HAL
open science

Writing a Reproducible Article

Luka Stanisic, Arnaud Legrand

To cite this version:

Luka Stanisic, Arnaud Legrand. Writing a Reproducible Article. Pascal Felber; Laurent Philippe; Etienne Riviere; Arnaud Tisserand. ComPAS 2014 : conférence en parallélisme, architecture et systèmes, Apr 2014, Neuchâtel, Switzerland. 2014. hal-00994575

HAL Id: hal-00994575

https://inria.hal.science/hal-00994575v1

Submitted on 21 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Writing a Reproducible Article

Luka Stanisic and Arnaud Legrand

MESCAL team, LIG, Univ. of Grenoble

COMPAS, 22.04.2014

Context

- HPC applications nowadays use both multi-core CPUs and GPUs
- Managing efficiently computation execution and data transfer is extremely complex
- Need for portable performance $\rightsquigarrow$ Runtime system

Context

- HPC applications nowadays use both multi-core CPUs and GPUs
- Managing efficiently computation execution and data transfer is extremely complex
- Need for portable performance $\rightsquigarrow$ Runtime system

Many configuration parameters:

1. Task granularity
2. Scheduling strategies
3. Application structure

Emerging challenges:

1. Finding optimal combination of parameters for a given machine
2. Evaluate configurations on a wide variety of platforms
3. Quickly identify performance issues (e.g., bottlenecks)

Possible solution: Simulation

Our proposal

StarPU

Dynamic runtime for hybrid architectures. StarPU execution consists in scheduling a graph of tasks with data dependencies on the different computing resources

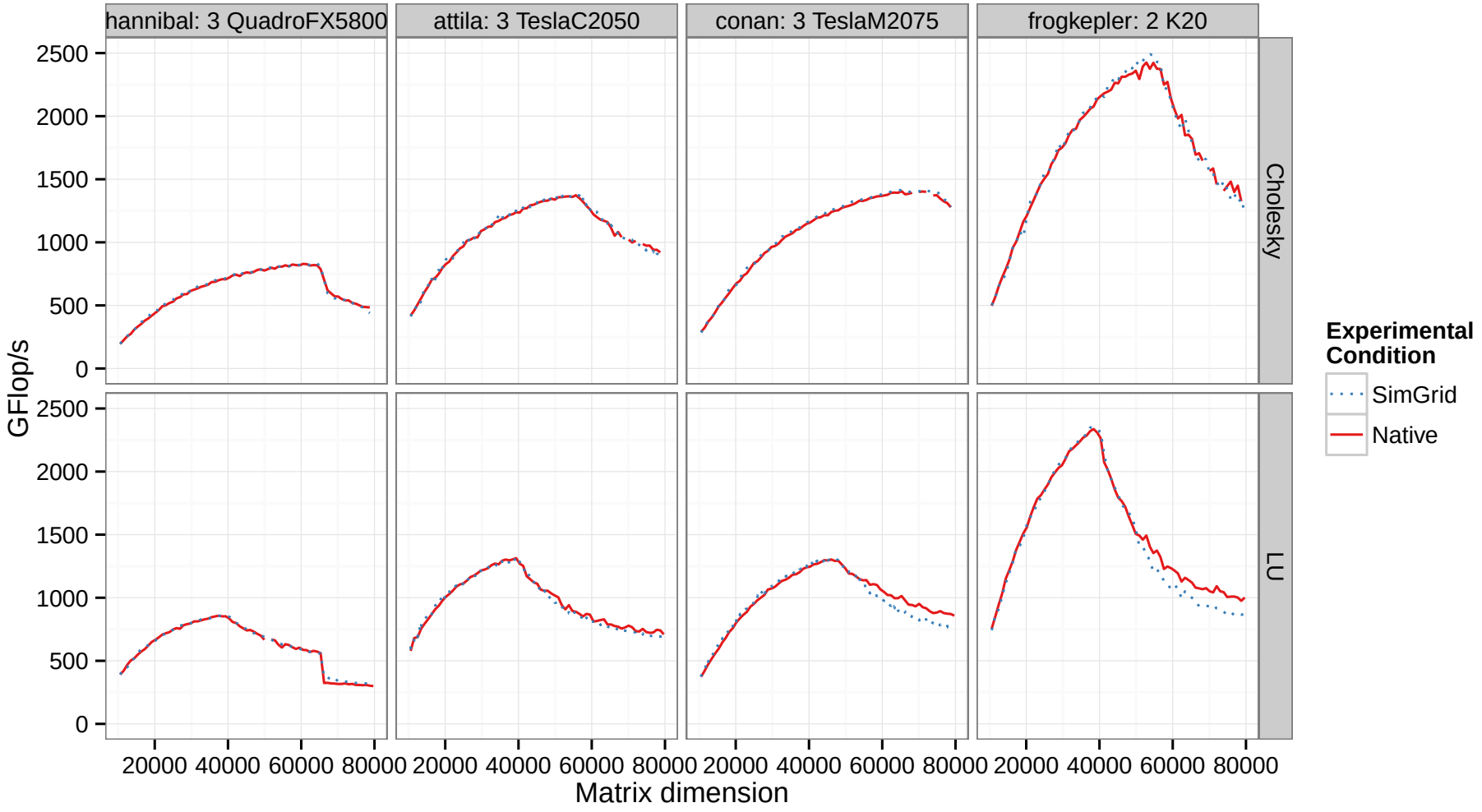
Simgrid

Versatile simulator for distributed systems

Implementation:

- StarPU applications and runtime are emulated
- All operations related to thread synchronization, actual computations and data transfer are simulated
- Control part of StarPU is modified to dynamically inject computation and communication tasks into the simulator
- StarPU calibration and platform description is used by Simgrid

Results


hannibal: 3 QuadroFX5800
attila: 3 TeslaC2050
conan: 3 TeslaM2075
frogkepler: 2 K20
Cholesky
LU
GFlop/s
2500
2000
1500
1000
500
0
20000 40000 60000 80000
Matrix dimension
Experimental Condition
SimGrid
Native

Results

Results

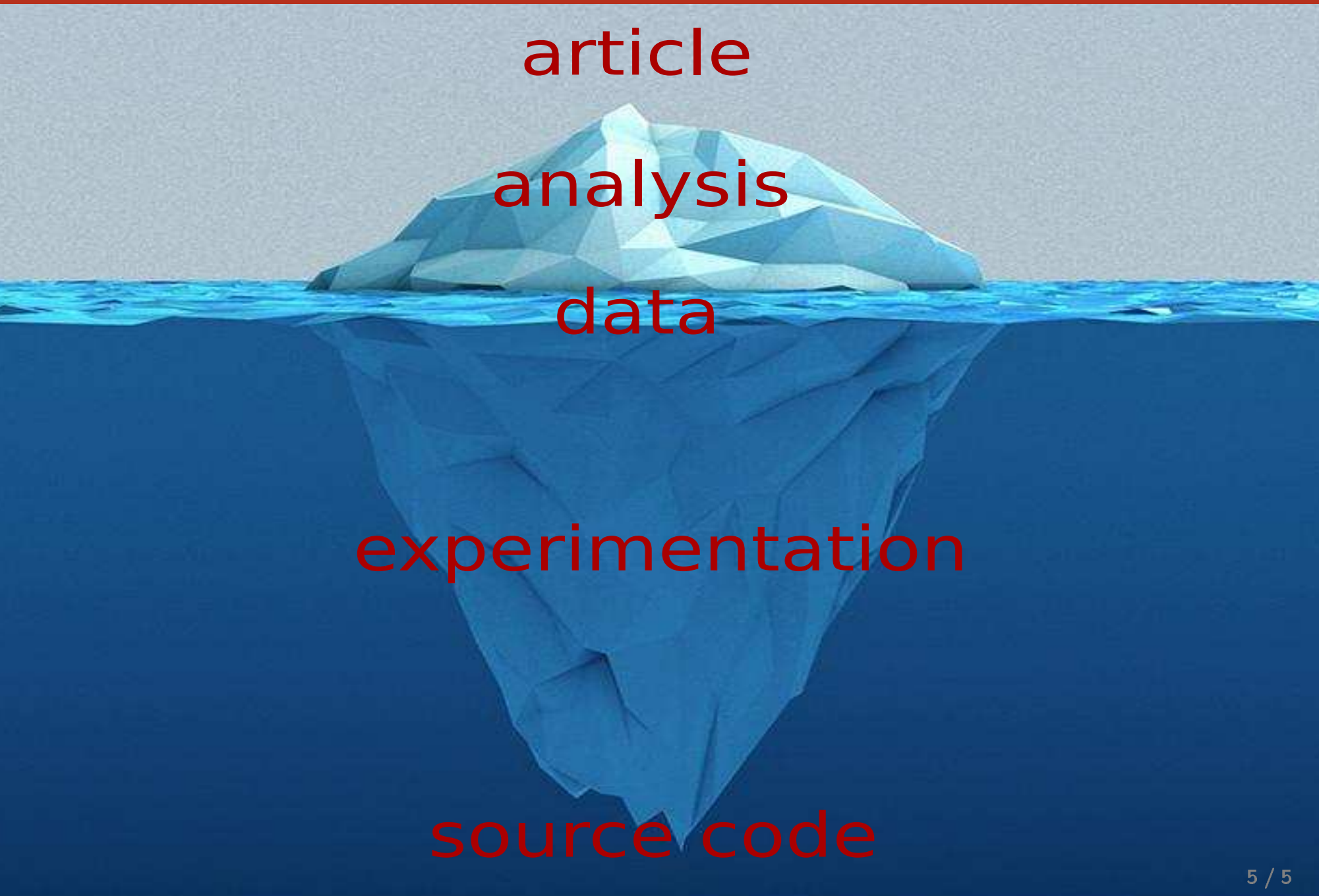
- Works fine now, but coming to this point was not easy
- We had to do many iterations of:
  1. Running complex beta code on several not always dedicated machines
  2. Comparing with simulations, debugging, understanding, remodeling and going back to step 1 until not satisfied

- With good results, we decided to make a reproducible article
  1. From outside it looks like any other pdf paper
  2. From inside ...

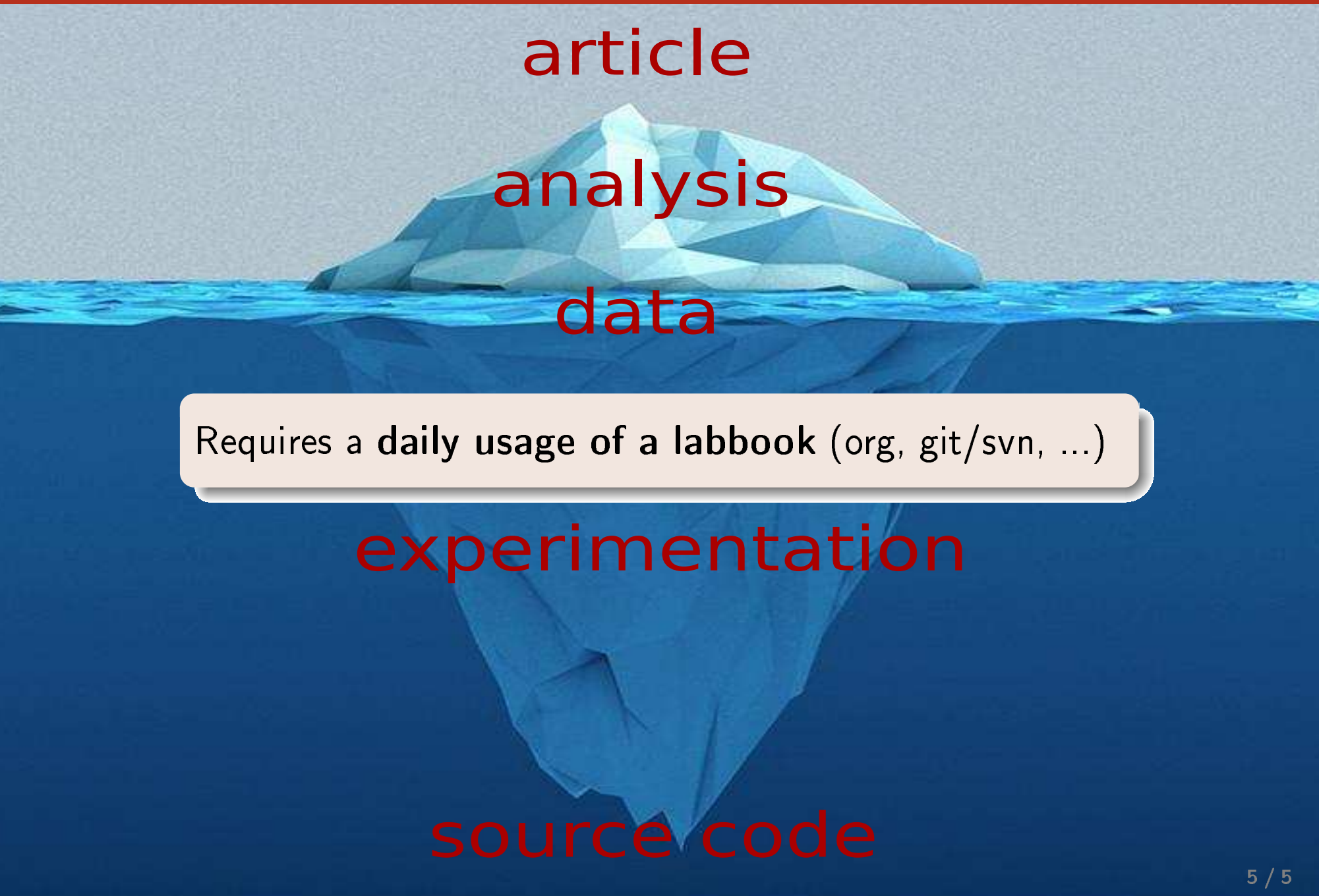
I will try to convince you that our article is not only reproducible but also readable and understandable!

http://dx.doi.org/10.6084/m9.figshare.928338

Reproducible Research



Reproducible Research



Requires a **daily usage of a labbook** (org, git/svn, ...)