BALÁZS KÉGL
LAL / CNRS

DATA SCIENCE

Design of automated methods

to analyze massive and complex data

to extract knowledge

Paris-Saclay Center for Data Science

A multi-disciplinary initiative, building interfaces, matching people, helping them launching projects

345 affiliated **researchers**, 115 active, 29 in COPIL
17 of the 19 **partners**, 50 **laboratories**
2 core **departments** (STIC, Math), 4 strongly present (SPU, SDV, P2I, HSS)

Biology & bioinformatics
IBISC/UEvry
LRI/UPSud
Hepatinov
CESP/UPSud-UVSQ-Inserm
IGM-I2BC/UPSud
MIA/Agro
MIAj-MIG/INRA
LMAS/Centrale

Chemistry
EA4041/UPSud

Earth sciences
LATMOS/UVSQ
GEOPS/UPSud
IPSL/UVSQ
LSCE/UVSQ
LMD/Polytechnique

Economy
LM/ENSAE
RITM/UPSud
LFA/ENSAE

Neuroscience
UNICOG/Inserm
U1000/Inserm
NeuroSpin/CEA

Particle physics astrophysics & cosmology
LPP/Polytechnique
DMPH/ONERA
CosmoStat/CEA
IAS/UPSud
AIM/CEA
LAL/UPSud

Machine learning
LRI/UPSud
LTCI/Telecom
CMLA/Cachan
LS/ENSAE
LIX/Polytechnique
MIA/Agro
CMA/Polytechnique
LSS/Supélec
CVN/Centrale
LMAS/Centrale
DTIM/ONERA
IBISC/UEvry
LIST/
Visualization
INRIA
LIMSI

Signal processing
LTCI/Telecom
CMA/Polytechnique
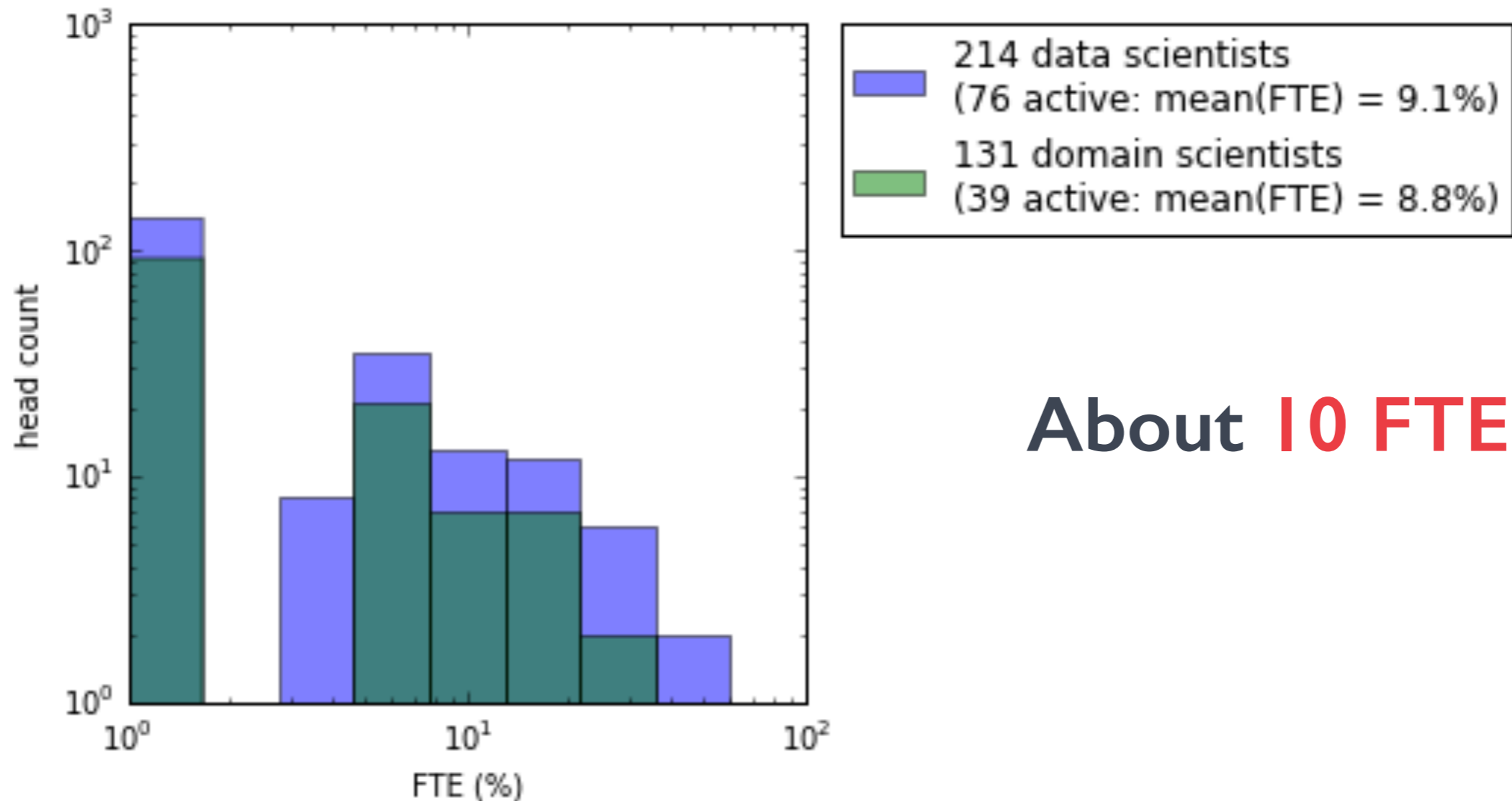CVN/Centrale
LSS/Supélec
CMLA/Cachan
LIMSI
DTIM/ONERA

Statistics
LMO/UPSud
LS/ENSAE
LSS/Supélec
CMA/Polytechnique
LMAS/Centrale
MIA/AgroParisTech

Paris-Saclay Center for Data Science

345 affiliated **researchers**, 115 active, 29 in COPIL

17 of the 19 **partners**, 50 **laboratories**

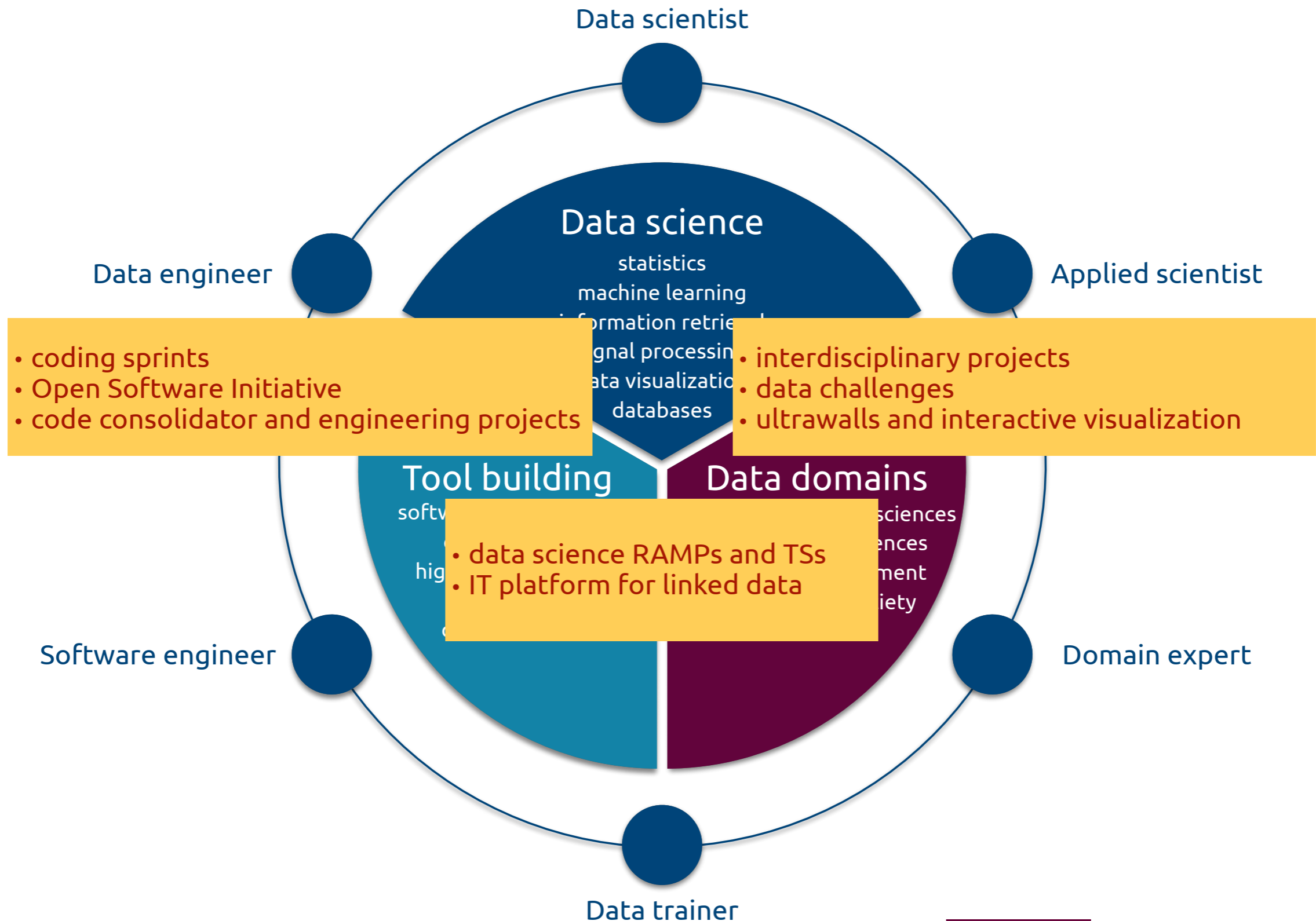2 core **departments** (STIC, Math), 4 strongly present (SPU, SDV, P2I, HSS)

214 data scientists (76 active: mean(FTE) = 9.1%)

131 domain scientists (39 active: mean(FTE) = 8.8%)

About **10 FTEs**

CHALLENGES

- (The lack of) manpower

  - especially at the interfaces

  - industrial brain-drain

- Incentives

  - data scientists are not incentivized to work on domain science

  - scientists are not incentivized to work on tools

- Access

  - no well-developed channels to **identify the right experts** for a given problem

- Tools

  - few **tools** that can help domain scientists and data scientists to **collaborate efficiently**
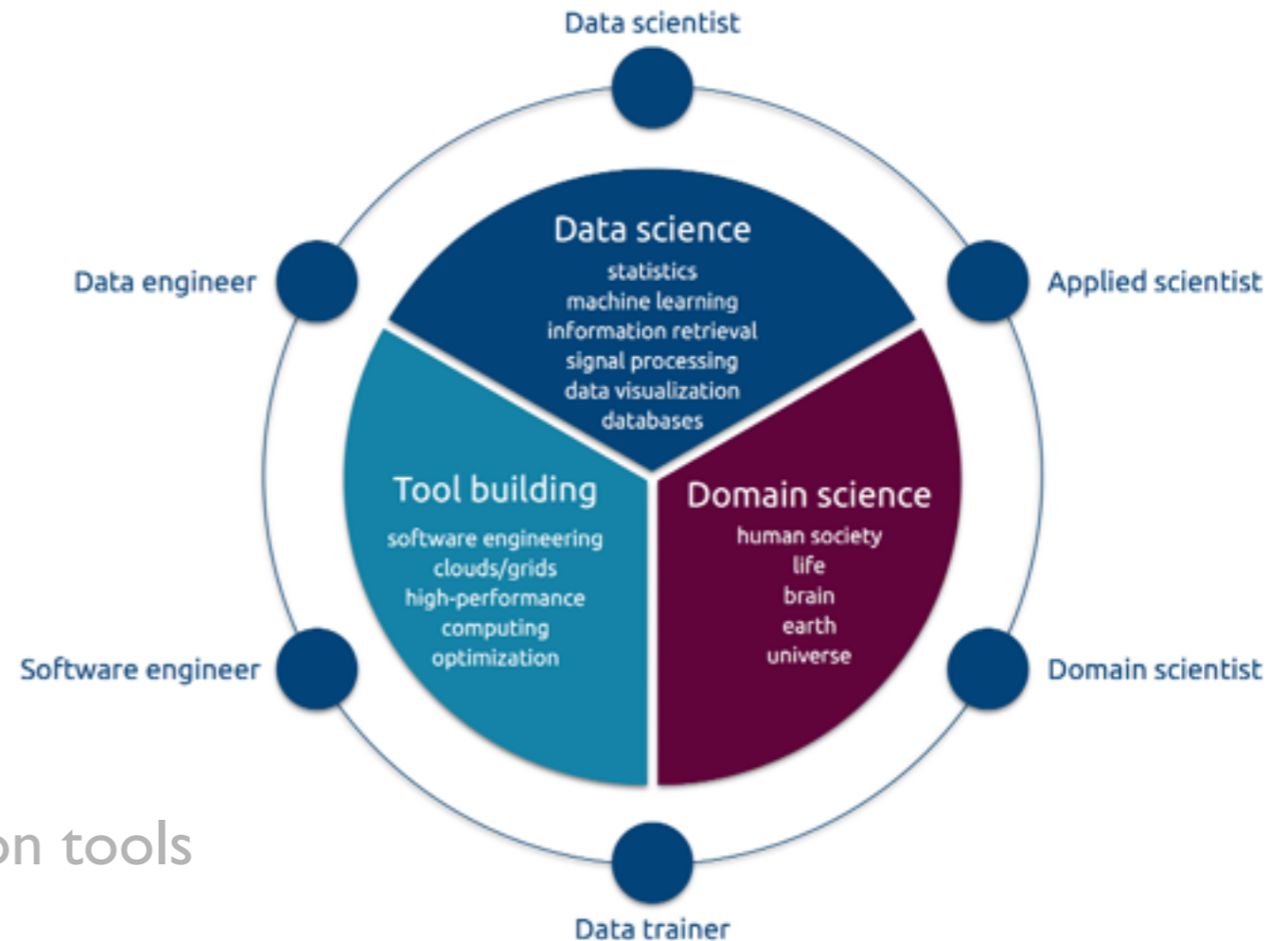
CDS 1.0: A SET OF INNOVATIVE TOOLS AND PROCESSES TO CONNECT COMMUNITIES, TO LAUNCH AND ACCOMPANY PROJECTS

Data scientist

Data engineer

Applied scientist

Data science
statistics
machine learning
information retrieval
signal processing
data visualization
databases

Tool building
software

high

Data domains
sciences
ences
ment
iety

Software engineer

Domain expert

Data trainer

- coding sprints
- Open Software Initiative
- code consolidator and engineering projects

- interdisciplinary projects
- data challenges
- ultrawalls and interactive visualization

- data science RAMPs and TSs
- IT platform for linked data

université PARIS-SACLAY

Paris-Saclay
Center for Data Science

GENERATING HIGH-VALUE SCIENTIFIC PROJECTS USING THE TOOLS WE BUILT

- Our projects are **not classical research projects**

- The CDS builds and runs **management and software tools** to accompany projects **starting at the project definition phase**

  - CDS 1.0 focused on the **design of the tools & interfaces**

  - CDS 2.0 will focus on **project generation**

universite
PARIS-SACLAY

Paris-Saclay
Center for Data Science

CDS 2.0: A SCIENTIFIC STUDIO

- Running the CDS as a **scientific studio**: building projects of **high potential impact**

- Building a strong **core support of platforms and tools**: a critical mass of **core researchers and engineers**

- **Matching** domain science demand to data science supply

- Accompanying projects starting at the building phase: semestrial **pitching days**, bi-weekly consultation
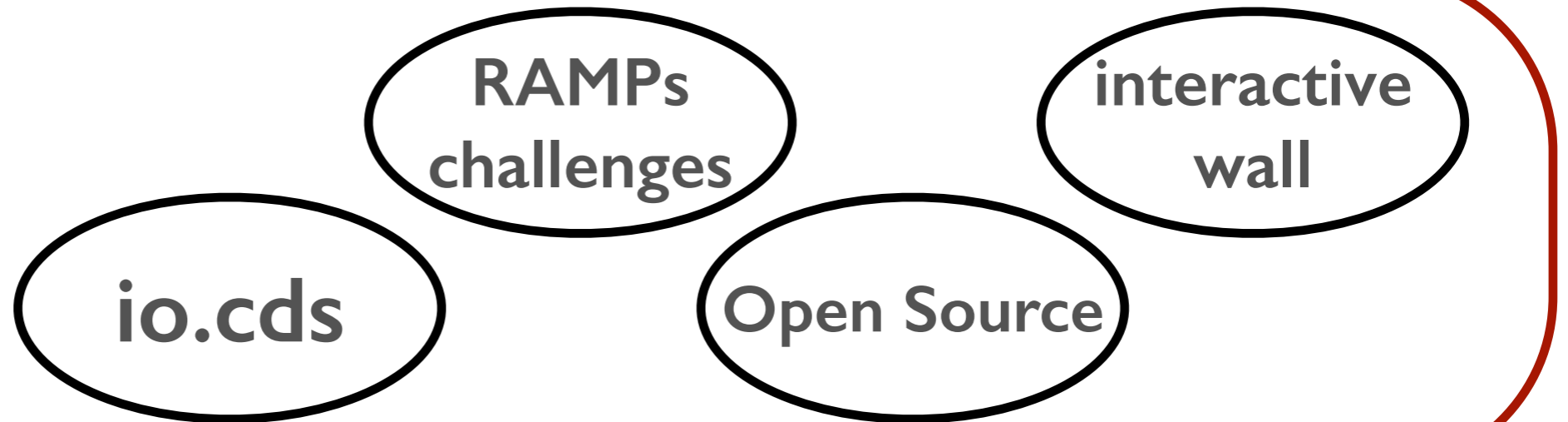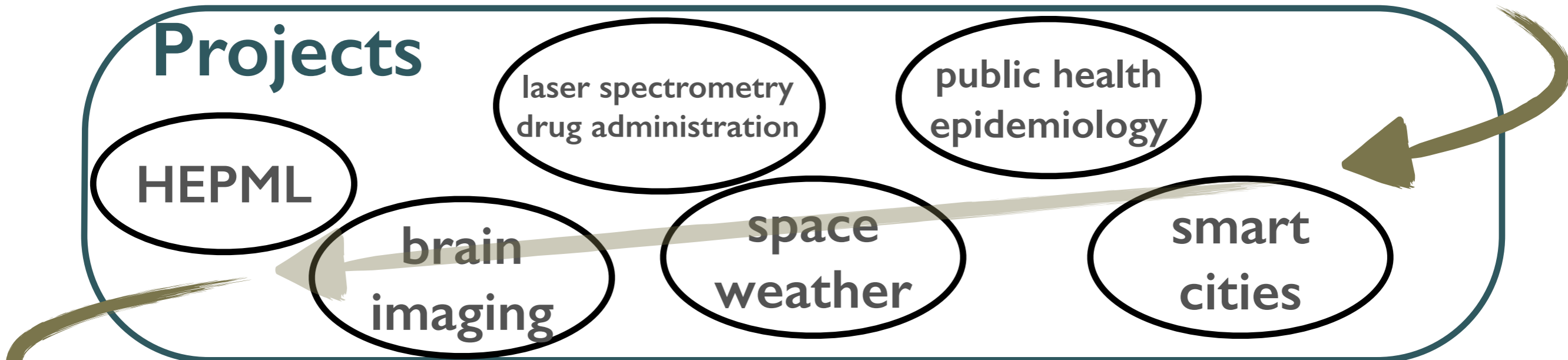
université
PARIS-SACLAY

Paris-Saclay
Center for Data Science

CDS 2.0: A SCIENTIFIC STUDIO

- A two-stage project

  - first stage until **2017 December, 500K€**

- What kind of **support** do we give?

  - **in-kind** support by one of the core engineers and researchers on **using our tools and platforms**

  - **master internships** (200+ data scientists per year at Saclay)

  - standalone **postdoc** or **engineer**

  - bi-weekly open **consulting**

CDS 2.0: A SCIENTIFIC STUDIO

- Pitching day

  - For finding **matching partners**

  - For **directing projects towards adequate type of support**: platforms, postdoc, internship, project-building workshops, external tools

- **No formal call with a deadline**, call will be open **continuously**, projects can be **submitted when they are ready**, evaluated **~once per month**

- CDS Wednesdays

  - here in **Turing**, **every second Wednesday** we will have an informal meetup, **impromptu talks**, **networking**, meet our engineers and researchers, **consulting**

Thank you!