The Swiss Data Science Center (SDSC), a joint-venture between ETH Zurich and EPFL

SDSC team (Stan Schymanski, Eric Bouillet, Olivier Verscheure *et al*.)

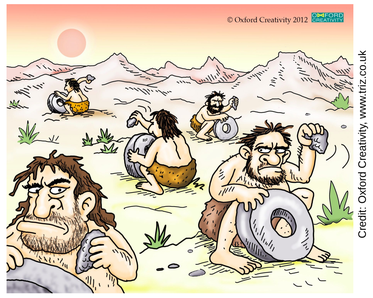
Swiss Data Science Center (SDSC), Swiss Federal Institute of Technology (ETH), Zurich & EPFL, Lausanne, Switzerland.
stan.schymanski@datascience.ch

1 Motivation and objectives

(a) *Status Quo*

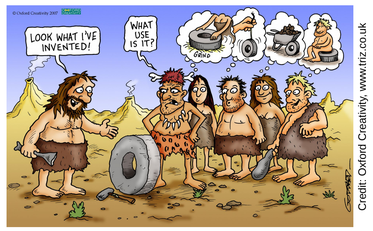
1. Growing disconnect between data scientists, domain experts and data providers
2. Research not always verifiable
3. Methods and data in publications not often reusable

Scientific progress hampered by frequent re-inventions of the wheel



(b) Objectives

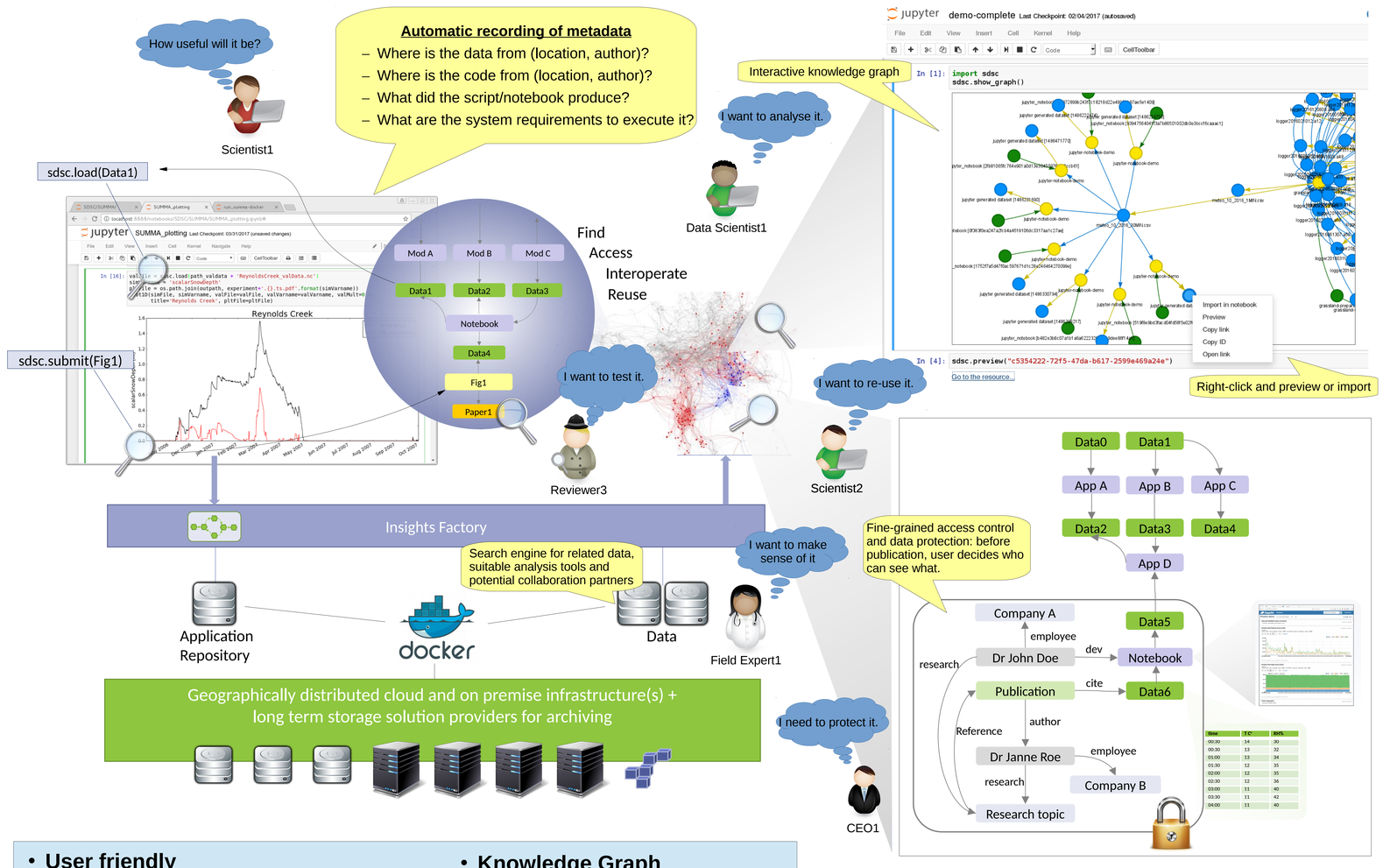
- Create scalable open-science framework for handling and tracking of data and algorithms, from raw data and first principle equations to final data products & visualizations
- Offer user-friendly tools and services, fostering intra and inter-domain collaboration, productivity and excellence
- Facilitate reproduction, reuse and impact tracking of data and algorithms



2 Reproducible and re-usable scientific workflow

(1) Download and convert data from external resource (logger, repository, etc.): `sdsc.load(Data1)`
(2) Import any third-party code: `sdsc.import(Mod1)`
(3) Interactive development, analysis and documentation in jupyter notebook, using your preferred analysis tool (e.g. python, R, matlab...)
(4) Create publication-quality figures
(5) Create docker containers for new components: `sdsc.dockerize(components)`
(6) Submit to sdsc insights factory: `sdsc.submit(project1)`
(7) Control access rights: `sdsc.share(project1, reviewers)`
(8) Submit paper to scientific journal and link to project: `sdsc.addlink(project1, paper1)`
(9) Publish project components: `sdsc.publish(project1, target=zenodo.org)`

3 A platform made for and developed by the scientific community



- **User friendly**
  - Work with your preferred tools (python, R, matlab, ...)
  - Automated version control and tracking
- **Reproducible and re-usable research**
  - See the (versioned) algorithms and data
  - Replay, clone and modify a workflow using dockerized components
- **Knowledge Graph**
  - Data popularity, H-index
  - Who is using the data?
  - For what?
- **IP Protection**
  - Decide what, when and who to share with
  - Track and share as you go, publish when ready

We would like to hear YOUR thoughts!

Stay tuned and follow us online:

www.datascience.ch
contact@datascience.ch
@SDSCdatascience