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

14y

Academia
“YEP... GOT MY CELLPHONE, MY PAGER, MY INTERNET LINK, MY WIRELESS FAX, AND THANKS TO THIS NIFTY SATELLITE NAVIGATION SYSTEM, I KNOW PRECISELY WHERE I AM AT ALL TIMES!”

BY LOWE FOR THE SUN-SENTINEL. FLO
A fragmented ecosystem

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

How can I best match the right drug with the right dosage to the right patient at the right time?
How is my data protected?
How private is it?
How exactly is it used?

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

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

Foster adoption of data science both in academia and industry

Data scientists

Domain experts

Data providers

Swiss Data Science Center (SDSC)
Where does SDSC fit?

- Environmental Sciences
- Personalized Health
- Manufacturing intelligence
- Digital Humanities

- Smart sensing
- Data security & privacy
- Statistics
- Machine learning
- Operations
- Visualization

Swiss Data Science Center
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
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
Facilitate communication to foster innovation
Foster multidisciplinary collaborations
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

- Jupyter Notebook
- R studio client
- Command line
- (…)

RENGA ADD-ONS
- Open Science
- Notifications
- Recommender
- ML Toolkits
- Insight Discovery
- …

RENGA CORE
- Identity
- Explorer
- Task Manager
- Knowledge Graph
- Authorization
- Storage
- Foreign policy

Compute connector
- Compute engine
  - Applications

App-SDK
- Storage connector
- Storage

Backends

SDSC
RENGA – Built on Proven Technology
Lineage in Data Science
Available as Open Source (Apache v2)

http://get-renga.io
Roamap – FEDERATED RENGA

- Open (Data) Science in an privacy-conscious world
- Controlled data sharing
Use Case in Biomedical Research

Federated Parties

Genome Center

Proteome Center

USZ

SDSC
Use Case in Biomedical Research

One stop shop

login

RENGA

Federated Parties

Genome Center

Proteome Center

USZ
Use Case in Biomedical Research

One stop shop

Federated Parties

- Genome Center
- Proteome Center
- USZ

RENGA

SDSC
Use Case in Biomedical Research
Use Case in Biomedical Research
Use Case in Biomedical Research

One stop shop

Federated Parties

Genome Center

Proteome Center

USZ

RENGA
Use Case in Biomedical Research
Use Case in Biomedical Research

One stop shop

Federated Parties

Genome Center

Proteome Center

USZ

SDSC
Use Case in Biomedical Research

One stop shop

Federated Parties

Genome Center

Proteome Center

USZ

RENGA
Use Case in Biomedical Research
Use Case in Biomedical Research

One stop shop

notify

RENGA

Federated Parties

Genome Center

Proteome Center

USZ

SDSC
Explainable AI – What Are We Trying To Do?

Today

- Training Data → Learning Process → Learned Function → Output → User with a Task

• Why did you do that?
• Why not something else?
• When do you succeed?
• When do you fail?
• When can I trust you?
• How do I correct an error?

Tomorrow

- Training Data → New Learning Process → Explainable Model → Explanation Interface → User with a Task

• I understand why
• I understand why not
• I know when you’ll succeed
• I know when you’ll fail
• I know when to trust you
• I know why you erred

This is a cat:
- It has fur, whiskers, and claws.
- It has this feature:
THANK YOU!

http://www.datascience.ch
Twitter: @SDCdataScience