Domain Scientists and Computer Scientists
Synergies and Implementation

F. Suter

June 3, 2015

Domain Scientists and Computer Scientists

Domain Scientists

- Use computing "stuff" to produce scientific results in their domain
  - Raise interesting challenges for Computer Scientists

Domain Scientists and Computer Scientists

Domain Scientists

- Use computing "stuff" to produce scientific results in their domain
  - Raise interesting challenges for Computer Scientists

Computer Scientists

- Study/Propose computing "stuff" to produce results in computer science
  - Can offer interesting solutions to Domain Scientists

Domain Scientists and Computer Scientists

Domain Scientists

- Use computing "stuff" to produce scientific results in their domain
  - Raise interesting challenges for Computer Scientists

Computer Scientists

- Study/Propose computing "stuff" to produce results in computer science
  - Can offer interesting solutions to Domain Scientists

Theory vs. Practice

- **Theory:** Virtuous cycle with benefits for all
- **Practice:** Mismatching focuses

Domain Scientists and Computer Scientists

Domain Scientists

- Use computing "stuff" to produce scientific results in their domain
  - Raise interesting challenges for Computer Scientists

Computer Scientists

- Study/Propose computing "stuff" to produce results in computer science
  - Can offer interesting solutions to Domain Scientists

Theory vs. Practice

- **Theory:** Virtuous cycle with benefits for all
- **Practice:** Mismatching focuses

How to bridge this gap?

- Develop synergies and foster collaborations
  - In a way that is useful for domain scientists
  - But also publishable for computer scientists (at some point)
- Find the appropriate implementation in the specific context of IN2P3

Dos and Don'ts for Collaborations

What doesn't work

- There is no silver bullet or magical software that solve every issue
  - Codes are too big and intricate
- Thinking a Computer Scientist can understand the computing of Domain Scientists
  - Not with a fair level of (repeated) explanations at least
- All the solutions cannot come from stacks of existing software
  - Might prevent the investigation of upstream trends
- All the knowledge cannot be found within a single community
  - Might lead to wheel reinvention or bad habits propagation

Dos and Don'ts for Collaborations

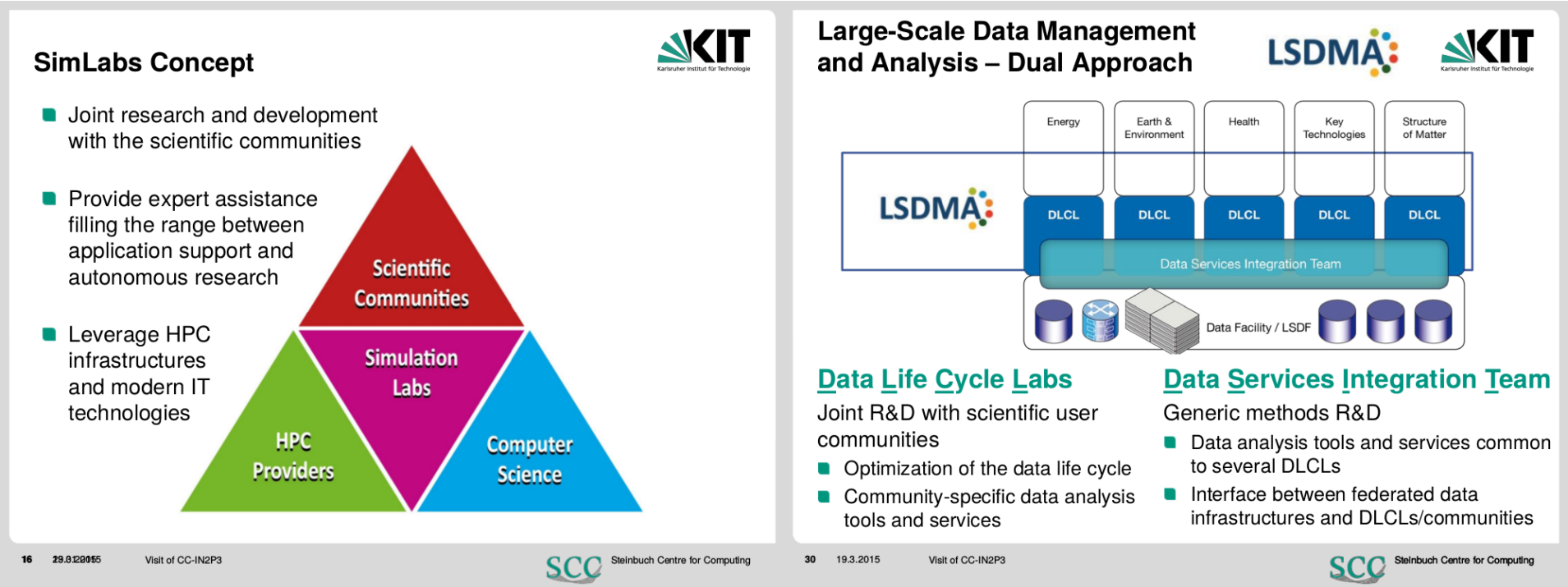
What doesn't work

- There is no silver bullet or magical software that solve every issue
  - Codes are too big and intricate
- Thinking a Computer Scientist can understand the computing of Domain Scientists
  - Not with a fair level of (repeated) explanations at least
- All the solutions cannot come from stacks of existing software
  - Might prevent the investigation of upstream trends
- All the knowledge cannot be found within a single community
  - Might lead to wheel reinvention or bad habits propagation

What could work

- Set up a task force with the right mix of actors
- Focus on a specific problem with a clear potential outcome
  - That satisfies all the actors involved
- Adopt a methodology of collaboration $\rightsquigarrow$ Scientific-centered design
  - Short interaction loop $\rightsquigarrow$ small but frequent improvements

Ideas for Implementations



People involved in a SimLab/DLCL

- Members of the different components (DS, CS, Industrial partner, IT)
- One project leader (permanent staff)
- One hired people (Post-doc/Ph.D.)

Ideas for Implementations (2)

BIDS: Berkeley Institute for Data Science

- http://bids.berkeley.edu/ (led by S. Perlmutter)
- $12.5M over five-years!
- "A central hub of research and education designed to facilitate and nurture data-intensive science"

BIDS Working groups

- Career Paths and alternative metrics (beyond publications)
- Education ad Training (programming, statistics, open science)
- Software tools and Environment (bridge gap between academia and open source)
- Reproducibility and Open Science
- Working Space (to help cross-fertilization)