



SWISS DATA SCIENCE CENTER

A DIGITALIZATION JOURNEY

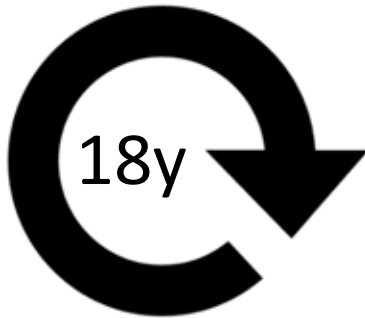
Prof. Edouard Bugnion

Co-Academic Director, Swiss Data Science Center

Professor & Vice-President, EPFL

Open Data in Sciences Workshop, Brussels 31.01.2018

About me



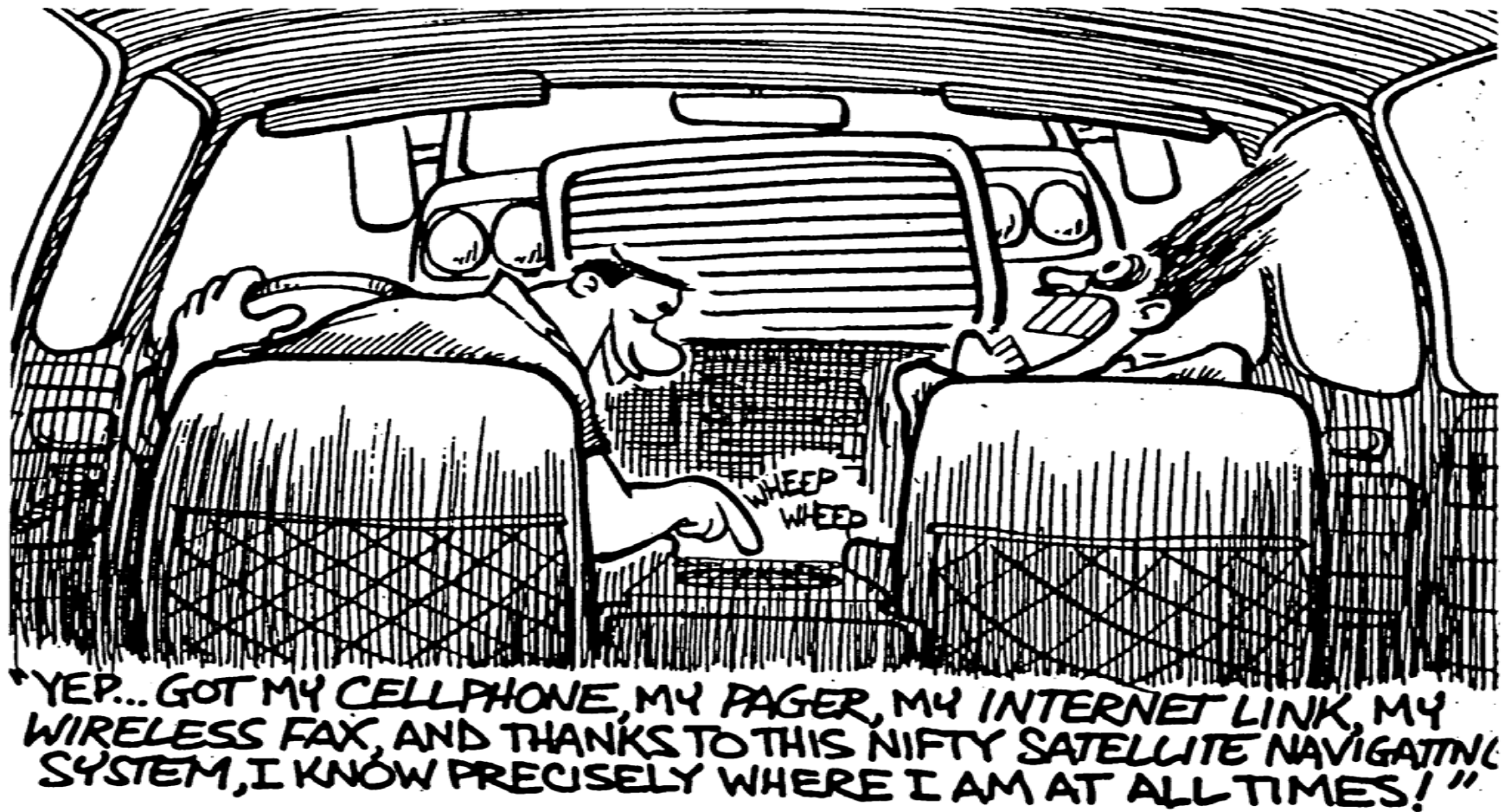
vmware[®]Industry



Academia

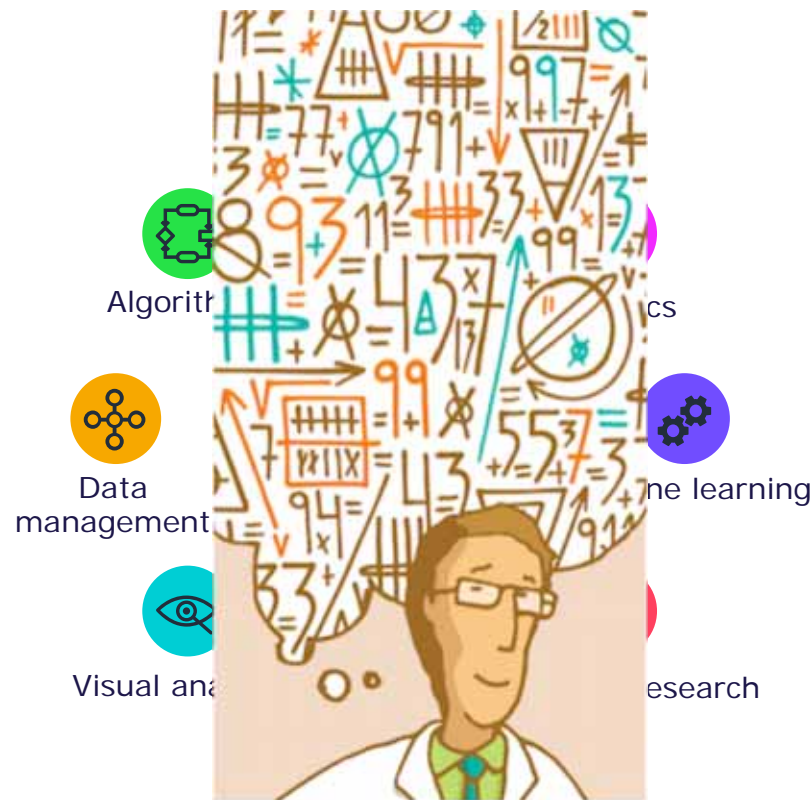


Big Data, Bad Data



BY LOWE FOR THE SUN-SENTINEL, FLOR

A fragmented ecosystem



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

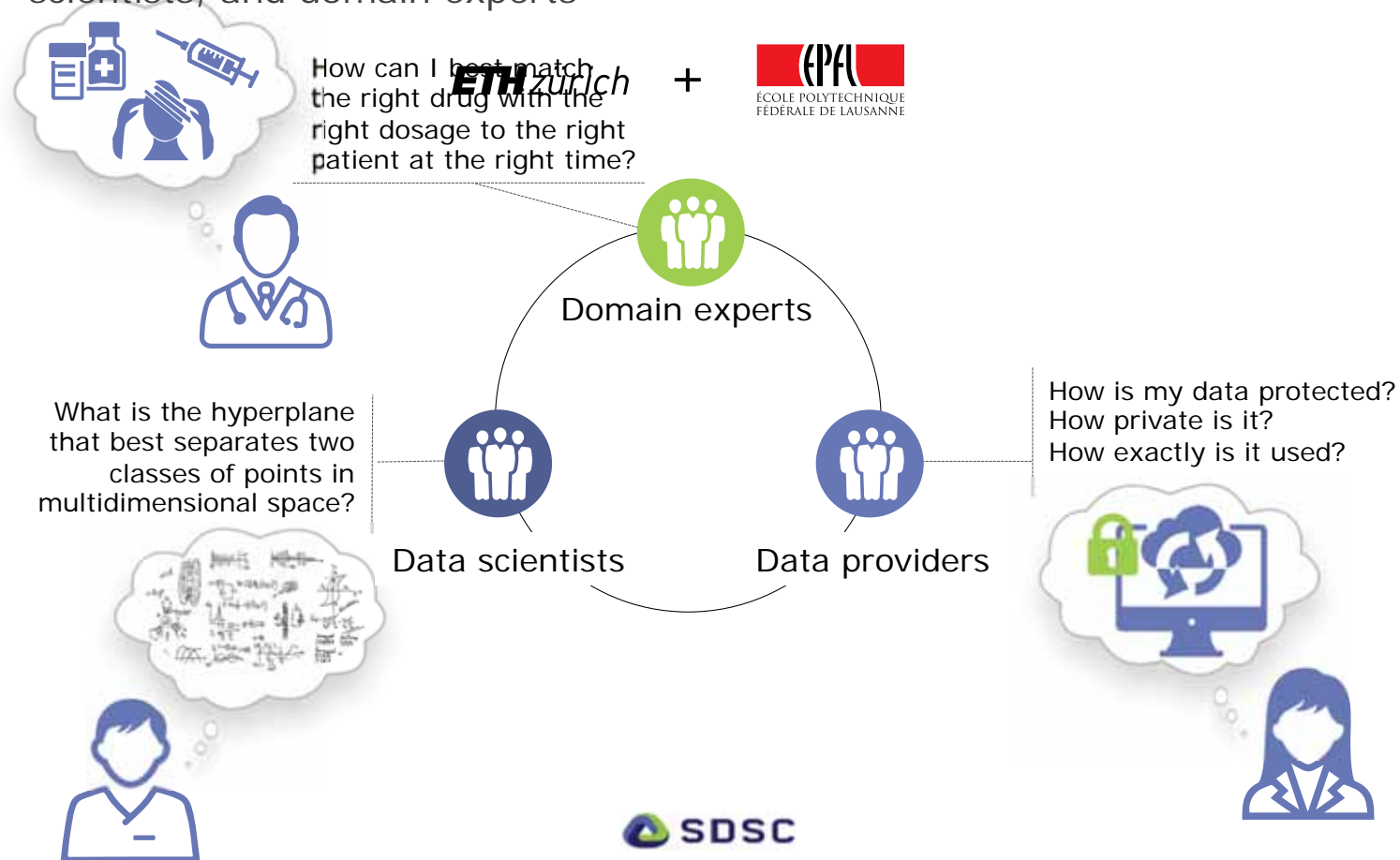
↔
GAP



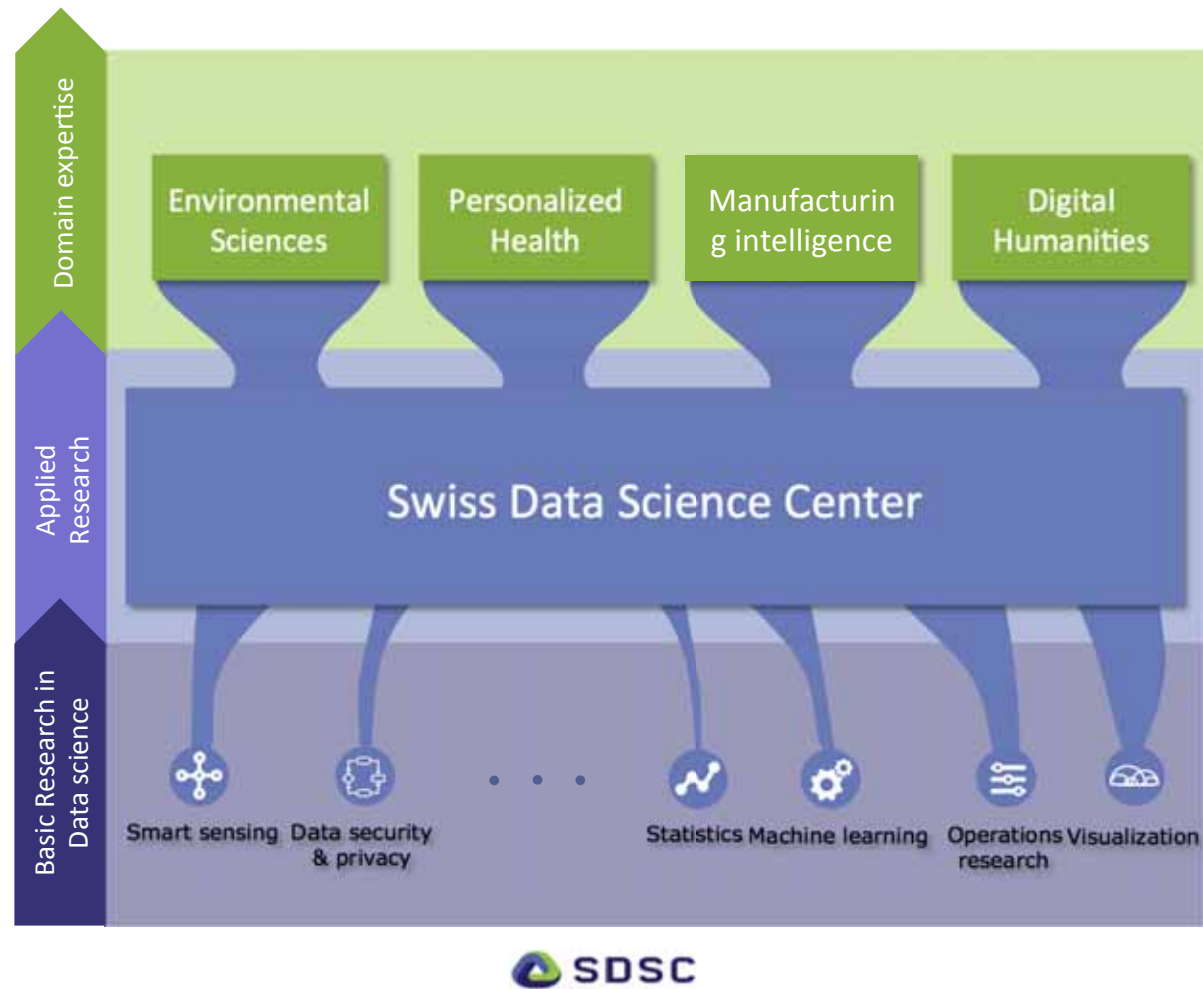
How can I best match the right drug with the right dosage to the right patient at the right time?

Swiss Data Science Center (SDSC)

Multi-disciplinary team of 40 from both 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.

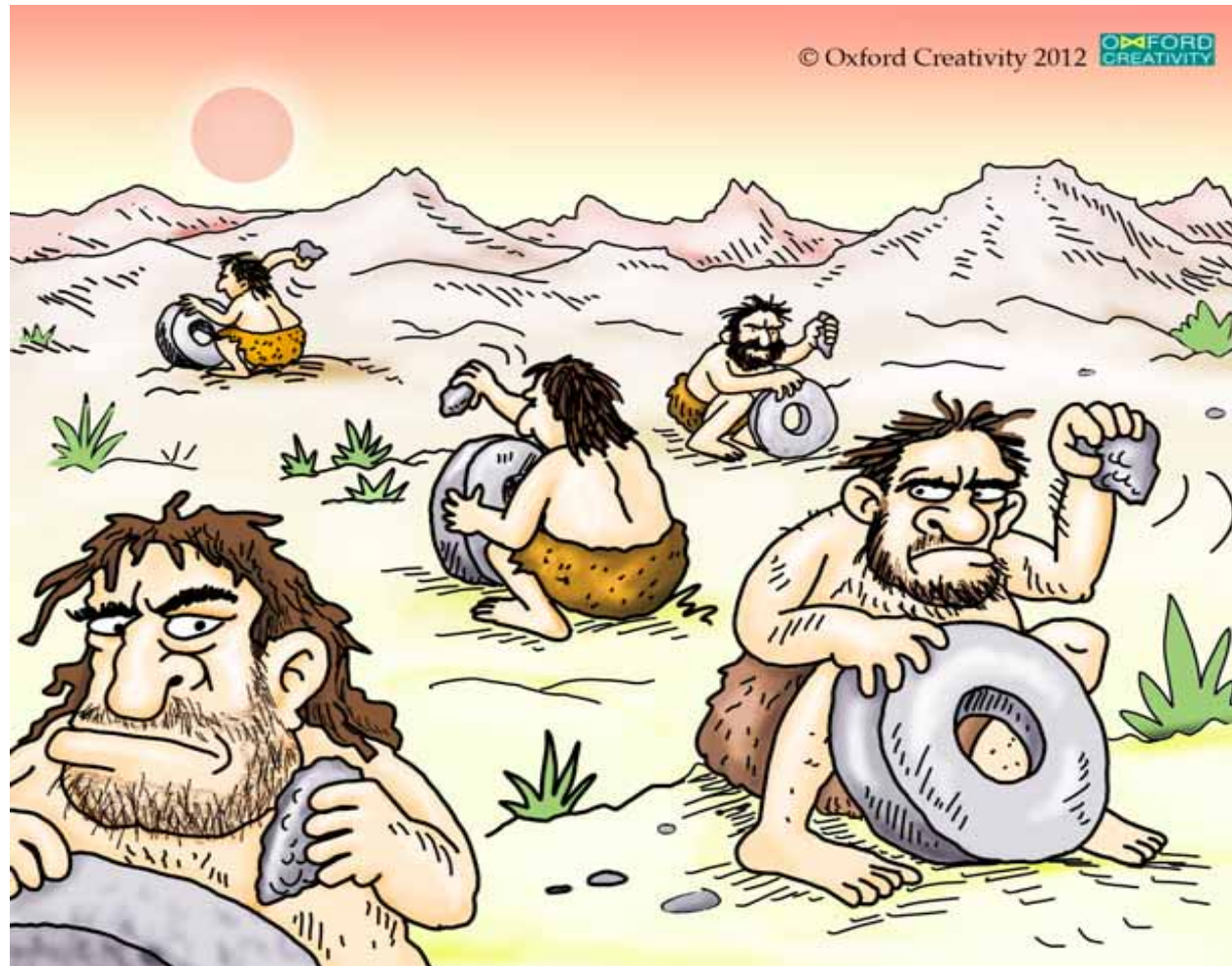


Open (Data) Science

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

SDSC Analytics Platform

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**

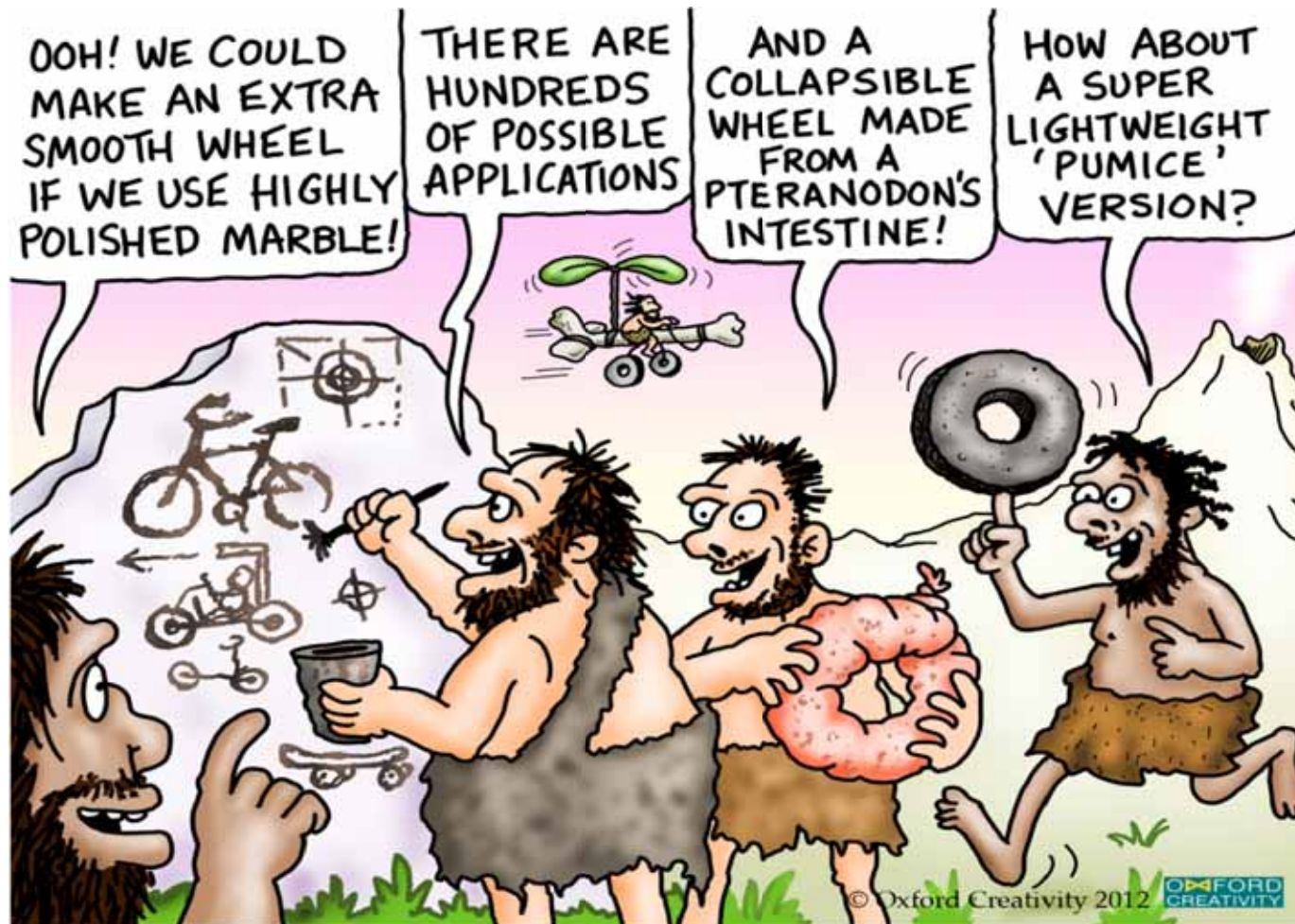


credit: oxford creativity, <https://www.triz.co.uk/>

Facilitate communication to foster innovation



Foster multidisciplinary collaborations



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

RENGA – SDSC Analytics Platform

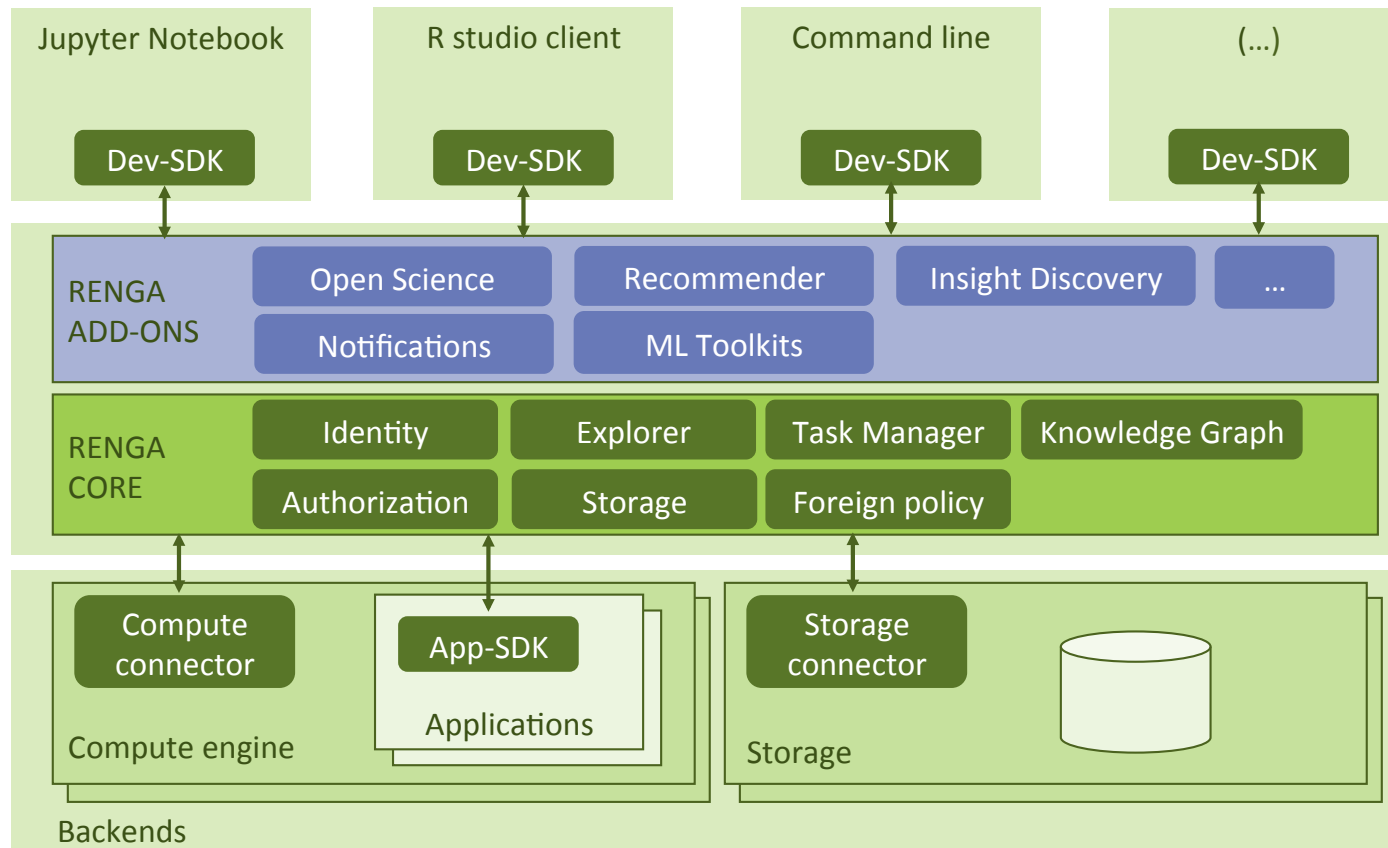
Renga (連歌), plural **renga**, a genre of Japanese linked-verse poetry in which two or more poets supply alternating sections of a poem linked by verbal and thematic associations.

—*Encyclopædia Britannica*

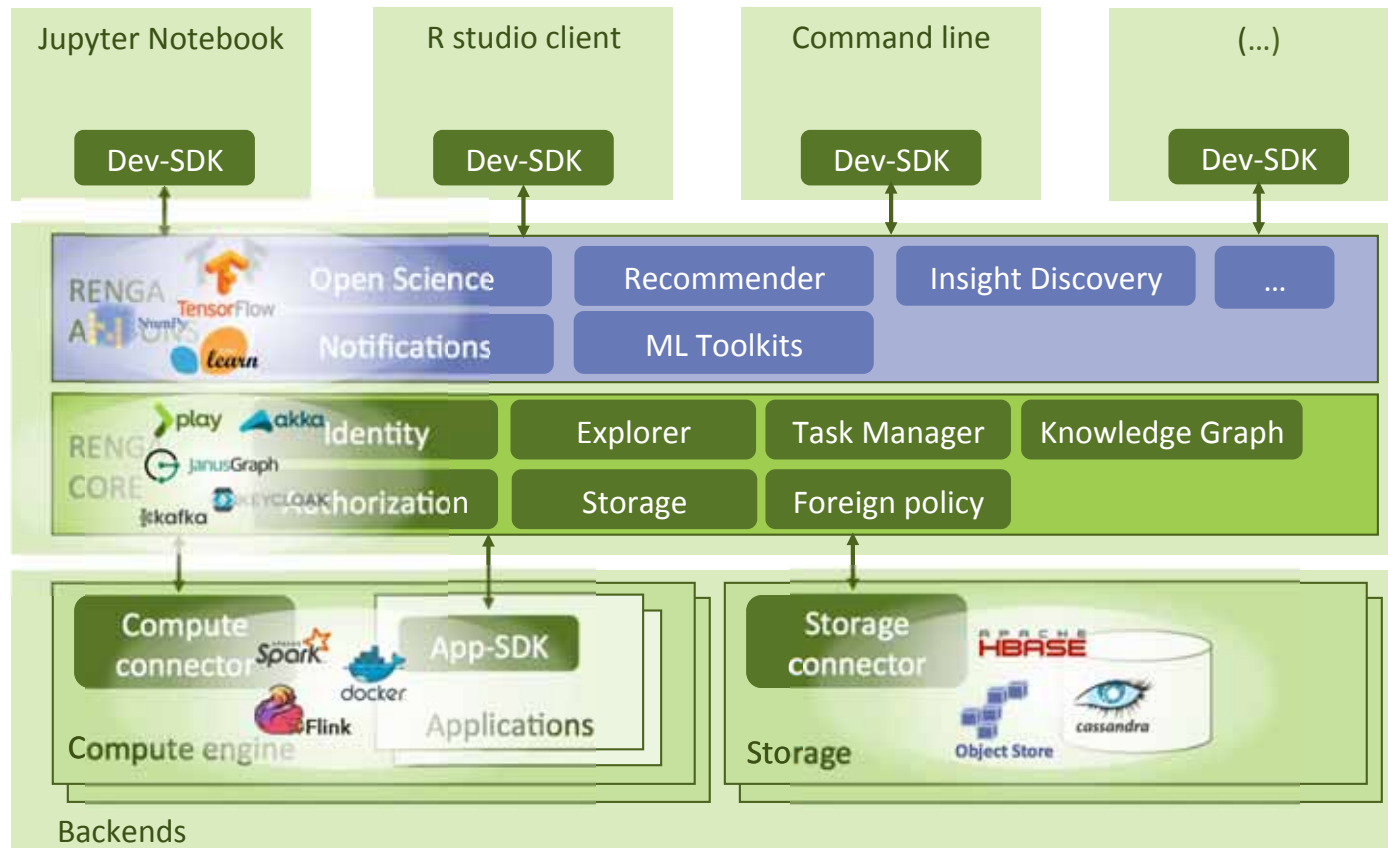
- Renga is a highly-scalable & secure open software platform designed to foster multidisciplinary (data) science collaborations across mutually untrusted academic / industrial institutions
- The platform allows practitioners to:
 - Securely manage, share, find and process large-scale data across untrusted 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



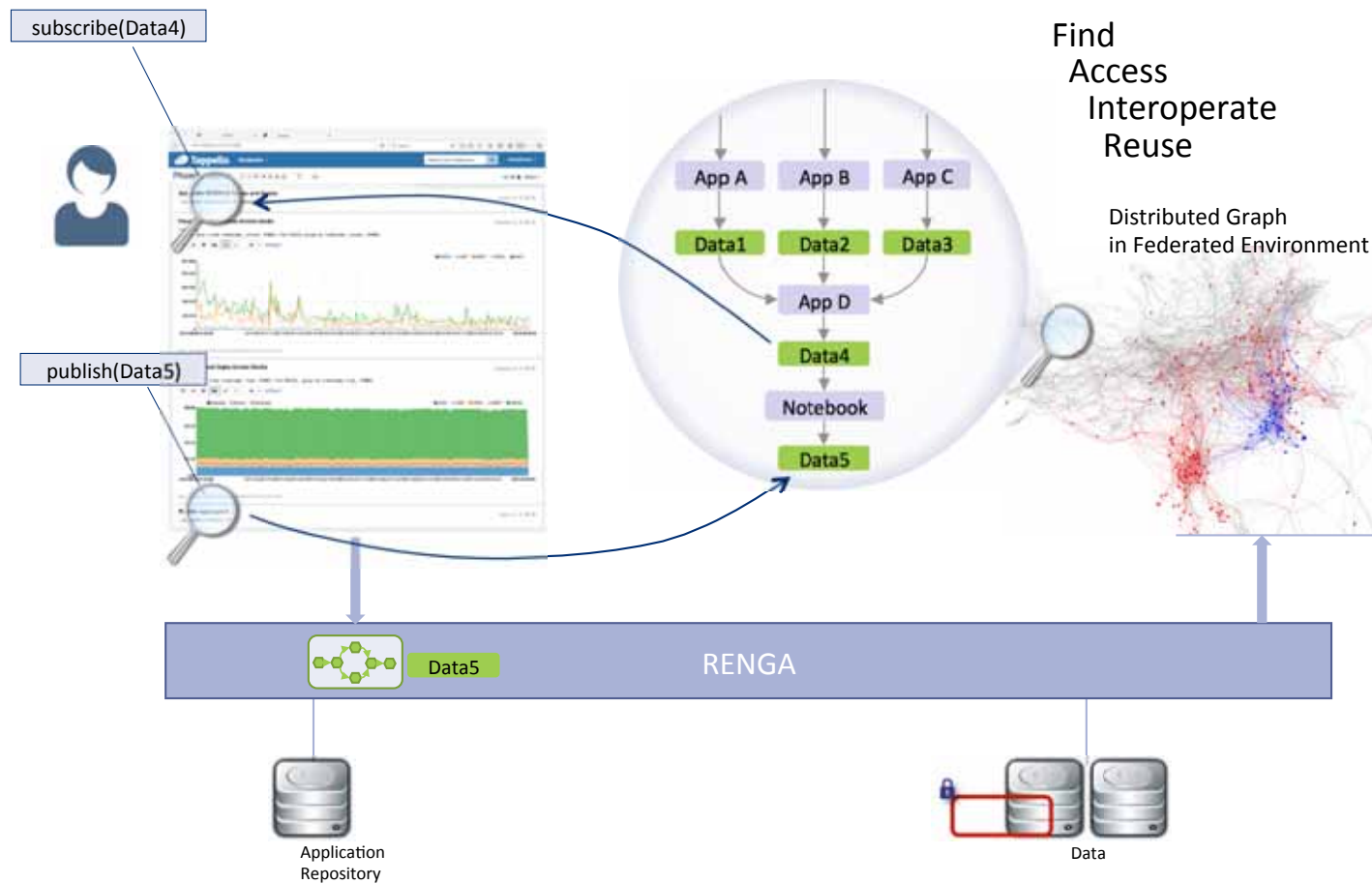
RENGA – Highly Modular Architecture



RENGA – Built on Proven Technology



Lineage in Data Science

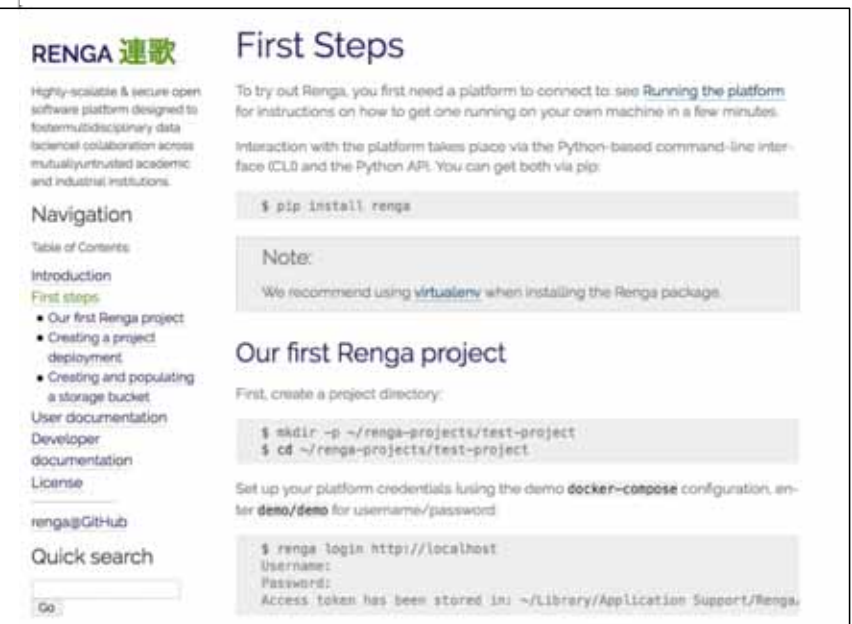


Available as Open Source (Apache v2)



The screenshot shows the Renga website homepage. At the top left is the SDSC logo. In the center is the Renga logo, a stylized 'A' made of two overlapping shapes, followed by the word 'RENGA' in a bold, sans-serif font. Below the logo is the tagline 'MULTIDISCIPLINARY DATA SCIENCE COLLABORATIONS MADE EASY'. There are two buttons: 'GETTING STARTED' and 'DOWNLOAD S.I.S.'. Below these buttons is a note: 'Preview beta 0.1.0 released on September 15, 2017'. Further down, there is a paragraph about Renga (連歌), plural renga, a genre of Japanese linked-verse poetry. Below this is another paragraph stating that Renga is a highly-scalable & secure open software platform designed to foster multidisciplinary data (science) collaborations. At the bottom, there are three diamond-shaped boxes with the following content:

- SECURE**: Securely manage, share and process large-scale data across untrusted parties operating in a federated environment.
- TRACEABLE**: Capture complete lineage automatically up to original raw data for detailed traceability (auditability & reproducibility).
- FAIR**: Work in a Findable, Accessible, Interoperable and Re-usable (FAIR) principles environment.



The screenshot shows the Renga documentation page. The title is 'RENGA 連歌'. The main heading is 'First Steps'. The text says: 'To try out Renga, you first need a platform to connect to: see [Running the platform](#) for instructions on how to get one running on your own machine in a few minutes. Interaction with the platform takes place via the Python-based command-line interface (CLI) and the Python API. You can get both via pip:

```
$ pip install renga
```

Below this is a 'Note' box: 'We recommend using [virtualenv](#) when installing the Renga package.'

The next section is 'Our first Renga project'. It says: 'First, create a project directory:

```
$ mkdir -p ~/renga-projects/test-project
$ cd ~/renga-projects/test-project
```

Then it says: 'Set up your platform credentials using the demo [docker-compose](#) configuration, enter [demo/demo](#) for username/password

```
$ renga login http://localhost
Username:
Password:
Access token has been stored in: ~/Library/Application Support/Renga.
```

On the left side of the documentation page, there is a 'Navigation' section with a 'Table of Contents' link. Below that is an 'Introduction' section with a 'First steps' link. The 'First steps' link is highlighted. Below it is a list of links: 'Our first Renga project', 'Creating a project deployment', 'Creating and populating a storage bucket', 'User documentation', 'Developer documentation', 'License', 'renga@GitHub', and 'Quick search'.

<http://get-renga.io>

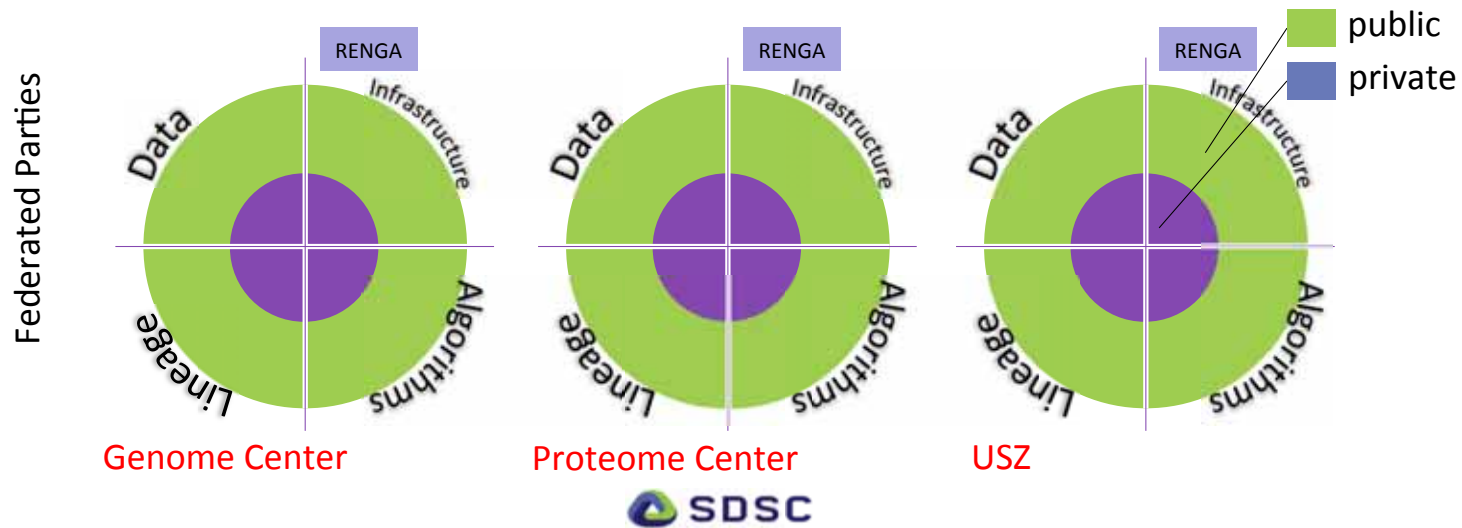




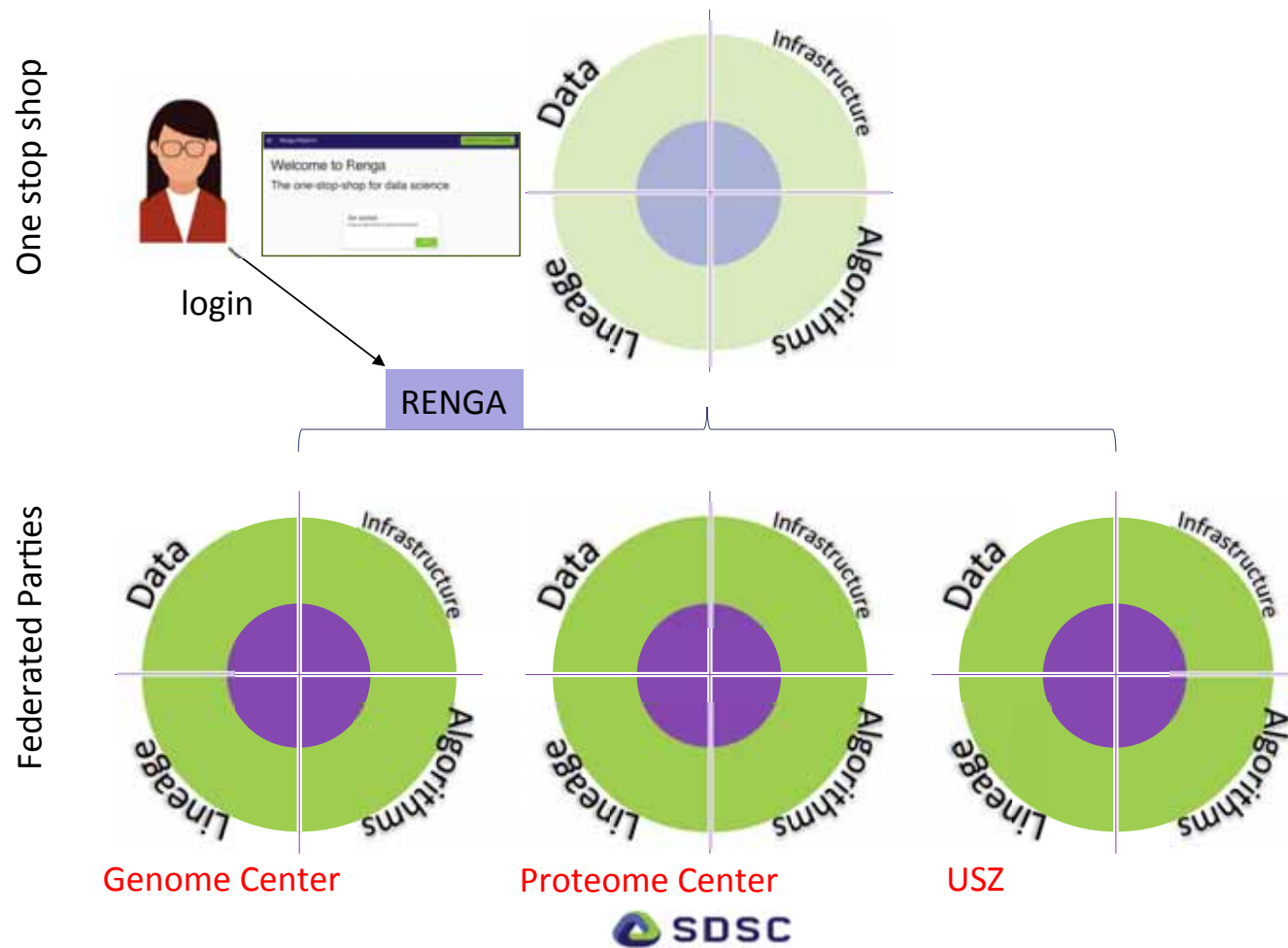
Roamap – FEDERATED RENGAs

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

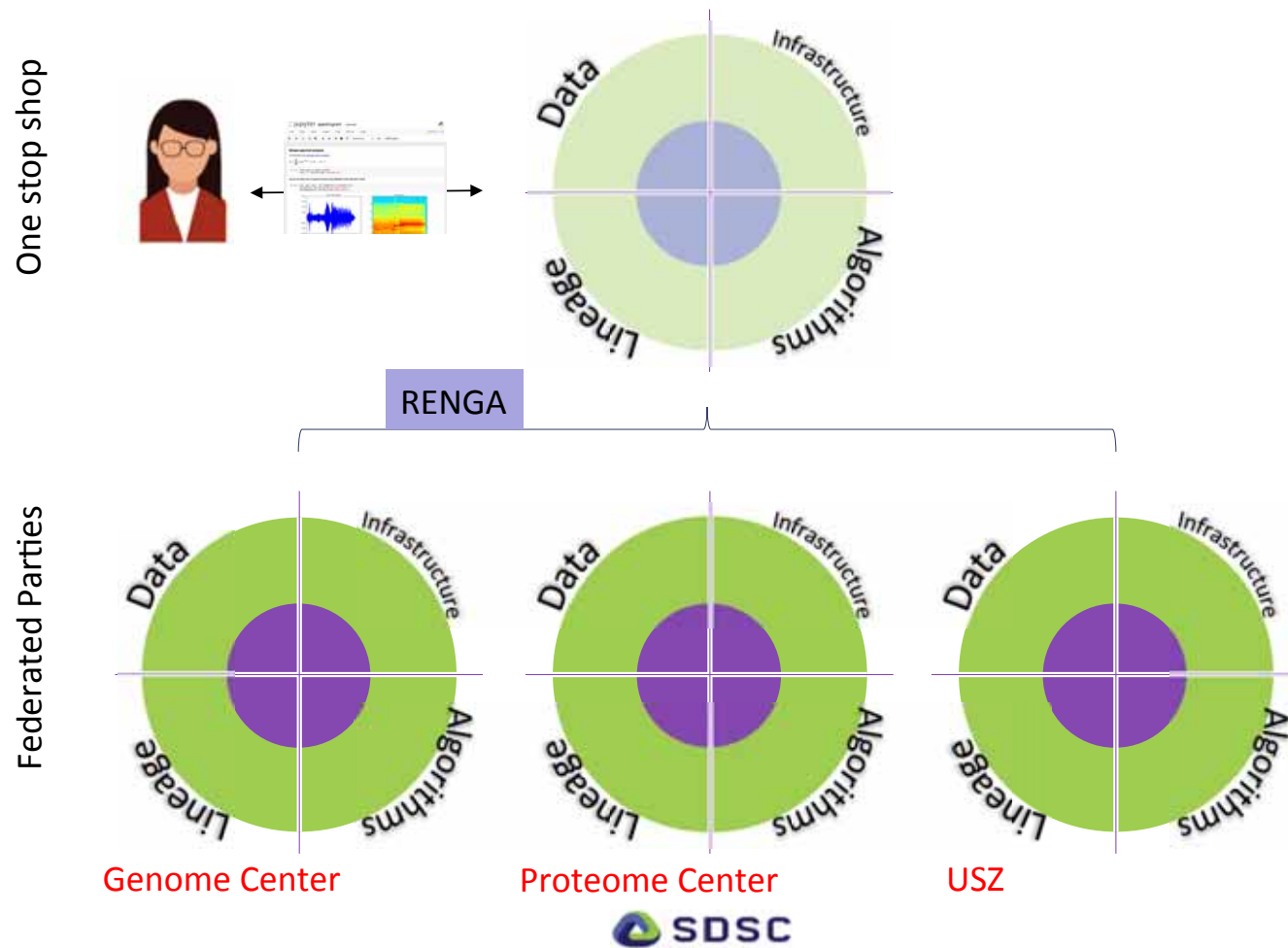
Use Case in Biomedical Research



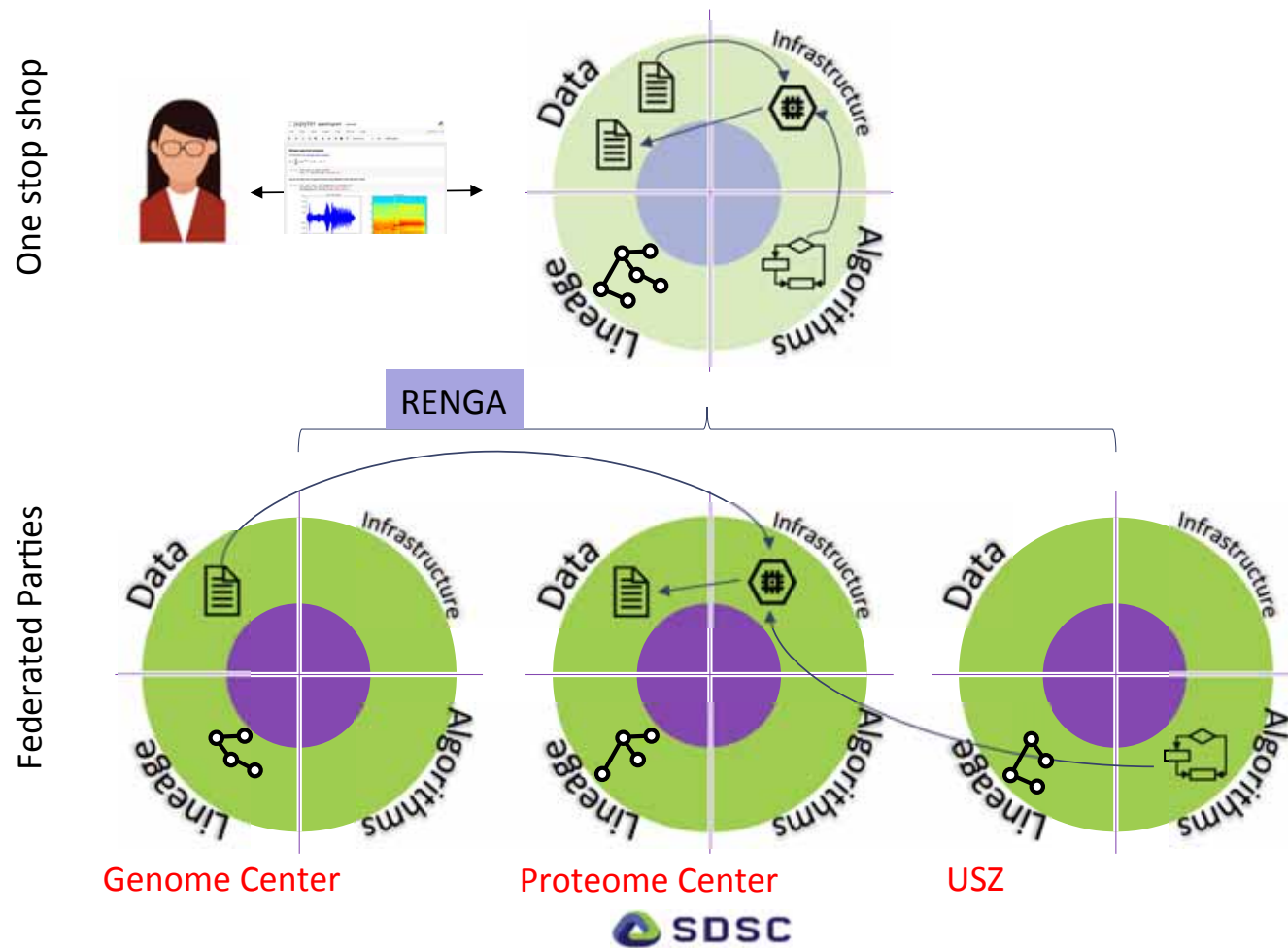
Use Case in Biomedical Research



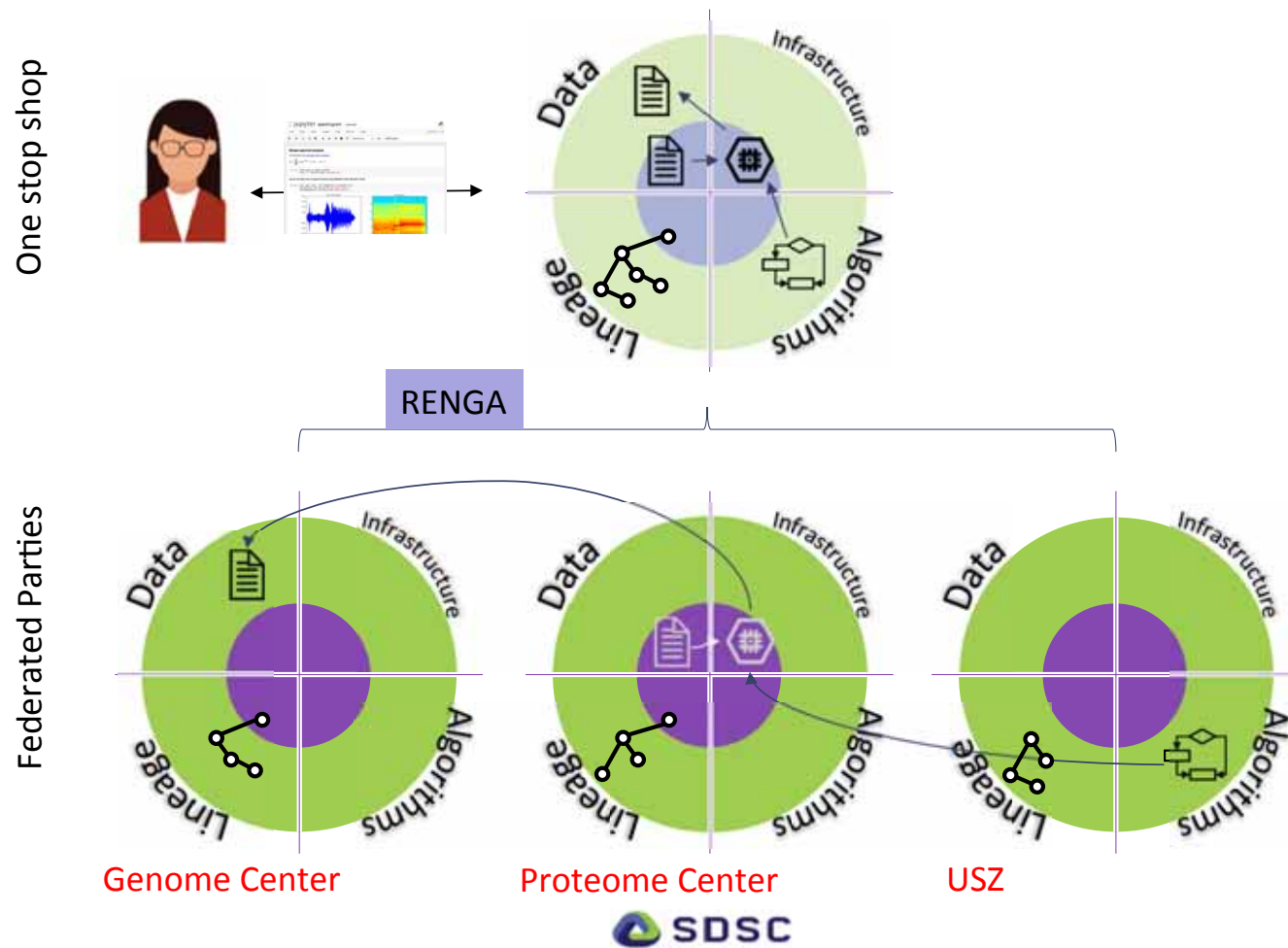
Use Case in Biomedical Research



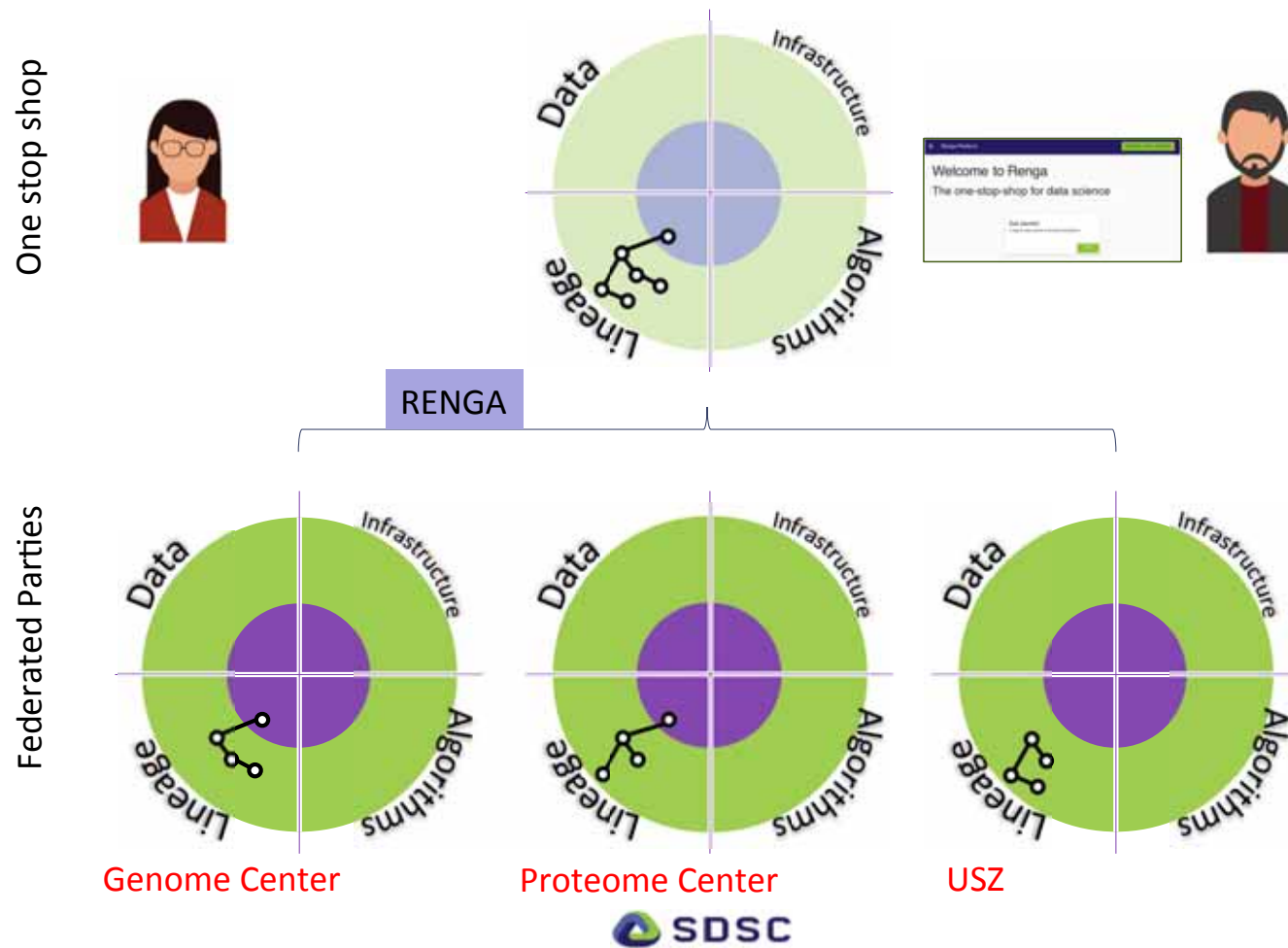
Use Case in Biomedical Research



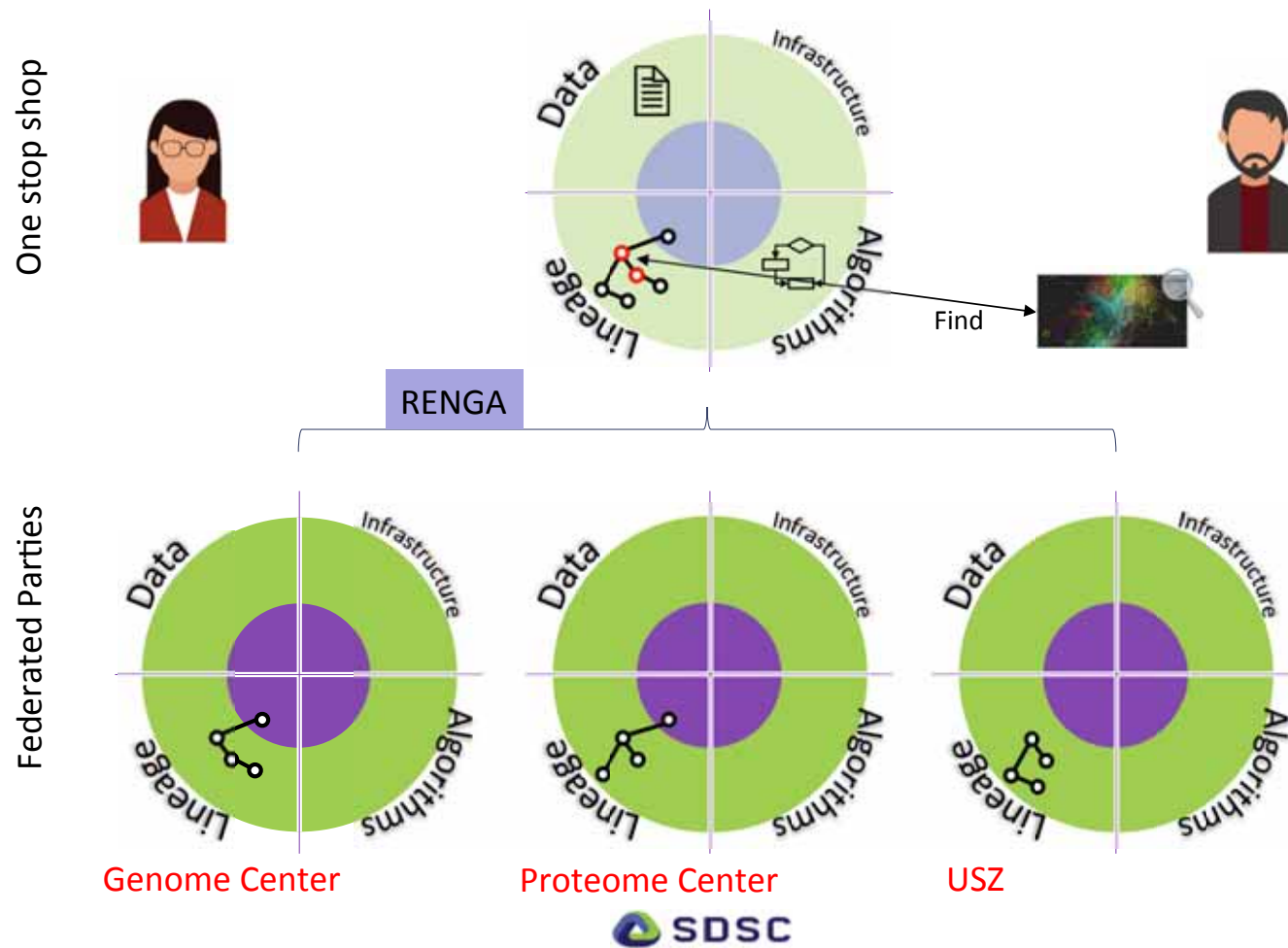
Use Case in Biomedical Research



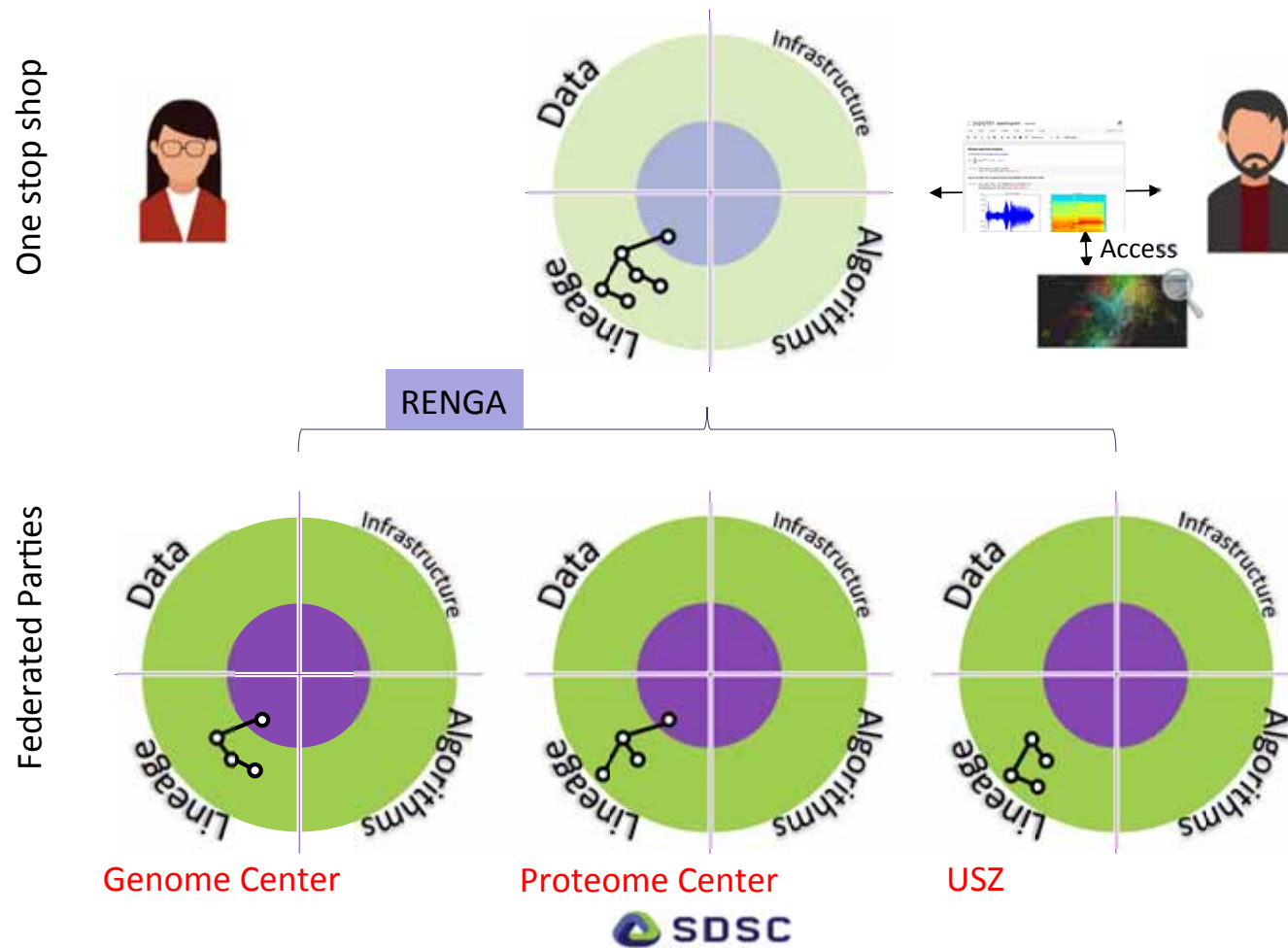
Use Case in Biomedical Research



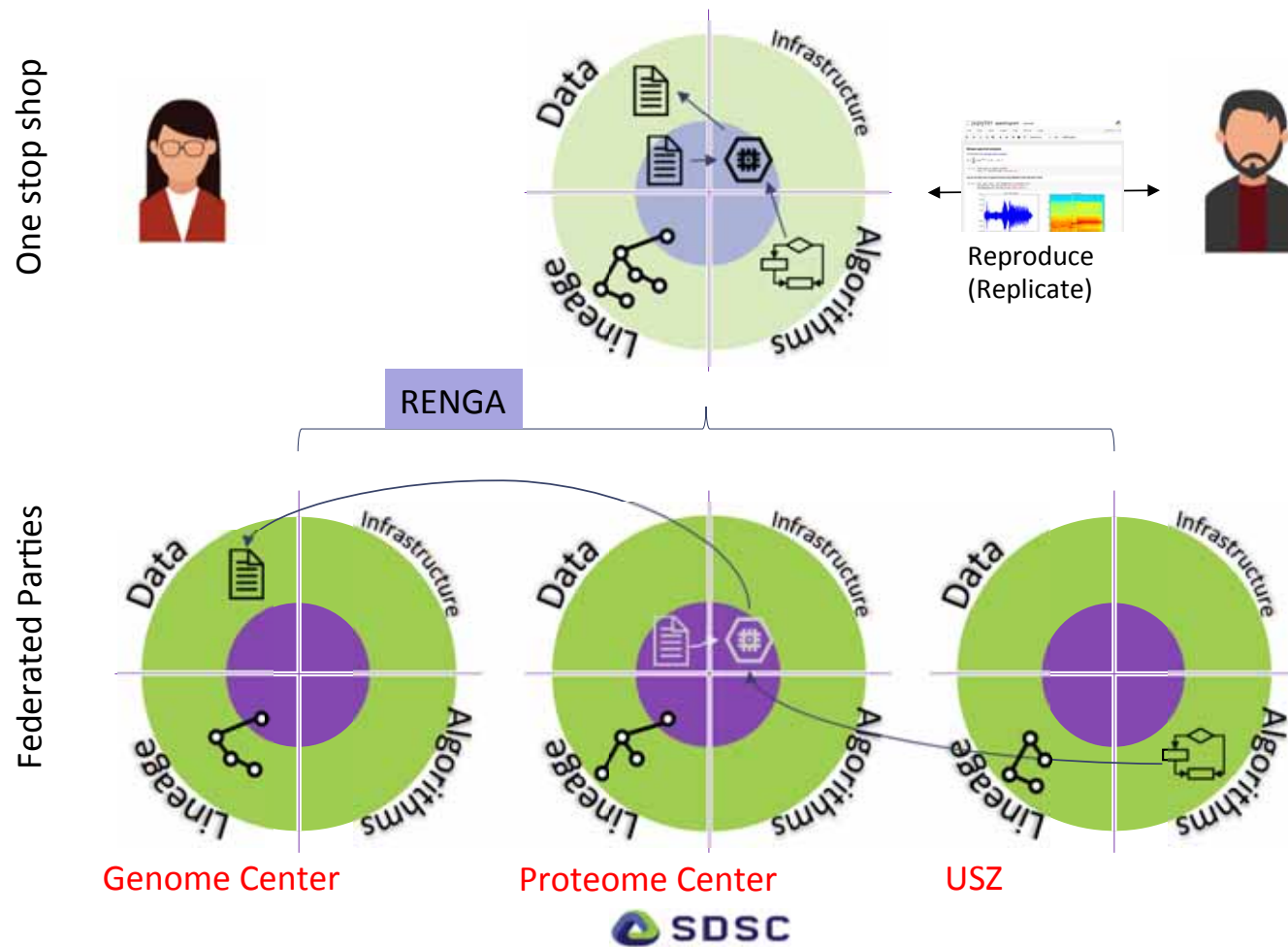
Use Case in Biomedical Research



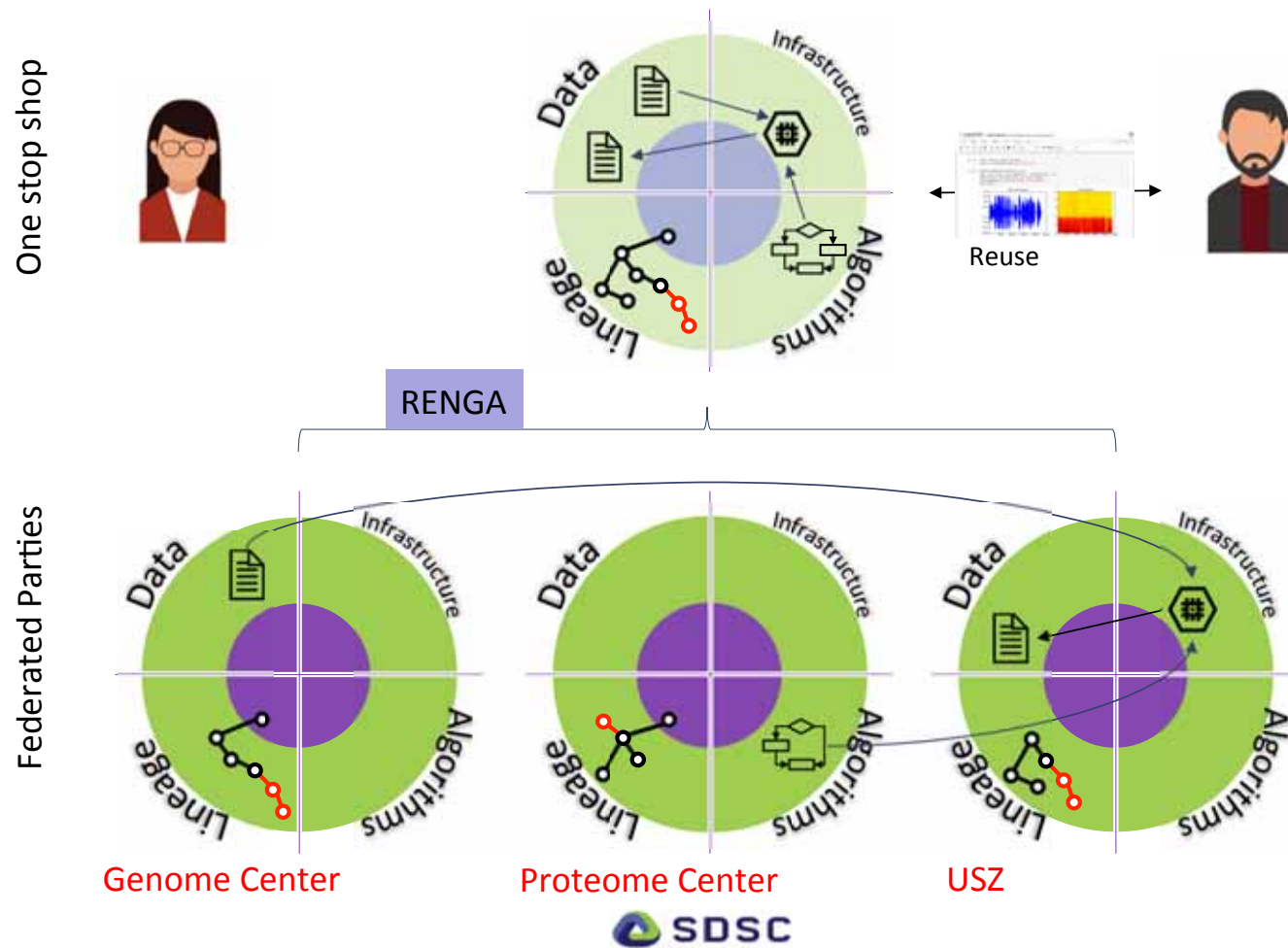
Use Case in Biomedical Research



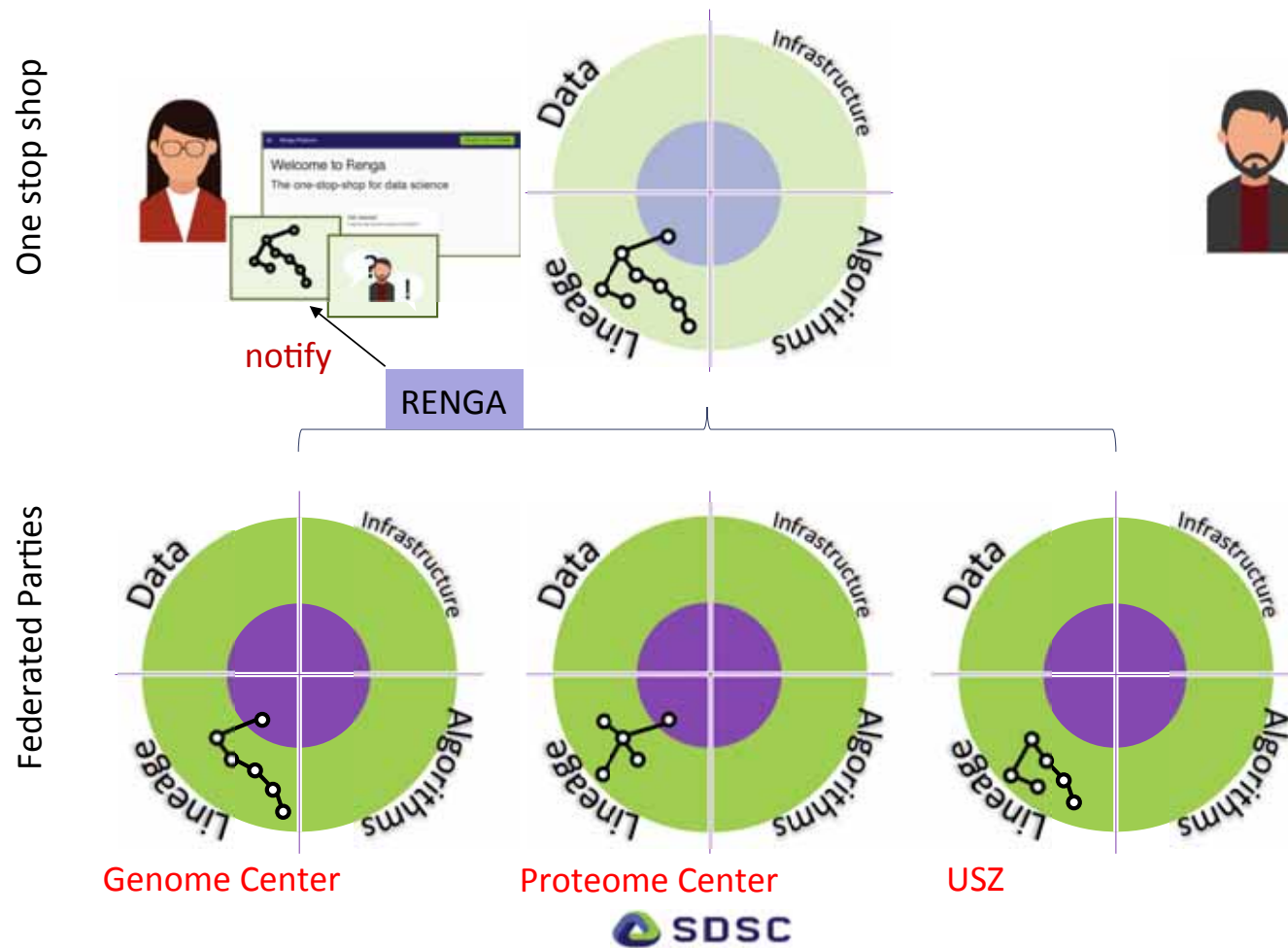
Use Case in Biomedical Research



Use Case in Biomedical Research

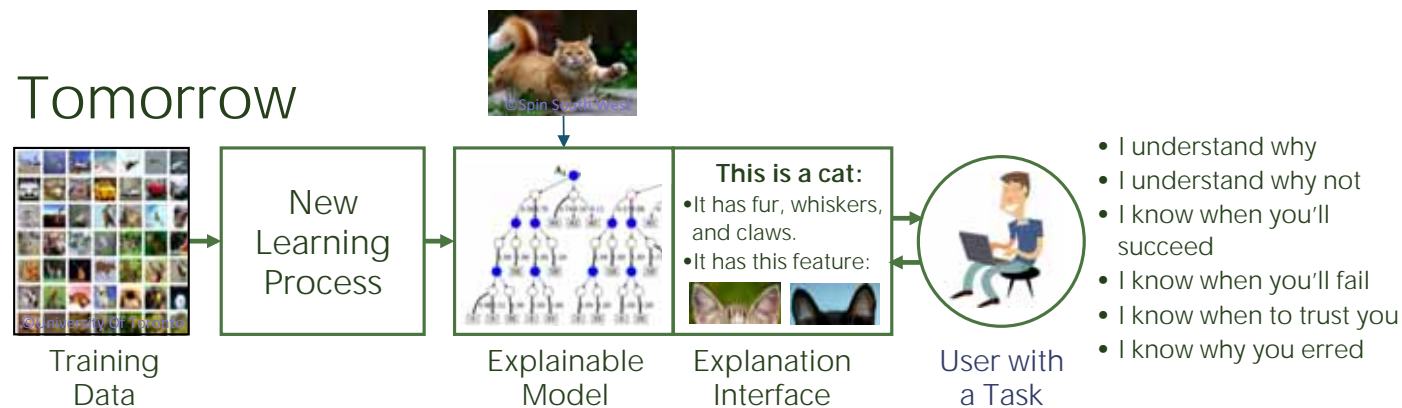
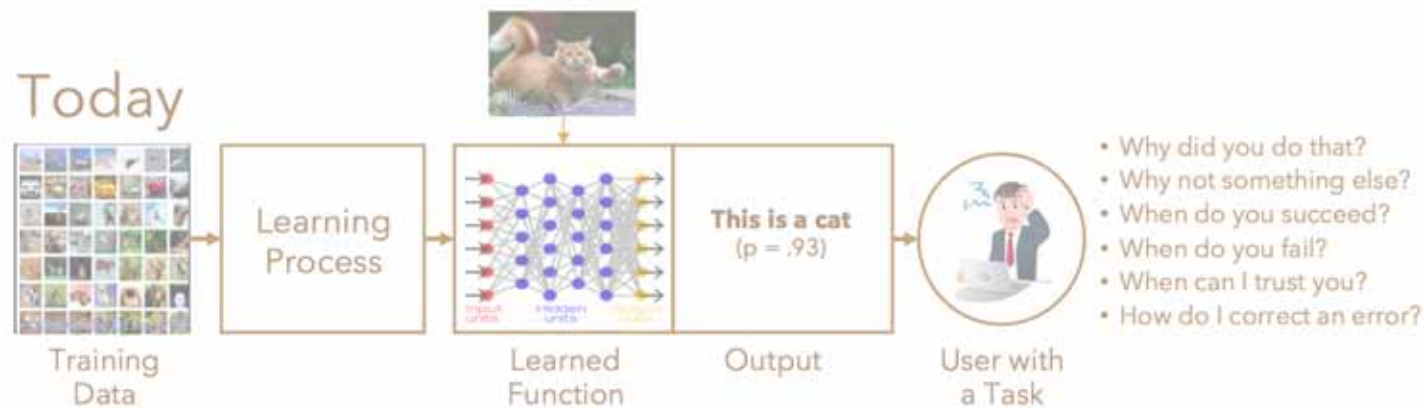


Use Case in Biomedical Research





Explainable AI – What Are We Trying To Do?





THANK YOU!

<http://www.datascience.ch>

Twitter: @SDSCdatascience