Paris-Saclay Center for Data Science

Balázs Kégl

Linear Accelerator Laboratory and Computer Science Laboratory
CNRS/IN2P3 & University Paris-S{ud,aclay}

LAL, June 30, 2014

Paris-Saclay
Center for Data Science

Campus Paris Saclay
Fondation de Cooperation Scientifique

Data Science

Data Science

Design of automated methods

Data Science

Design of automated methods

to analyze massive and complex data

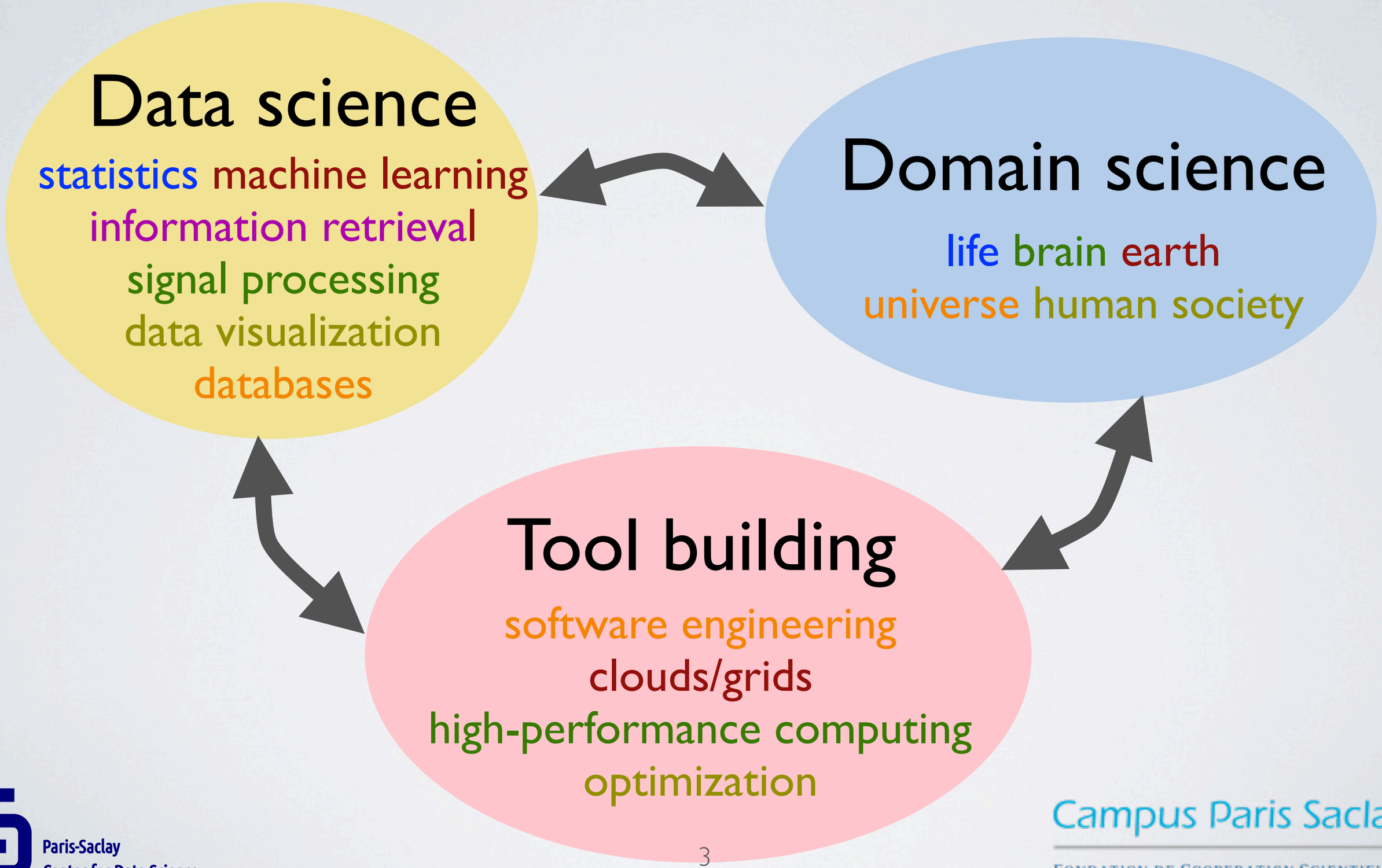
Data Science

Design of automated methods

to analyze massive and complex data

in order to extract useful information from them

DATA SCIENCE - DATA IN SCIENCE



GOALS AND TOOLS

GOALS AND TOOLS

- Build a community at Saclay around data science
  - an agora where researchers and engineers can meet and talk
  - a culture of crossing the disciplinary aisles
  - get to know each other's expertise and data analysis problems

GOALS AND TOOLS

- Build a community at Saclay around data science
  - an agora where researchers and engineers can meet and talk
  - a culture of crossing the disciplinary aisles
  - get to know each other's expertise and data analysis problems
- Tools
  - an interactive web portal
  - workshops, thematic days, and brainstorming sessions
  - summer school(s)

GOALS AND TOOLS

GOALS AND TOOLS

- Get interdisciplinary collaborations off the ground
  - seeding larger projects for outside funding (ANR, Europe)
  - embedding data scientists in domain science labs and vice verse

GOALS AND TOOLS

- Get interdisciplinary collaborations off the ground
  - seeding larger projects for outside funding (ANR, Europe)
  - embedding data scientists in domain science labs and vice verse
- Tools
  - financing postdoctoral projects and (incoming) sabbatical/visiting stays
  - financing data challenges

GOALS AND TOOLS

GOALS AND TOOLS

- Support software tool building
  - primarily open source
  - development, maintenance, reusability across disciplines

GOALS AND TOOLS

- Support software tool building
  - primarily open source
  - development, maintenance, reusability across disciplines
- Tools
  - the Open Software Initiative for interns and doctoral students
  - "code consolidator" and engineering projects
  - coding sprints

GOALS AND TOOLS

GOALS AND TOOLS

- Data science IT platform
  - open data, open access, reproducible research (see IPOL)
  - disseminating and sharing data and software
  - connecting the CDS to data centers (e.g. Virtual Data)

GOALS AND TOOLS

- Data science IT platform
  - open data, open access, reproducible research (see IPOL)
  - disseminating and sharing data and software
  - connecting the CDS to data centers (e.g. Virtual Data)
- Tools
  - interactive web portal
  - work group on open data/data centers

GOALS AND TOOLS

GOALS AND TOOLS

- Design strategies
  - the institutional structure to stabilize the CDS
  - the ideal structures for data science research
  - projecting it onto existing structures (e.g., hotel à projets)

GOALS AND TOOLS

- Design strategies
  - the institutional structure to stabilize the CDS
  - the ideal structures for data science research
  - projecting it onto existing structures (e.g., hotel à projets)
- Tools
  - work group on strategy

GOALS AND TOOLS

GOALS AND TOOLS

- Making national and local decision-makers aware of the challenges
  - lack of incentives for interdisciplinary research and tool building
  - unprecedented brain drain into industrial research

GOALS AND TOOLS

- Making national and local decision-makers aware of the challenges
  - lack of incentives for interdisciplinary research and tool building
  - unprecedented brain drain into industrial research
- Tools
  - lobbying, explaining, thinking together

GOALS AND TOOLS

GOALS AND TOOLS

- Creating a map of Saclay data science masters
  - making the masters transparent

GOALS AND TOOLS

- Creating a map of Saclay data science masters
  - making the masters transparent
- Tools
  - web portal
  - work group on educational relationships
  - initiating an annual meeting on data science education

GOALS AND TOOLS

GOALS AND TOOLS

- Making the CDS a contact point to big data industry

GOALS AND TOOLS

- Making the CDS a contact point to big data industry
- Tools
  - web portal
  - building on existing tools (e.g. CIFRE theses, SystemX, Cap Digital)
  - work group on industrial relationships

THEMES AND PARTICIPANTS

Roughly 250 researchers in 35 laboratories

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

chemistry
EA4041/UPSud

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

visualization
INRIA LIMSI

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

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

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

statistics
LMO/UPSud LS/ENSAE
LSS/Supélec MIA/Agro
CMA/Polytechnique
LMAS/Centrale

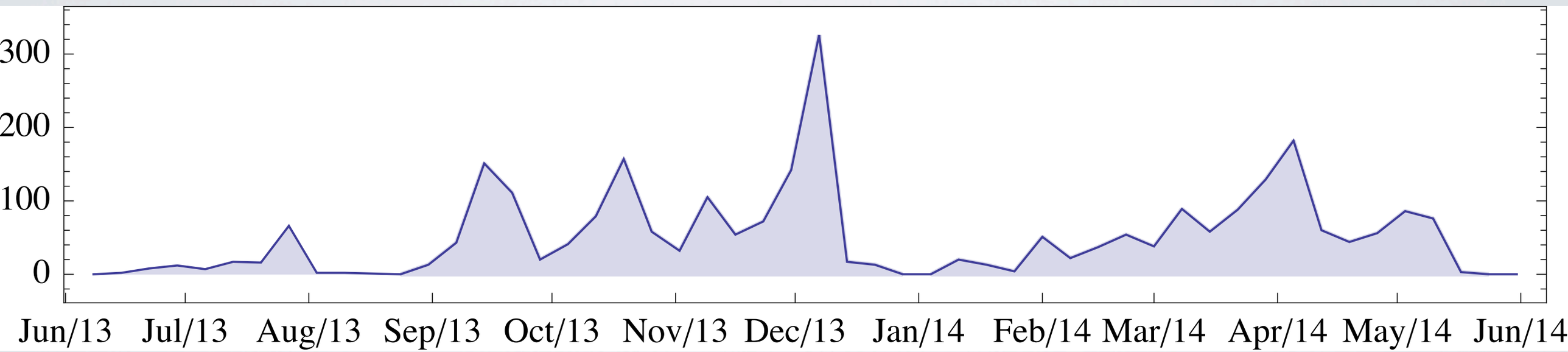
OUR SHORT HISTORY

OUR SHORT HISTORY

2700 mails

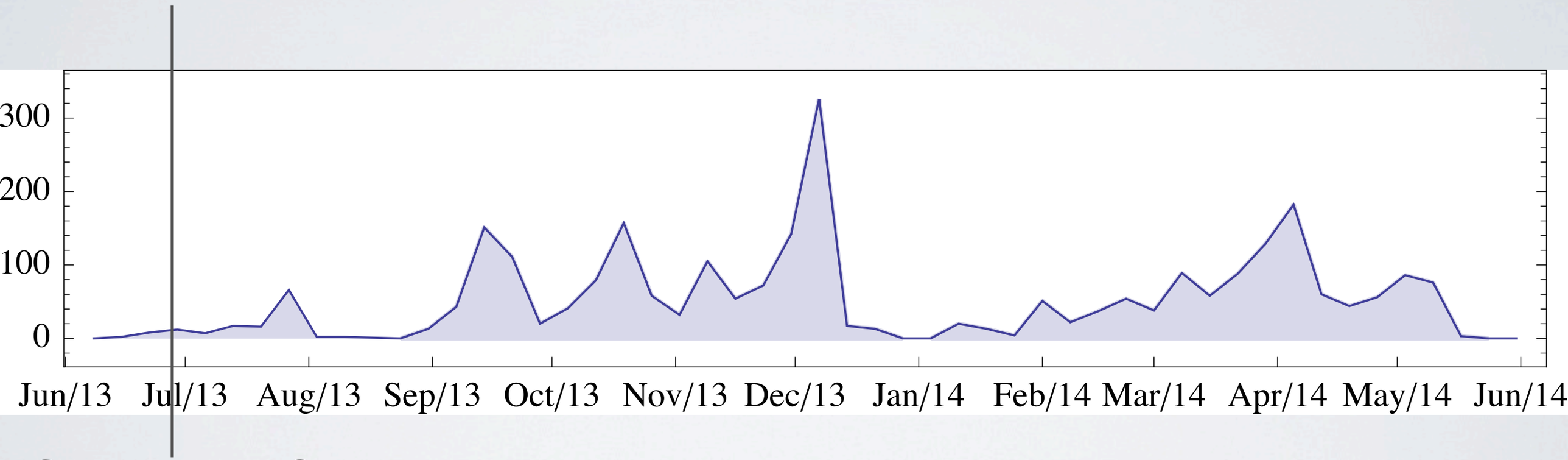
OUR SHORT HISTORY

2700 mails



OUR SHORT HISTORY

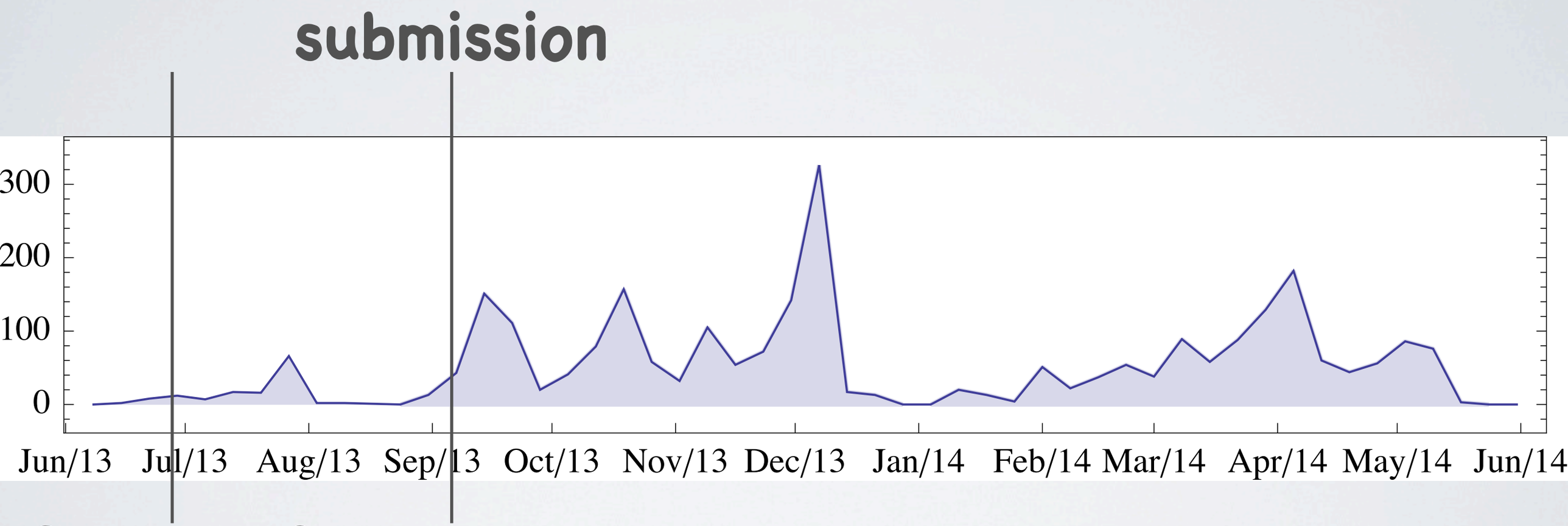
2700 mails



first draft

OUR SHORT HISTORY

2700 mails

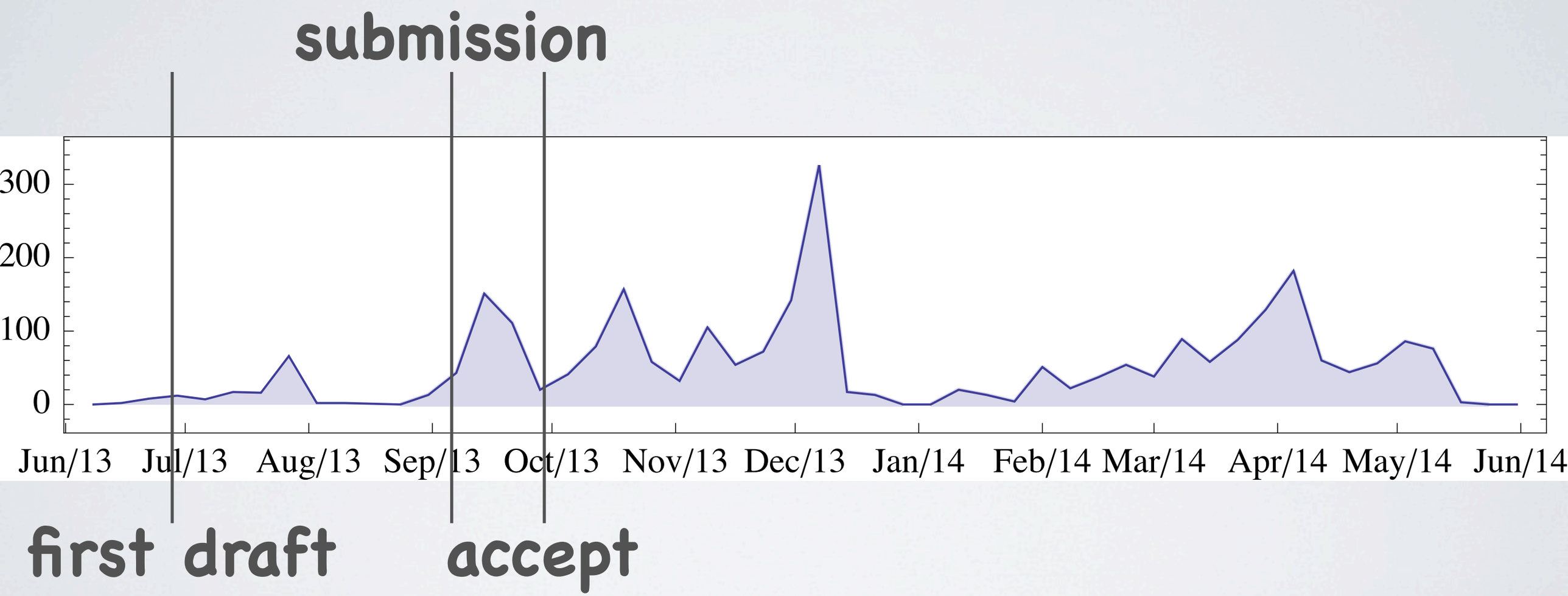


Paris-Saclay
Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE

OUR SHORT HISTORY

2700 mails

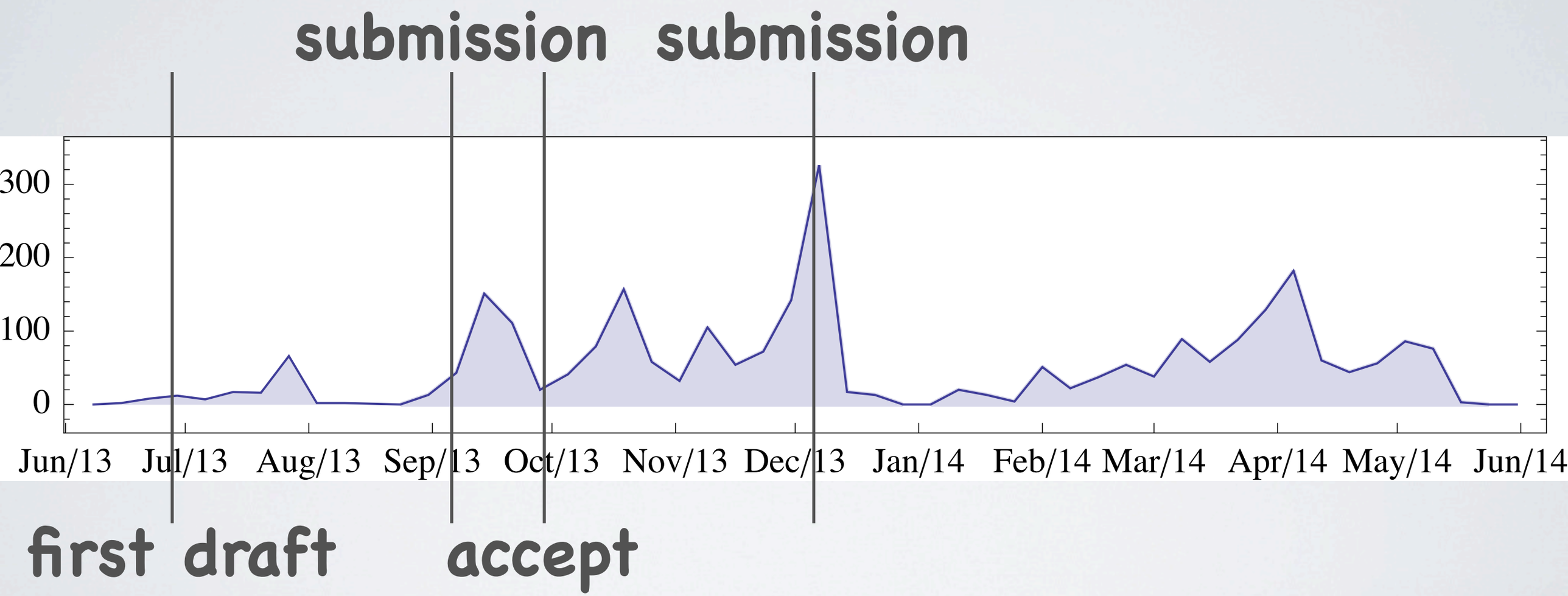


Paris-Saclay Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE

OUR SHORT HISTORY

2700 mails

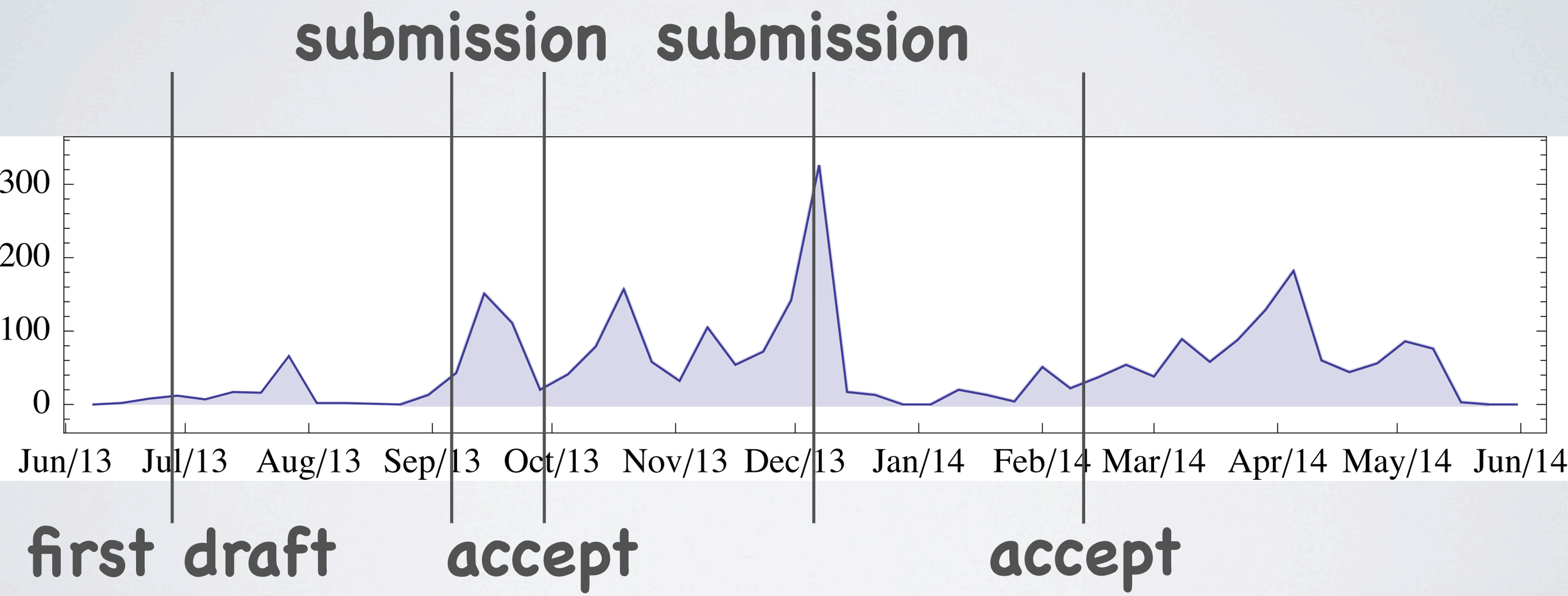


Paris-Saclay Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE

OUR SHORT HISTORY

2700 mails

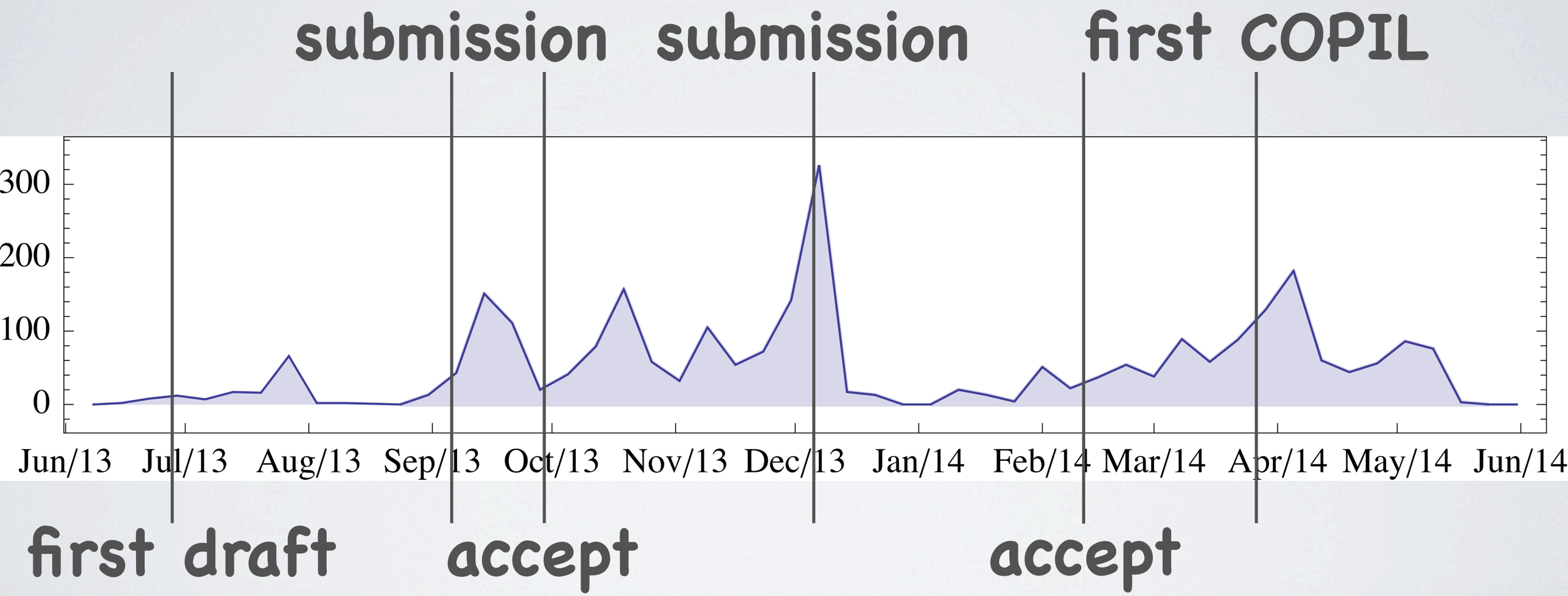


Paris-Saclay Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE

OUR SHORT HISTORY

2700 mails

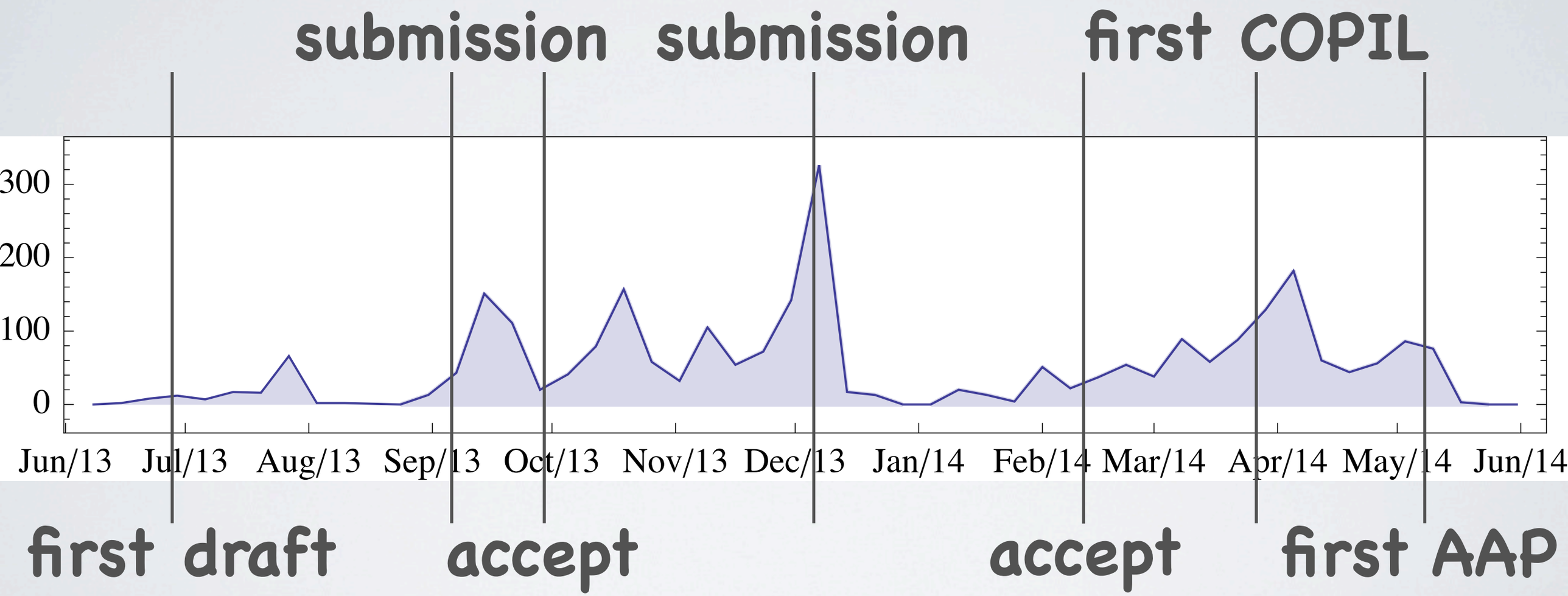


Paris-Saclay Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE

OUR SHORT HISTORY

2700 mails



Paris-Saclay Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE

Main project calls

Main project calls

- First call, ended June 9
  - we select 10-12 projects (out of 22) by mid July

Main project calls

- First call, ended June 9
  - we select 10-12 projects (out of 22) by mid July
- 3-6 theses, without candidates, september 2014

Main project calls

- First call, ended June 9
  - we select 10-12 projects (out of 22) by mid July
- 3-6 theses, without candidates, september 2014
- Second postdoc/engineer call, february 2015

Main project calls

- First call, ended June 9
  - we select 10-12 projects (out of 22) by mid July
- 3-6 theses, without candidates, september 2014
- Second postdoc/engineer call, february 2015
- We may leave the code consolidator and sabbatical/visiting researcher calls permanently open

THANK YOU

- Kick-off organizers:
  - A. Dalalyan, A. Gramfort, B. Kégl
- LAL team:
  - V. Brouillard, G. Dreneau, S. Soullard, D. Benbouzid, S. Mishra
- FCS comms team:
  - A. Dacosta, M-P. Gacoin

Paris-Saclay
Center for Data Science

Campus Paris Saclay
FONDATION DE COOPERATION SCIENTIFIQUE