Sustainable software as an infrastructure

Oxana Smirnova
Lund University / NeIC

Generic Components of the eScience Infrastructure Ecosystem Workshop
at the 14th IEEE eScience Conference, 29 October 2018, Amsterdam

Data matures like wine
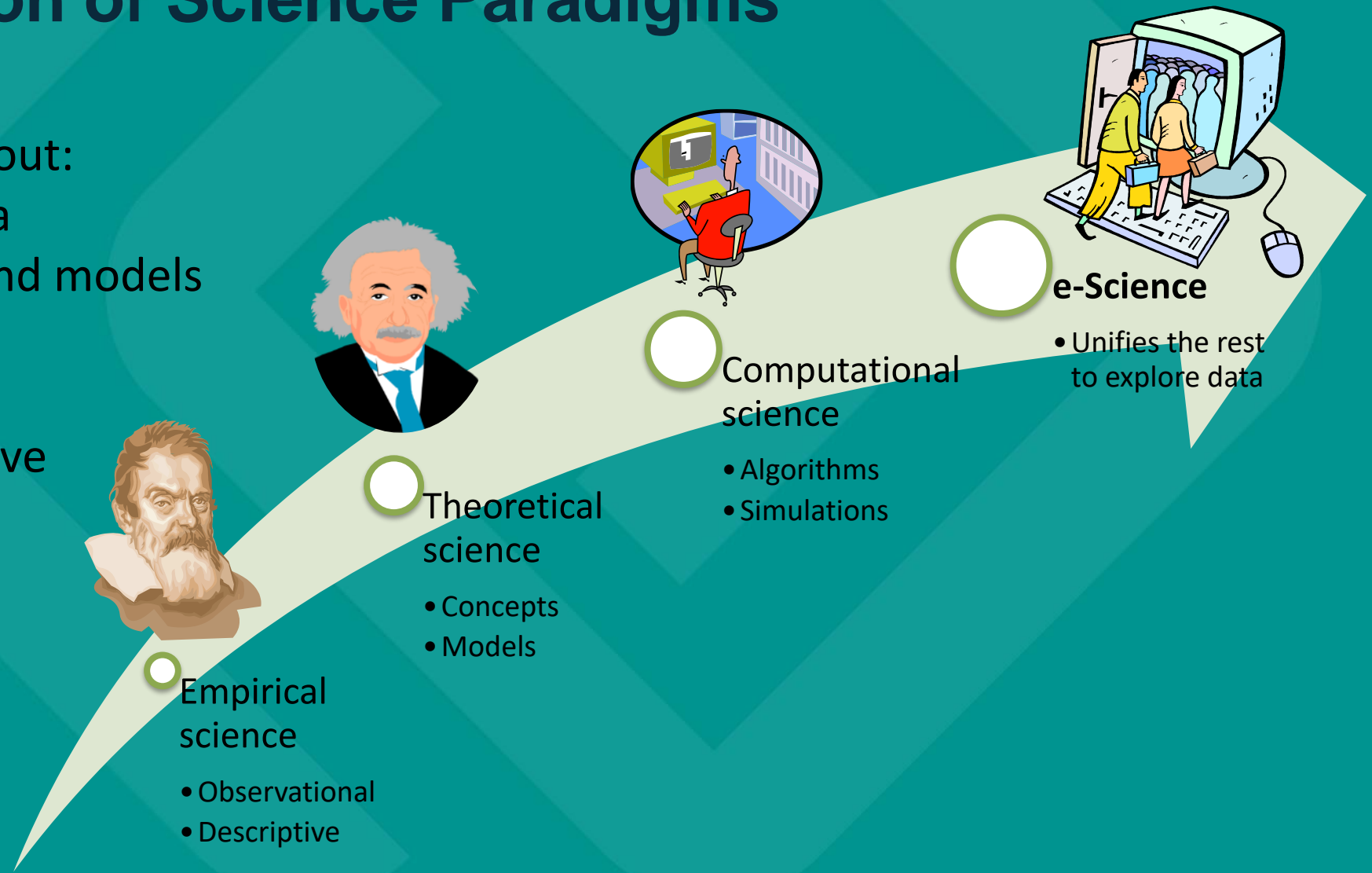# Software matures like fish

A cyberspace adage

Evolution of Science Paradigms
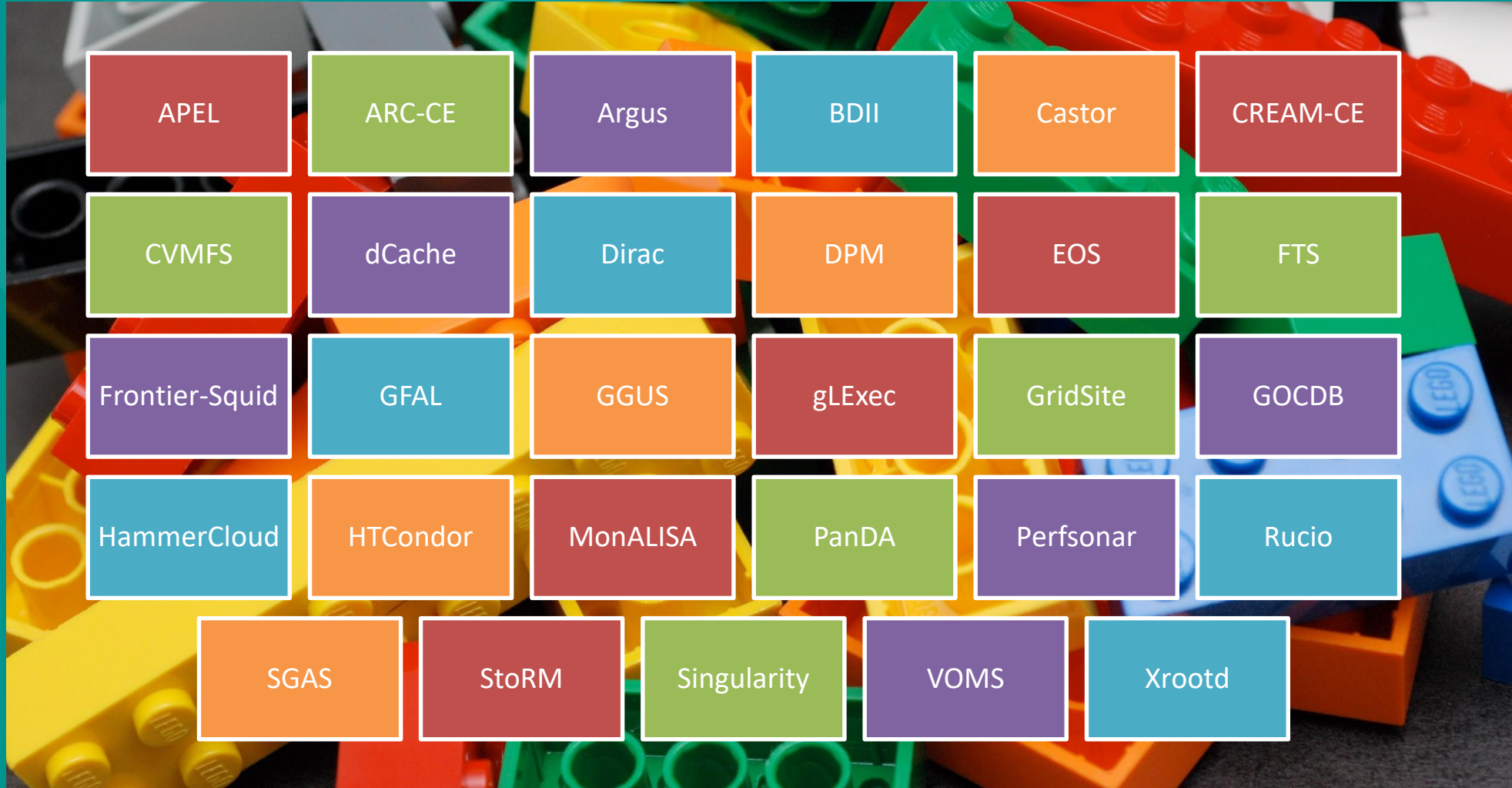
- E-Science is about:
  - Digitised data
  - Algorithms and models
  - Data centers
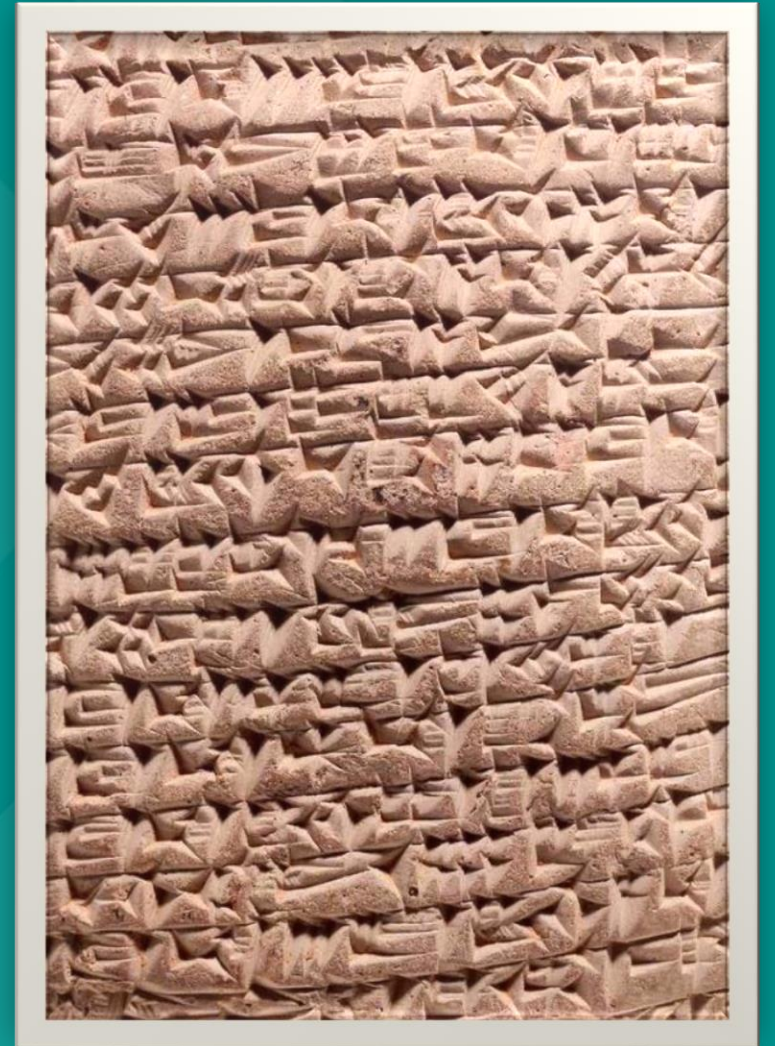 - **Software** for
    all of the above

e-Science
- Unifies the rest to explore data

Computational science
- Algorithms
- Simulations

Theoretical science
- Concepts
- Models

Empirical science
- Observational
- Descriptive

Context: WLCG and its software
A very incomplete list of components:

APEL	ARC-CE	Argus	BDII	Castor	CREAM-CE
CVMFS	dCache	Dirac	DPM	EOS	FTS
Frontier-Squid	GFAL	GGUS	gLExec	GridSite	GOCDB
HammerCloud	HTCondor	MonALISA	PanDA	Perfsonar	Rucio
	SGAS	StoRM	Singularity	VOMS	Xrootd

The challenge:

Will people be able to use LHC data when all this software and infrastructure is gone?

Software as an intangible asset

A kind of data?

Some of it is kept in repositories, and some of it is inseparable from data	Unlike data, it **evolves**, often in a backwards-incompatible manner

A kind of publication?

It is made public, has authors, can be copyrighted and can even get a DoI	Unlike papers, needs **maintenance**, each release is a different object

Everybody wants it for free!
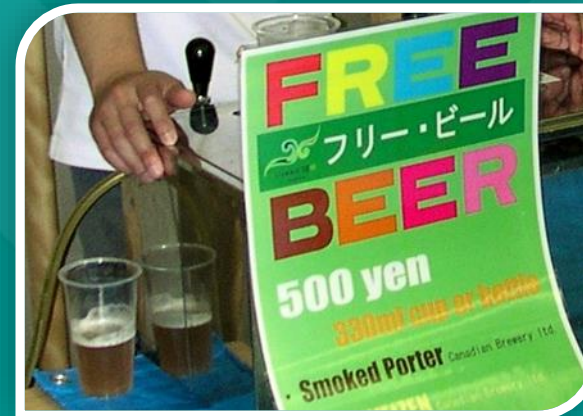
Survey time:

Go now to www.menti.com and enter code 11 88 38

Perhaps software is a research tool?

Yes, if you ask me: indeed, like any device, software needs	Not, if you ask funding agencies
• Design, prototyping, construction / implementation and commissioning / testing • Permanent support, maintenance and upgrades • Documentation and usage protocols • Domain-specific expertise	• A tool is tangible, we always know who uses it (RoI) • Software often requires constant funding flow, tool is paid upfront • Maintenance and upgrade of software does not amount to research – neither in physics, nor in CS

WLCG middleware as a limited-scope software

- Multimillion investments into R&D
- But Grid is only useful when an e-Infrastructure has more than ~5 sites
- For most e-Science areas it is cheaper to get an allocation at an HPC center or buy Cloud time

Non-existing market: only 266 certified sites in entire EGI as of today

High experienced support costs per customer (site/infrastructure)

Customers are not willing to foot the support bill – not now, not in future

Best-effort support, poor quality

Customers are increasingly switching to "free" or "home-made" solutions

How did we get there?

Many failed to realise that infrastructure-enabling software is a **scientific tool**

Illusion #1: scientific software can have a business use case
- Some can, but not much more than a collider

Illusion #2: software is free like air
- It is only free like a toll-free road

For a long while there was little or no support for e-Infrastructure enabling software R&D
- We got stuck with ancient technologies
- New technologies were not developed for us

HL-LHC is a challenge that can not be ignored
- HEP Software Foundation emerged
- First positive results funding-wise
- Still no breakthrough sustainability-wise

Software now and then
Data management systems as an example

Rucio

- Evolution of experiment-specific tool
- Supported by specialist community
- Rapidly evolves into a generic tool

LHC File Catalogue (LFC)

- Generic, experiment-agnostic
- Supported via EU funding
- Too basic to be useful – dead

A very personal take on infrastructures

Do we need modular Lego-like software to build e-Infrastructures?

Or is software an infrastructure itself?

Device-specific software

Serves a very specific purpose

High maintenance, but on a small scale

No need to scale out

As sustainable as the device

Public domain software

- Low maintenance
- Robust
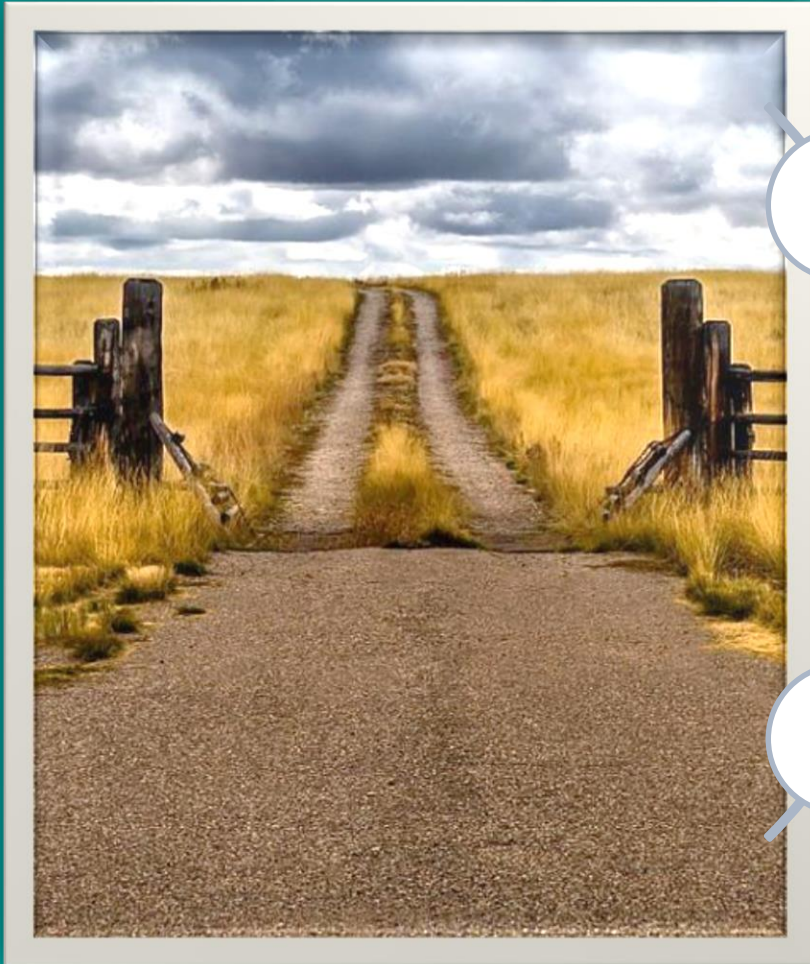- Self-sustainable if new features are not needed

Commercial software

- Profit-oriented
- Multi-purpose, scalable
- Sustainable as per business model

Community software

- Tailored for different purposes
- Different communities use different models
- Maintenance levels differ
- Sustainable as long as communities support it

Grid software

- Strictly structured
- Heavy-duty, but quickly dilapidates
- Does not suit every vehicle
- Problematic sustainability

Application software

- A-mazingly elaborate
- Organically grown
- What is "sustainability"?

Software of your dream?

Free and smooth

No glitches or traffic jams

Supported by a philanthropist, forever

Or maybe we all should buy highway stickers?