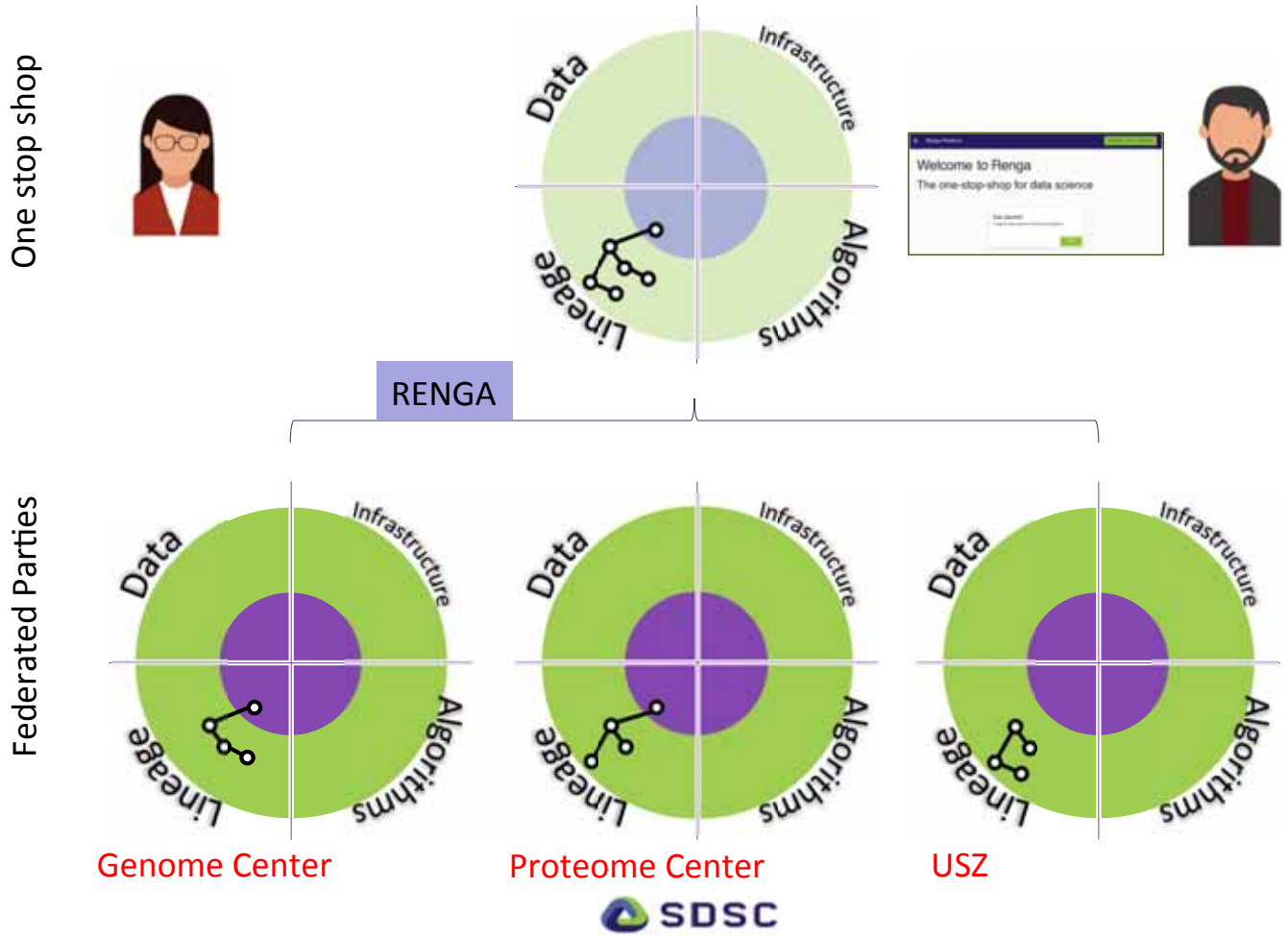
Use Case in Biomedical Research



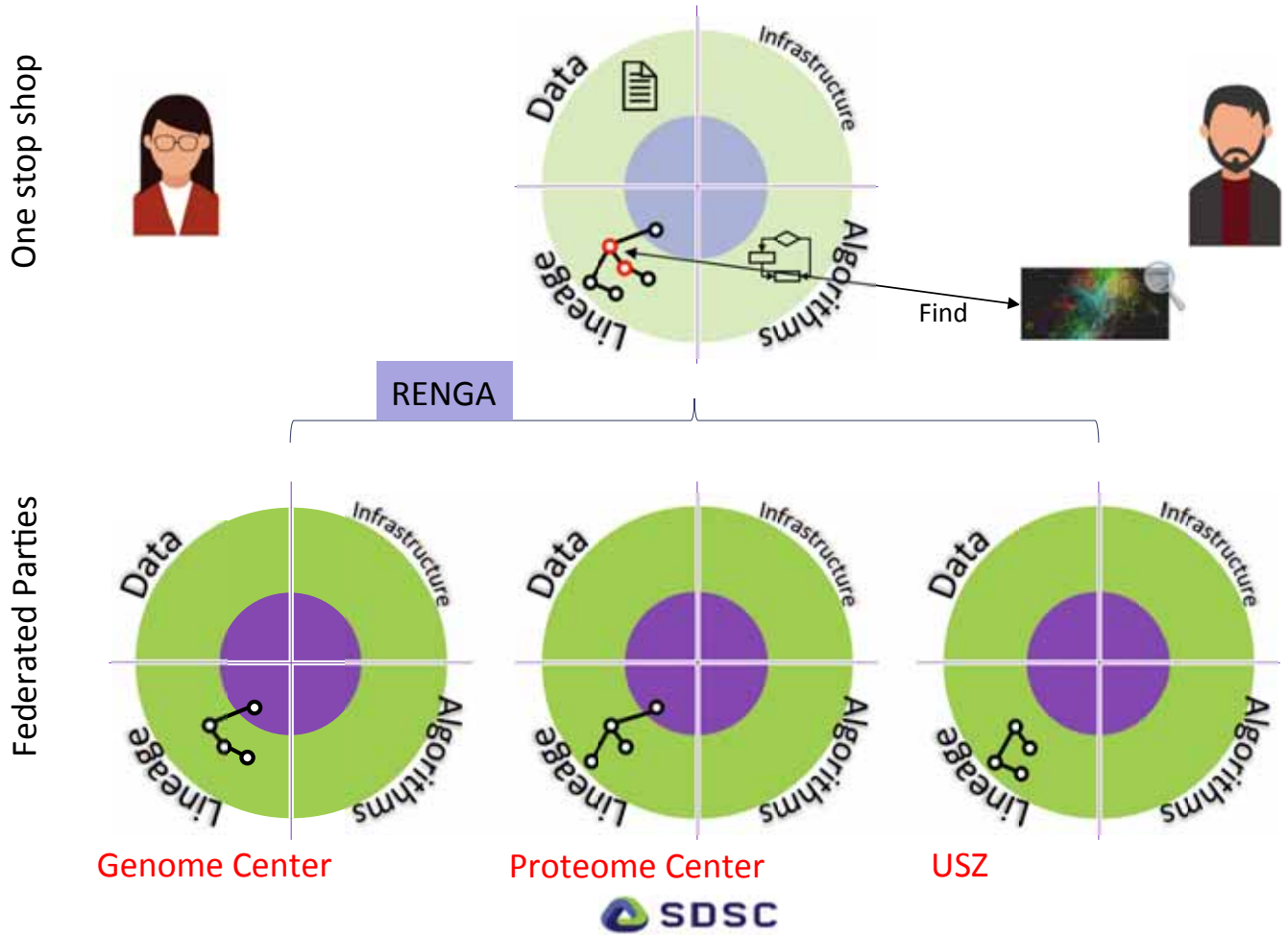
Use Case in Biomedical Research



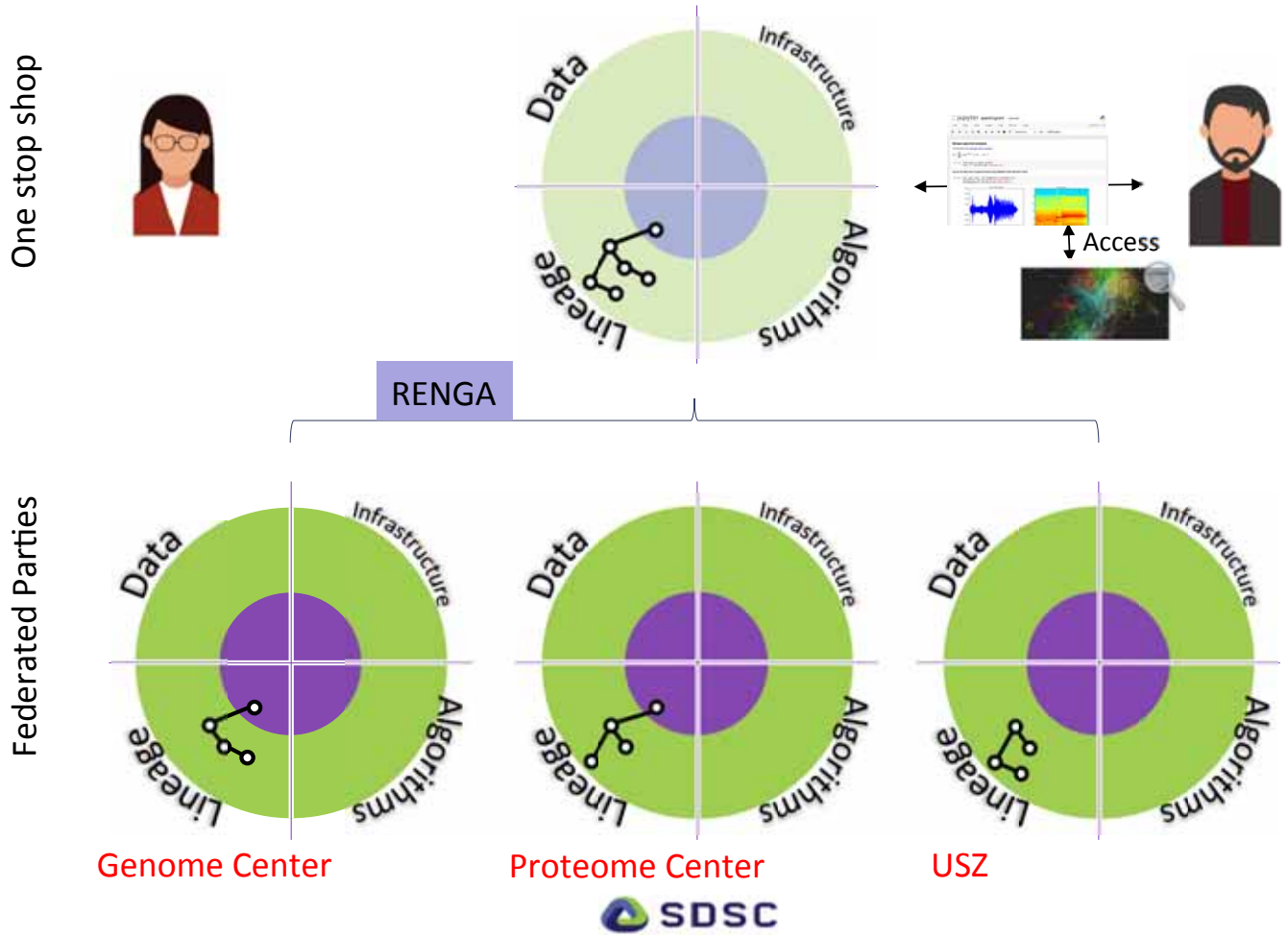
Use Case in Biomedical Research



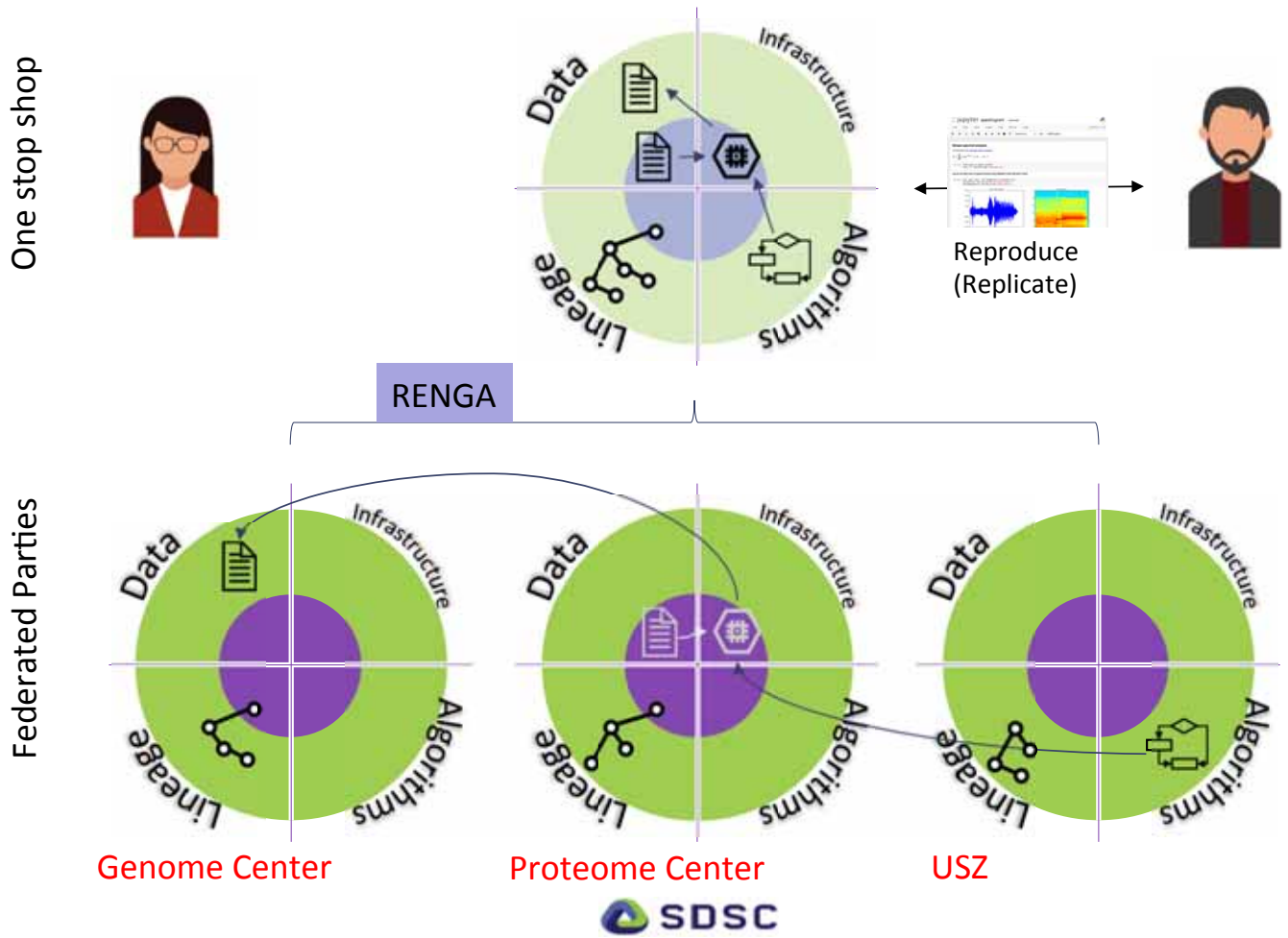
Use Case in Biomedical Research



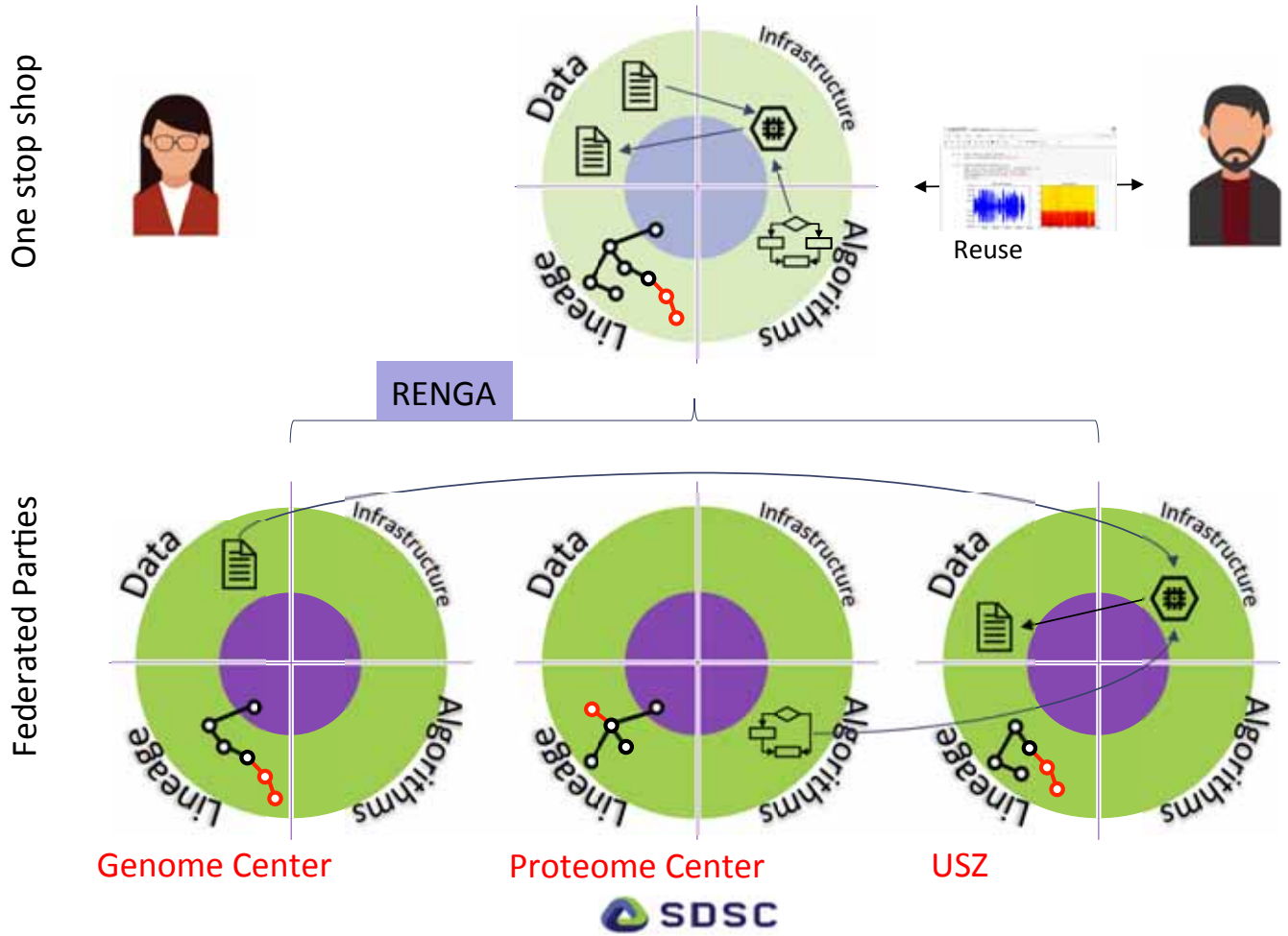
Use Case in Biomedical Research



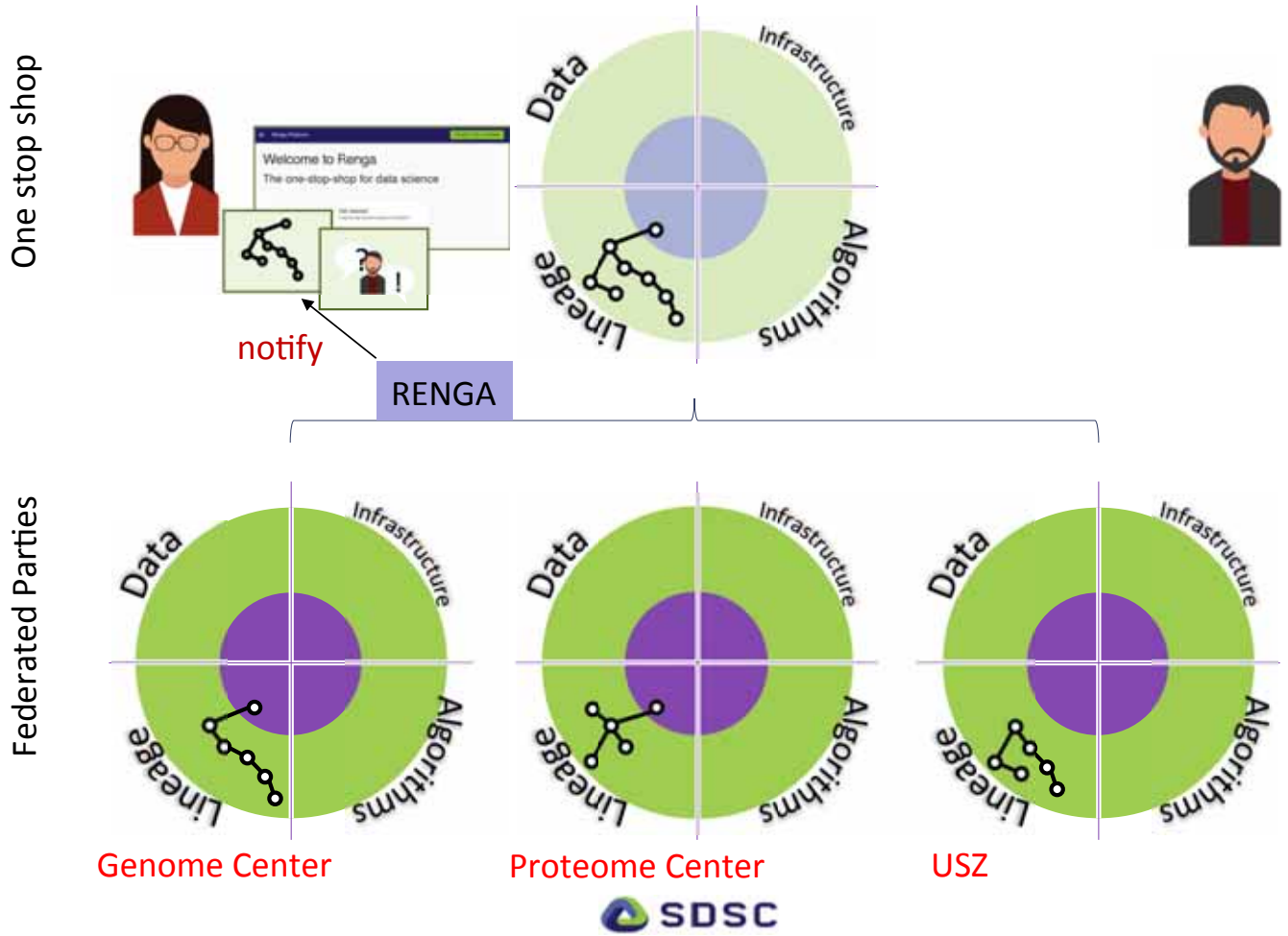
Use Case in Biomedical Research



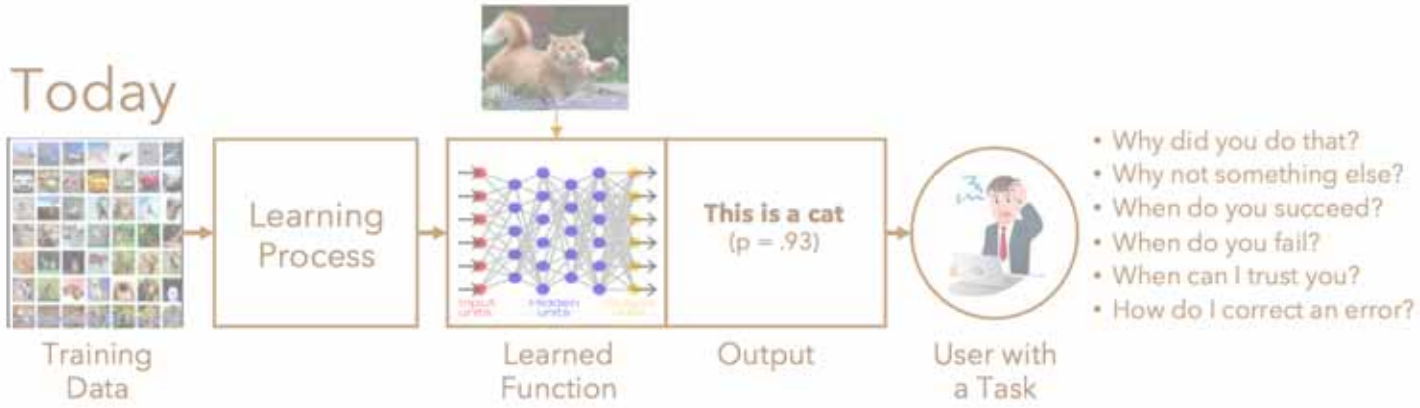
Use Case in Biomedical Research

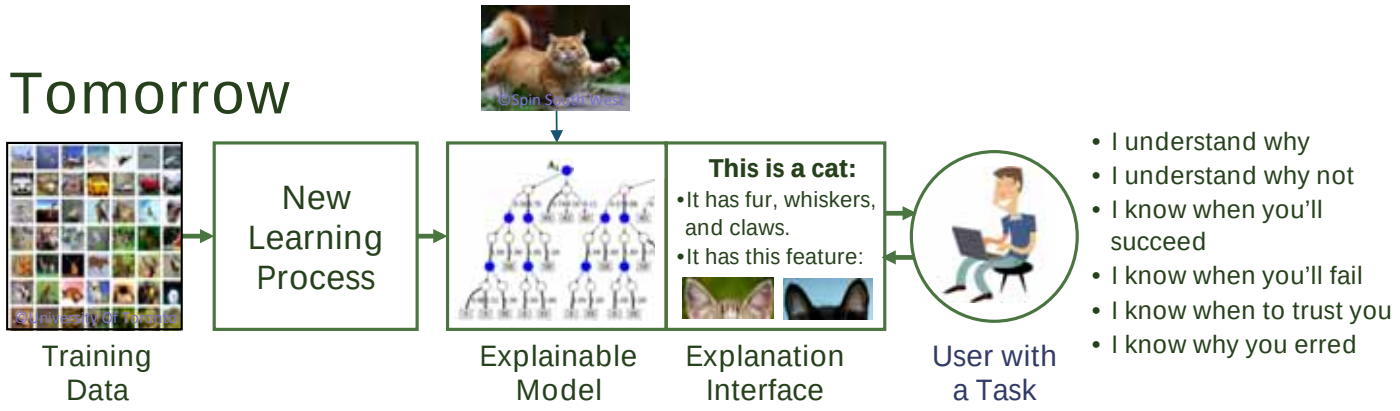


Use Case in Biomedical Research



Explainable AI – What Are We Trying To Do?





SDSC

THANK YOU!

http://www.datascience.ch

Twitter: @SDSCdatascience