ATLAS Open Data Project: HEP for everyone

Meirin Oan Evans, University of Manchester,
On behalf of the ATLAS Collaboration
39th International Conference on High Energy Physics, Seoul,
July 6th 2018
What & Who?

• What is ATLAS Open Data?
  – education, training, outreach
  – public data, education tools
  – software, platforms, analysis examples
  – analyse ATLAS data, learn HEP techniques

• Who? Those with basic HEP knowledge
  – through web or labs organised by experts
ATLAS Open Data portal

- opendata.atlas.cern

ATLAS Open Data website introduction video

- opendata.cern.ch
Current datasets

- 1 fb$^{-1}$ of 8 TeV 2012 tuple data
- MC:

  \[ W^\pm \rightarrow \ell^\pm \nu_\ell \]

  \[ Z \rightarrow \ell^+ \ell^- \]

  low-mass Drell-Yan

  \[ t\bar{t} \rightarrow \ell + X, \ t\bar{t} \rightarrow \text{jets} \]

  \[ W^+W^-, \ W^\pm Z, \ ZZ \]

  single top production

  \[ H \rightarrow W^+W^-, \ H \rightarrow ZZ \]

  \[ Z' \rightarrow t\bar{t} \]
• Dataset contents:
  – final state lepton, jet 4-momenta
  – truth matching booleans
• Single lepton trigger
• No systematics nor truth 4-momenta
• Example analyses:
Analysis Tools

- Different ways to get the data & tools
- Animations & apps

ROOT Notebooks
(ROOTbooks)

Virtual Machines

Visualisation of ATLAS detector
created by ROOT
GitBooks & Links with other groups

- GitBook documentation, similar to markdown-kind adopted by ATLAS collaboration
  - example Open Data GitBooks [1], [2]
- Constant contact with other groups looking for public data, like ATLAS Machine Learning
  - participated in 2016 & 2017 ATLAS Machine Learning Workshops
  - plan to continue to be present there
Review of ATLAS Open Data Dataset

Review of ATLAS Open Data 8 TeV datasets, tools & activities

- Open Data in master’s theses [1], [2]
• Produce multimedia materials with communications team
• Show the web platform & tutorials

**Videos**

- People using & creating ATLAS Open Data & Tools
- ATLAS Open Data website Data & Tools video
- ATLAS Open Data Zll intro notebook
• **CERN website** gives updates on ATLAS Open Data

• **ATLAS Open Data description** on CERN Open Data portal

ATLAS' Higgs ML Challenge data open to public

Explore LHC Data on new ATLAS educational platform

Dortmund students work with ATLAS data
2015 Data 3.2 fb$^{-1}$

Other:

$W^\pm \to \ell^\pm \nu_\ell$

low-mass Drell-Yan

$t\bar{t} \to \ell + X$

$W^+W^-, W^\pm Z, ZZ$

single top production
Open Data users

California
Montréal
ZPATH
Lund
Dortmund
Göttingen

Montréal
ZPATH
Lund
Dortmund
Göttingen

Michigan
HYPATIA
Maastricht
Dresden

CEVALE2VE
ICTP
LIP
Creative Commons
Valencia
Open Data users

- ZPATH
  - Oslo
- Lab course
  - Dortmund
- Lab course
  - Göttingen
- Lab course
  - Manchester
- HYPATIA
  - Athens
- Web-based interactive learning tool
  - Georgia
- CEVALE2VE
  - Latin-America

ATLAS Open Data Project – Meirin Oan Evans – July 6th, 2018
Stats: Active users (1 Jan 18 - 23 Jun 18)

- **Users**: 1,121
- **New Users**: 1,049
- **Sessions**: 2,585
- **Number of Sessions per User**: 2.31
- **Pageviews**: 8,191
- **Pages / Session**: 3.17
- **Avg. Session Duration**: 00:04:02
- **Bounce Rate**: 43.91%

**Top Channels**
- Direct: 51.6%
- Referral: 33.9%
- Organic Search: 7.6%
- Social: (Other): 7.6%
Join us!

• Projects
  – Desktop
    • Masterclasses, lab courses, ATLAS & training workshops, self-service for external users
  – Web-based
    • HTML, GitBook, Jupyter, videos, Electron apps, “nteract”
• Ideas
  – app store, distribution of updates, provision in low-internet access locations
 Plans

• 13 TeV dataset development & production
  – hope to release all 2015 data by 2019
  – add $H\rightarrow\gamma\gamma$ data, MC, analysis
  – include systematics & truth info
• Focus on workshop development
• ATLAS Open Key
  – USB stick
• **8 TeV ATLAS Open Data ready to use!**
  – tools & documentation help users make their own discoveries
• Suitable for everyone interested in HEP
  – high school students to researchers
• Useful for education, outreach, public engagement, science communication, scientific research
• **13 TeV ATLAS Open Data on their way**
Thanks!

• [atlas.outreach.data.tools@cern.ch](mailto:atlas.outreach.data.tools@cern.ch) for general ATLAS Open Data questions

[opendata.atlas.cern](http://opendata.atlas.cern)
Backup
User acquisition (1 Jan 18 - 13 Jun 18)

- **New Visitor**: 339 Users (26.5%)
- **Returning Visitor**: 73.5%

### Acquisition

<table>
<thead>
<tr>
<th>Source</th>
<th>Users</th>
<th>New Users</th>
<th>Sessions</th>
<th>Bounce Rate</th>
<th>Pages / Session</th>
<th>Avg. Session Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>553</td>
<td>553</td>
<td></td>
<td>41.41%</td>
<td>3.64</td>
<td>00:04:06</td>
</tr>
<tr>
<td>Referral</td>
<td>363</td>
<td>363</td>
<td></td>
<td>45.76%</td>
<td>3.72</td>
<td></td>
</tr>
<tr>
<td>Organic Search</td>
<td>81</td>
<td>81</td>
<td></td>
<td>33.82%</td>
<td>3.84</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>72</td>
<td>72</td>
<td></td>
<td>52.99%</td>
<td>3.96</td>
<td></td>
</tr>
<tr>
<td>(Other)</td>
<td>2</td>
<td>2</td>
<td></td>
<td>100.00%</td>
<td>4.08</td>
<td></td>
</tr>
</tbody>
</table>
User demographics (1 Jan 18 - 13 Jun 18)

Gender
- Male: 75.8% of total users
- Female: 24.2% of total users

Age
- 18-24: 23.9% of total users
- 25-34: 75.8% of total users
- 35-44: 10.3% of total users
- 45-54: 6.8% of total users
- 55-64: 5.3% of total users
- 65+: 2.37% of total users

46.50% of total users
<table>
<thead>
<tr>
<th>Country</th>
<th>Users</th>
<th>% Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>163</td>
<td>15.52%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>107</td>
<td>10.19%</td>
</tr>
<tr>
<td>Switzerland</td>
<td>88</td>
<td>8.38%</td>
</tr>
<tr>
<td>Germany</td>
<td>85</td>
<td>8.10%</td>
</tr>
<tr>
<td>France</td>
<td>82</td>
<td>7.81%</td>
</tr>
<tr>
<td>Canada</td>
<td>52</td>
<td>4.95%</td>
</tr>
<tr>
<td>Brazil</td>
<td>47</td>
<td>4.48%</td>
</tr>
<tr>
<td>India</td>
<td>45</td>
<td>4.29%</td>
</tr>
<tr>
<td>Italy</td>
<td>32</td>
<td>3.05%</td>
</tr>
<tr>
<td>Portugal</td>
<td>25</td>
<td>2.38%</td>
</tr>
</tbody>
</table>

Worldwide use (1 Jan 18 - 13 Jun 18)
Worldwide use

Mar 1, 2017 - Feb 18, 2018

Switzerland
Users: 297
Active users (18 Feb 17 - 18 Feb 18)
<table>
<thead>
<tr>
<th>Country</th>
<th>Users</th>
<th>Sessions</th>
<th>Pages / Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>462 (17.61%)</td>
<td>957 (15.86%)</td>
<td>3.20</td>
</tr>
<tr>
<td>Switzerland</td>
<td>297 (11.32%)</td>
<td>771 (12.78%)</td>
<td>3.32</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>286 (10.90%)</td>
<td>548 (9.08%)</td>
<td>3.93</td>
</tr>
<tr>
<td>Germany</td>
<td>238 (9.07%)</td>
<td>625 (10.36%)</td>
<td>3.45</td>
</tr>
<tr>
<td>Italy</td>
<td>110 (4.19%)</td>
<td>289 (4.79%)</td>
<td>3.76</td>
</tr>
<tr>
<td>France</td>
<td>101 (3.85%)</td>
<td>126 (2.09%)</td>
<td>3.32</td>
</tr>
<tr>
<td>Brazil</td>
<td>100 (3.81%)</td>
<td>276 (4.57%)</td>
<td>3.58</td>
</tr>
<tr>
<td>Canada</td>
<td>88 (3.35%)</td>
<td>271 (4.49%)</td>
<td>2.87</td>
</tr>
<tr>
<td>India</td>
<td>72 (2.74%)</td>
<td>129 (2.14%)</td>
<td>3.71</td>
</tr>
<tr>
<td>Spain</td>
<td>68 (2.59%)</td>
<td>150 (2.49%)</td>
<td>3.12</td>
</tr>
</tbody>
</table>
User activity (1 Mar 17 - 18 Feb 18)

Users: 2,534
New Users: 2,460
Sessions: 6,035
Number of Sessions per User: 2.38

Pageviews: 19,562
Pages / Session: 3.24
Avg. Session Duration: 00:04:15
Bounce Rate: 43.96%
Other Open Data users

Lab course
California

Lab course
Montréal

Lab course
Lund

Lab course
Maastricht

Summer students
Michigan

Lab course
US high school

Lab course
(Portugal)

Conference
Creative Commons

Lab course
Trieste

Lab course
Valencia
• Makes accelerator physics accessible to blind & partially sighted learners
• Hope to use Open Data with Tactile Collider
• Could feel difference between signal & background
• Article on interview about ATLAS Open Data
• Making data and tools available for the world to see: Arturo Sanchez of CERN on why ATLAS uses CC0 data
• Appeared on Creative Commons website November 2016 (soon after 8 TeV dataset release)
• Successful in reaching non-physics audience interested in Open Data & software (potential users)
• Want to follow right licenses & procedures to release Open Source resources