Involving the Public in High Energy Physics through IT Challenges and Projects

Claire Adam-Bourdarios for the ATLAS outreach and computing groups
CERN & LHC ...
As presented on the Web

What is the universe made of? How did it start? Physicists at CERN are seeking answers, using some of the world's most powerful particle accelerators

LHC : 27km ring of superconducting magnets, 2 proton beams, collisions every 25 ns

2010 1st run starts
( 7 and 8 TeV – 3 years )

2012-2015  Discovery & study of the Higgs boson

2015 2nd run starts
( 13 TeV, higher intensity )

HL-LHC under study
ATLAS is...
an experiment at CERN
designed to explore the secrets of the universe

Discover ATLAS

Resources

Updates

ATLAS...
As presented on the Web
Target audiences:
- General public
- Policy makers in science and technology
- Students and teachers

Themes:
- Physics
- Collaboration
- Technology

Platforms:
- Web site (resources & updates)
- Social media
- Visits, virtual visits and events
- Master classes

ATLAS OUTREACH
Classical audiences and platforms

ATLAS-OREACH-PROC-2015-003
Innovating science communication
Moving towards citizen science?

ATLAS Software and Computing tasks in a nutshell

3000 physicists:

- 180 institutes in 38 countries
- 130 computing centres worldwide
- 200k to 250k jobs running in parallel, on the WLCG computing grid

“Typical” activities:

1. Simulation: extrapolate physics & detector knowledge to expected LHC running
2. Event reconstruction and filtering: tune algorithms to enhance signal/background ratio
3. Specific event picking and interpretation
Moving towards citizen science?
New themes, actors and projects


2. Signal extraction  https://higgsml.lal.in2p3.fr/


ATL-OREACH-PROC-2016-001
Claire Adam-Bourdarios
Volunteer computing

WHY?

In @home projects citizens voluntarily run scientific tasks on their home PC’s spare time:

- For ATLAS: (almost) free resources
direct contact with a very active and interested/-ing community
increased software and computing outreach opportunities

- For the Volunteers: get/stay engaged into science and high tech recent developments
competition “for fun” via a credit system
Volunteer computing

HOW?

DESIGN CONSIDERATIONS & SOLUTIONS

• Resources can come and go at any time:
  Run simulation and favour “non time critical tasks”
• Resources are not homogeneous:
  Virtual Machines, BOINC platform
• Nothing sensitive can be sent (e.g. certificates):
  ARC architecture

STRONG PROJECT INTEGRATION

• With other LHC projects, CERN IT department, BOINC developers community
• Within the ATLAS computing system: job submission, validation and data handling
• With the experiment outreach: specific resource web site designed for volunteers
Volunteer computing

As seen by ...

- Real simulation tasks
  - Full athena jobs
  - 25 events/job
  - Each job runs for 1-4 hours
  - Real Grid jobs ~12 hours (1000 events and 8 cores)
- Virtual image is 1.1GB (500MB compressed) and downloaded only once
- Input files (data file + small scripts) is 1-100MB
- Output is ~100MB
- Virtual machine memory is now 2GB (was 1GB initially, but now more complex jobs)
  - This is the limiting factor for many volunteers

1. Simulation

Volunteers

Physicists

- Install BOINC client and VirtualBox
  - Linux, Mac and Windows supported
  - Currently 80% of hosts have Windows
- In BOINC client choose ATLAS@Home and create an account
- That’s it!
- BOINC client can be configured to run whenever is convenient, e.g.:
  - After computer is idle for 5 mins
  - Only between 5pm and 8am
Volunteer computing
Declaring success!

Job statistics since May 2014

- Continuous increase in running jobs
- Now 11k parallel running jobs
- 2.5M completed jobs
- 5M CPU hours, 58M events
- 50% CPU efficiency
- Gaps are due to technical issues, not lack of volunteers

Volunteer growth

Currently 70k volunteers:
- 7500 ran at least one job
- 2500 currently active

Einstein@Home: 300k volunteers, 47k active
Seti@Home: 5 million volunteers, 150k active
Volunteer computing

Lessons learned?

FOR ATLAS COMPUTING

• The project architecture scales, and is an interesting alternative for small computing sites
• Physics process generation might be a good use case
• Replying to messages and continuous interaction with volunteers is ESSENTIAL

FOR ATLAS OUTREACH

• Badges, event display and web site are a 1st step, but we also need more Computing related stories: portraits, milestones, news, (virtual) visits...
• VISITORS : only pass by, attracted media
VOLUNTEERS : need arguments to chose our project
FANS : follow project, message boards, and help!

Unavoidable limitation: some volunteers may not want/need to know much about us?
The HiggsML challenge

WHY?

Top quark pair production at HL-LHC

Increasingly looking at more and more data, with complex event structures and to find very rare signals

One approach to meet these challenges is to use Machine Learning, at many points of chain from data taking to final analysis.
The HiggsML challenge

AIM:
• Benefit from the knowledge of the ML community
• Engage the public with the type of analysis work required to discover the Higgs Boson

SETUP:
• Organising team made of ATLAS physicists and ML experts
• 1 year of preparation: define problem, workshops path, data access conditions
• Financial support from Université Paris-Saclay
• Challenge hosted on the Kaggle platform – May to September 2014

SUCCESS:
• 1785 teams, 1943 players, 35772 entries
• HEP meets ML award, cash prices for best 3 competitors
• Signal significance enhanced by 20% with respect to traditional HEP tools
Follow up in the ML community:

• Dedicated workshop at NIPS 2014
• Simulated events used are now available publicly
• Work continues to integrate all types of experimental uncertainties

Follow up in the HEP community:

• Workshop & winners visit at CERN in May 2015
• Integration in the HEP tools & analysis takes time
• Track finding in dense environments is a good candidate for further exploration
Several ATLAS groups are currently developing university level lab courses for teaching Data analysis strategies such as selection optimisation, histogramming and statistics... as well as detector and accelerator physics.

ATLAS WILL RELEASE SOON:

• 1 fb-1 of data @8 TeV – i.e. $\sim 10^7$ events
• Simulated samples will include all standard processes, plus some Higgs and Z’
• Total size: $\sim 7$ Gb in storage
• Format: modified version of the mini ntuple used for ATLAS top quark analysis

*Future will say if it is enough for advanced Machine Learning studies? Probably not.*
Progress in ML is strongly and proactively supported the ATLAS Collaboration

**BUT**

In particle physics, one has to be part of the collaboration to access official real and simulated Data. All analysis paper are signed by the full collaboration (so 3000 for ATLAS!)

⇒ “Partnerships” need to be “structured”:

The status of “Short Term Associate” (STA), initially setup for theorists, was successfully used for 3 data scientists from the University of California Irvine... but is heavy.

A simplified status of “Analysis Consultant and Expert” (ACE) was recently setup, with access to simulation and relevant internal discussions.
3. Event interpretation

Higgs Hunters @ Zooniverse

Why?

EVENT PICKING & SCANNING:

- Is a genuine part of the physicists toolkit, when unusual events are found
- Is a genuine part of the programmer toolkit, to cross check algorithms performance
- Is a very powerful entry to HEP in Master Classes & education programs
Higgs Hunters @ Zooniverse
HOW?

A Web-based citizen science platform

• Volunteers help researchers:
  with huge amount of data
  where human eye is better than algorithm
• Volunteers engage in analysis and learn physics.

HIGGS HUNTERS:

• Look for “secondary vertices” created when a
  particle “flies” before decaying
• Physics case: peculiar decays of the Higgs boson,
  “exotic” models of new physics @LHC
• Interesting events are discussed on the forums
Higgs Hunters @ Zooniverse

WORK IN PROGRESS

• Very clear success very soon after the project launch:
  ~ 7000 people
  Some left, some persisted and still classify a large number of images

Major milestone in Jan 2016: 1 million classifications!!!!

• In citizen science people help & learn from each other:
  Interesting events are discussed
  Overall expertise grows

• Early analysis suggests that people are rather good
  at finding secondary vertices, and complementary
  to the computer algorithms. They also spot unusual
  features which the algorithms are not trained to spot.
A side remark
On Event display & interpretation

Our event displays are attractive: what about letting people explore 3D and learn with... virtual reality?

ATLAS RIFT: based on the Oculus technology + collaborative development with very engaged community.

Major outreach development?
The projects initiated in 2014 have been successful in many ways:

- Participation is a proof of the interest triggered by our scientific achievement and outreach effort.
- Enhanced visibility of the Software and Computing community, which is too often working behind the scenes.
- Many open questions will need interdisciplinary investigations: volunteers profiles, motivation, needs

How do we connect these communities?

Agora: central spot in ancient Greek city-states Gathering, assembly and market place.
Who are the volunteers?

Interdisciplinary dialog is not always easy

How does a Community grow and build up expertise?

In all cases, a rewarding but non negligible work is needed to build the communities... and keep them alive.