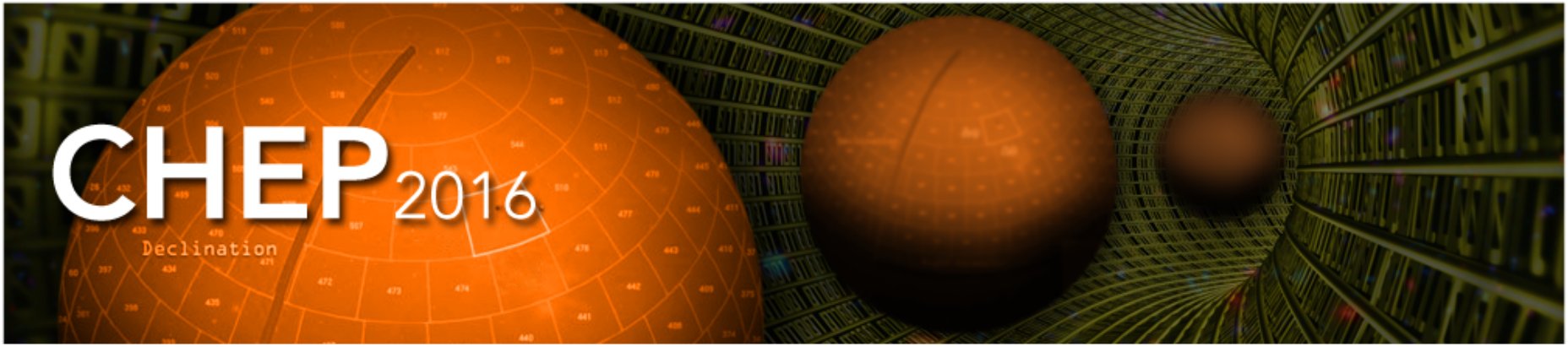
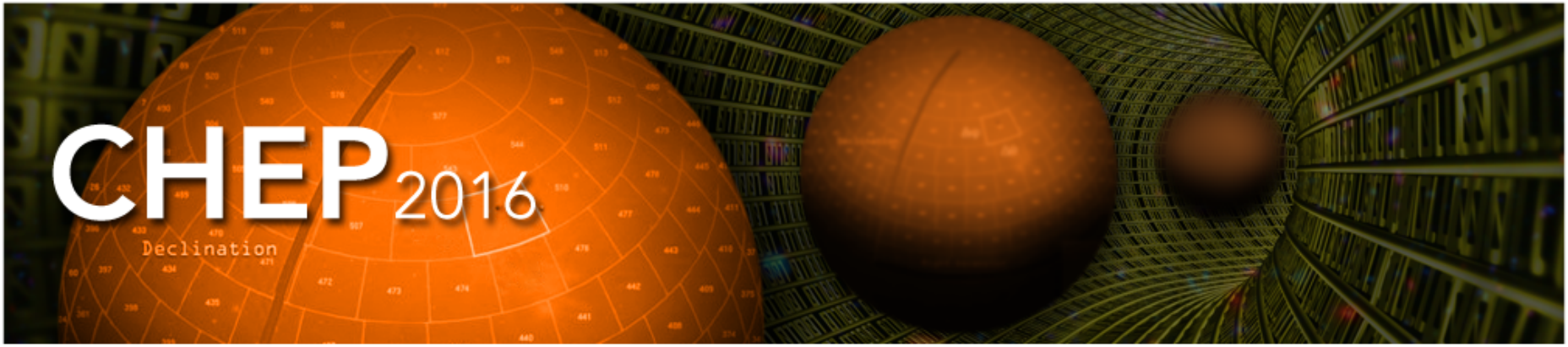


22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

Security, Policy and Outreach

Track 8 Highlights

David South, DESY
Torre Wenaus, BNL
Hannah Short, CERN



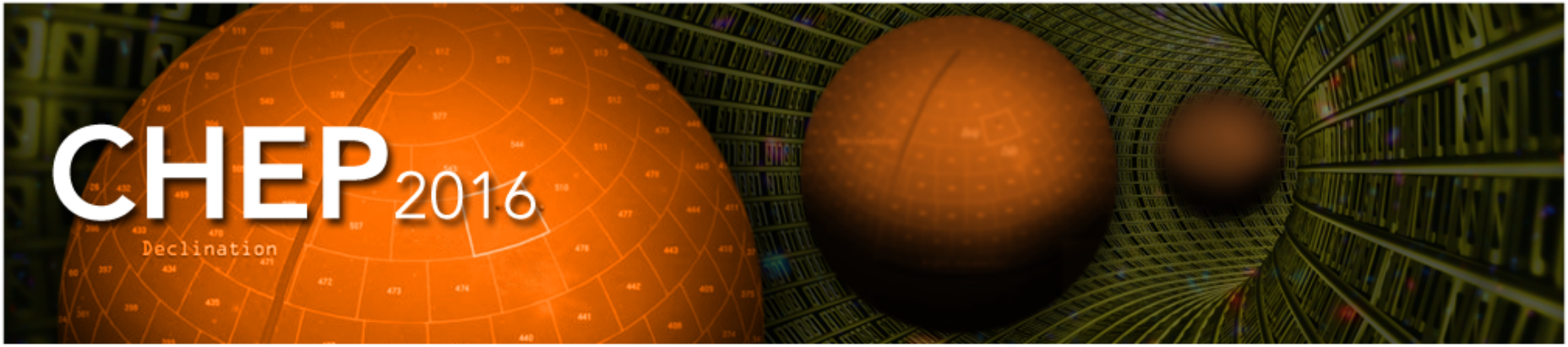
22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

Security, Reproducibility, Outreach, Authentication and Authorisation, Collaborative Tools…

A.K.A. Misc.


17 talks
Spread across 3 sessions
22 posters
On display throughout the week


Thank you to all our
contributors!



22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

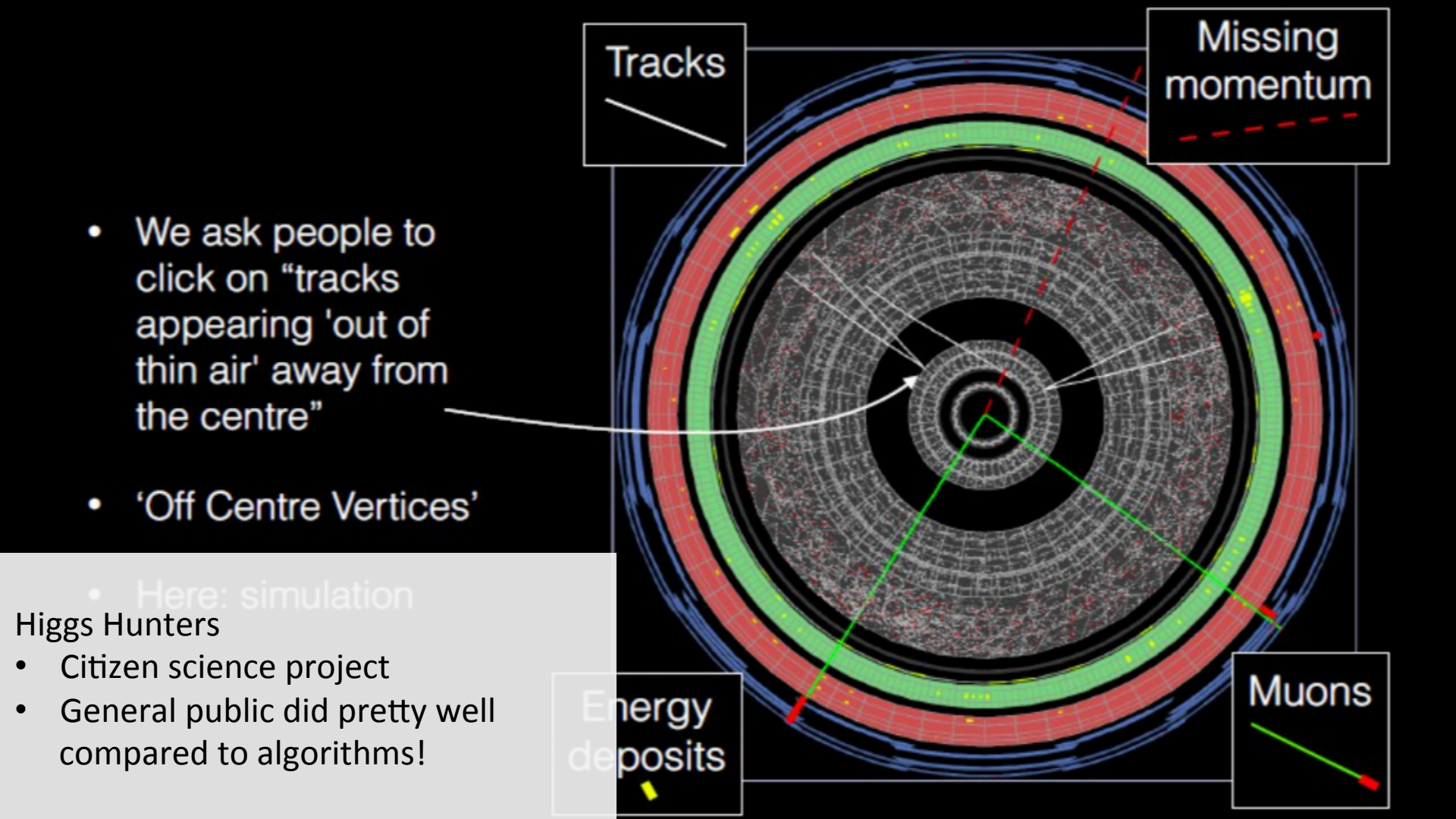
Within HEP, Long Tail of Science & Citizen Science

OUTREACH

CERN (2000)

CERN Virtual Visits

- Bringing CERN tours to the extended public
- Next step is CERN wide coordination

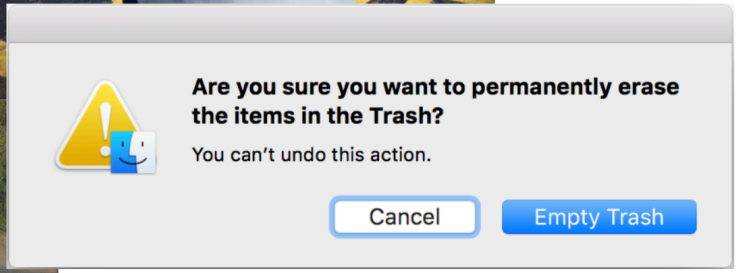


- We ask people to click on “tracks appearing 'out of thin air' away from the centre”
- ‘Off Centre Vertices’
- Here: simulation

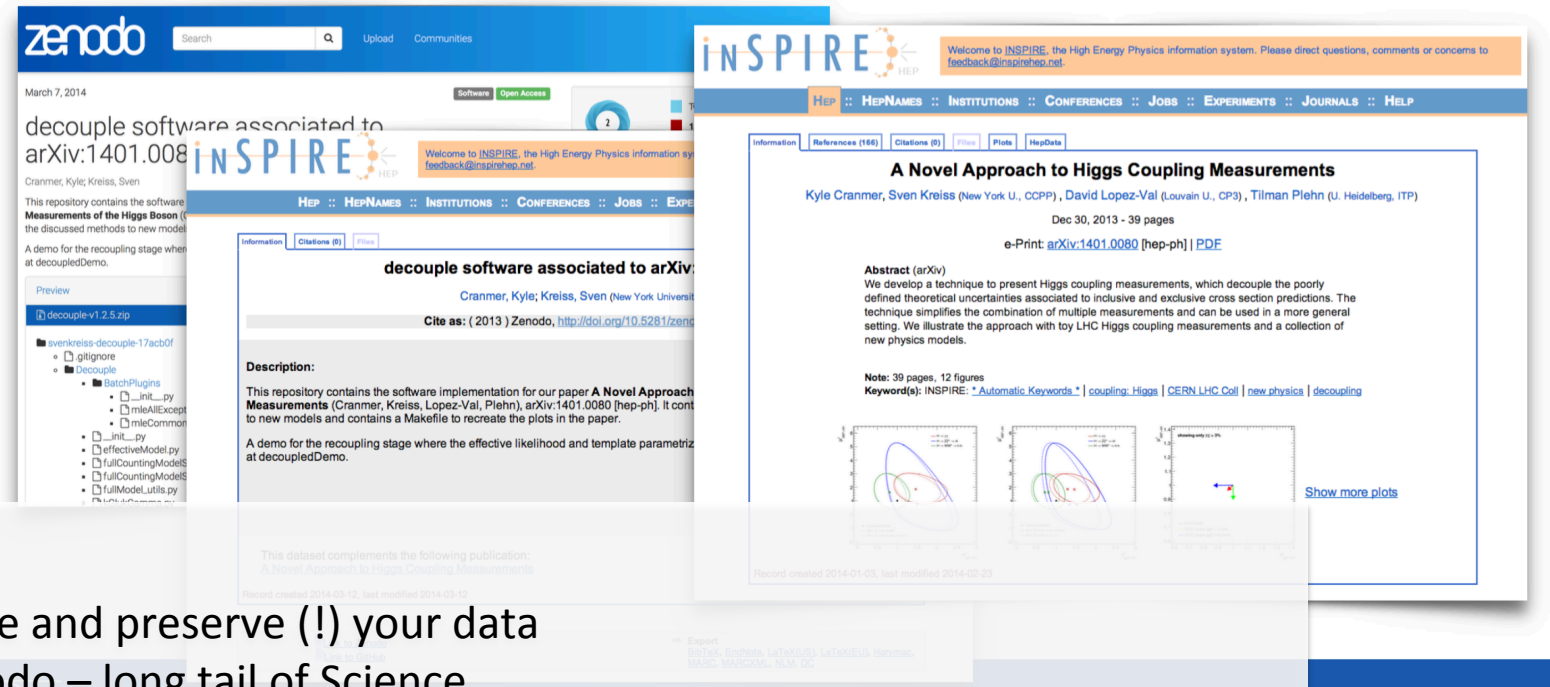
Higgs Hunters

- Citizen science project
- General public did pretty well compared to algorithms!


DRUCE
www.druce.com
Very important!
LAPTOP LOST
in the bus 345
CRUCIAL scientific data
+ many YEARS of
research work inside!


Are you sure you want to permanently erase the items in the Trash?
You can't undo this action.
Cancel
Empty Trash

Software meets INSPIRE



Invenio

- Share and preserve (!) your data
- Zenodo – long tail of Science
- CERN Open Data – including outreach projects
- DOIs registered and permanent, helping InSPIRE to identify data sets

The solution

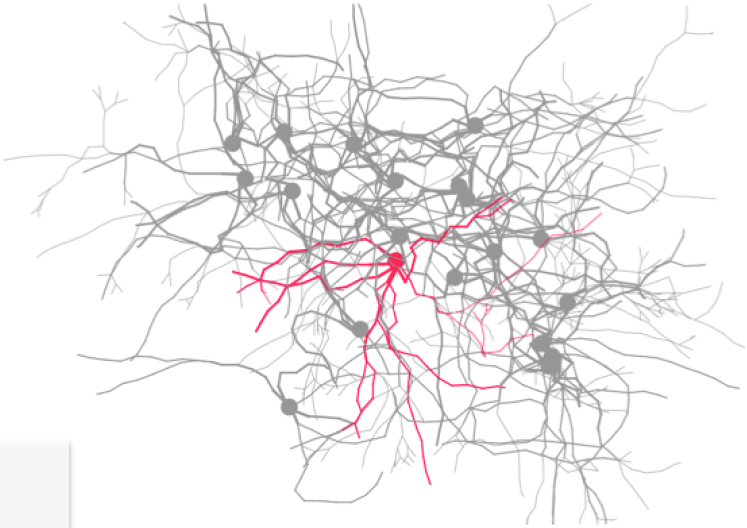
- The **Starterkit** team provides...
  - Online **tutorials**
  - Interactive **workshops**
- Goals
  - Improve software **literacy**
  - Teach **good practices**
  - **Socialisation** amongst collaboration members

LHCb Starterkit

- Teach computing essentials for new joiners
- Interest in reusing the content at other experiments

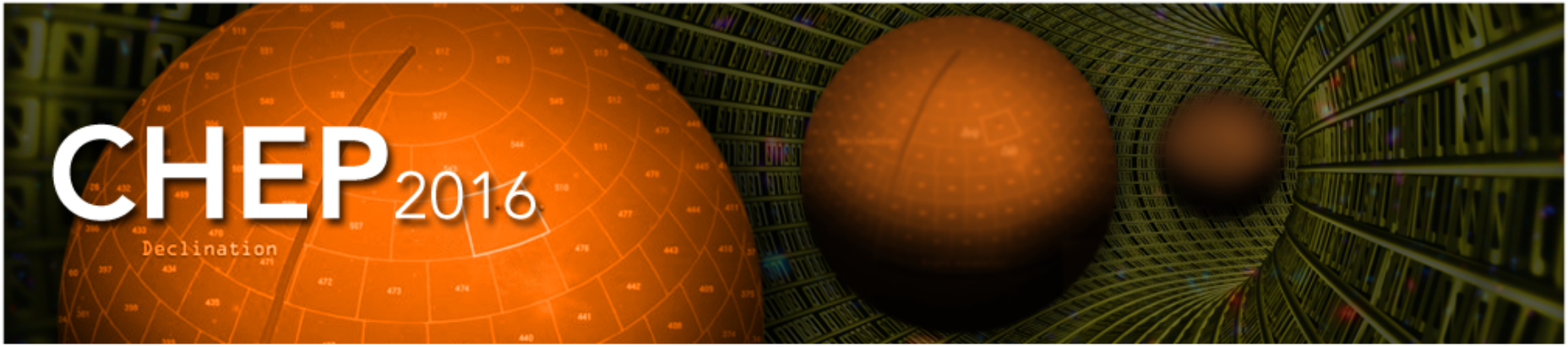
Why Simulate Brain Development

- Neuro scientific insights
  - How does the brain develop?
  - How do genetic and environmental cues interact?
  - How do neurons and brain regions communicate?
  - How does cognition, planning and memory work?
- Medicine
  - Understanding of brain diseases (epilepsy)
  - Tumor growth
  - Drug development



CERNopenlab Knowledge Transfer

- Sharing HEP expertise with non-HEP domains
- Genomics and Brain Simulation

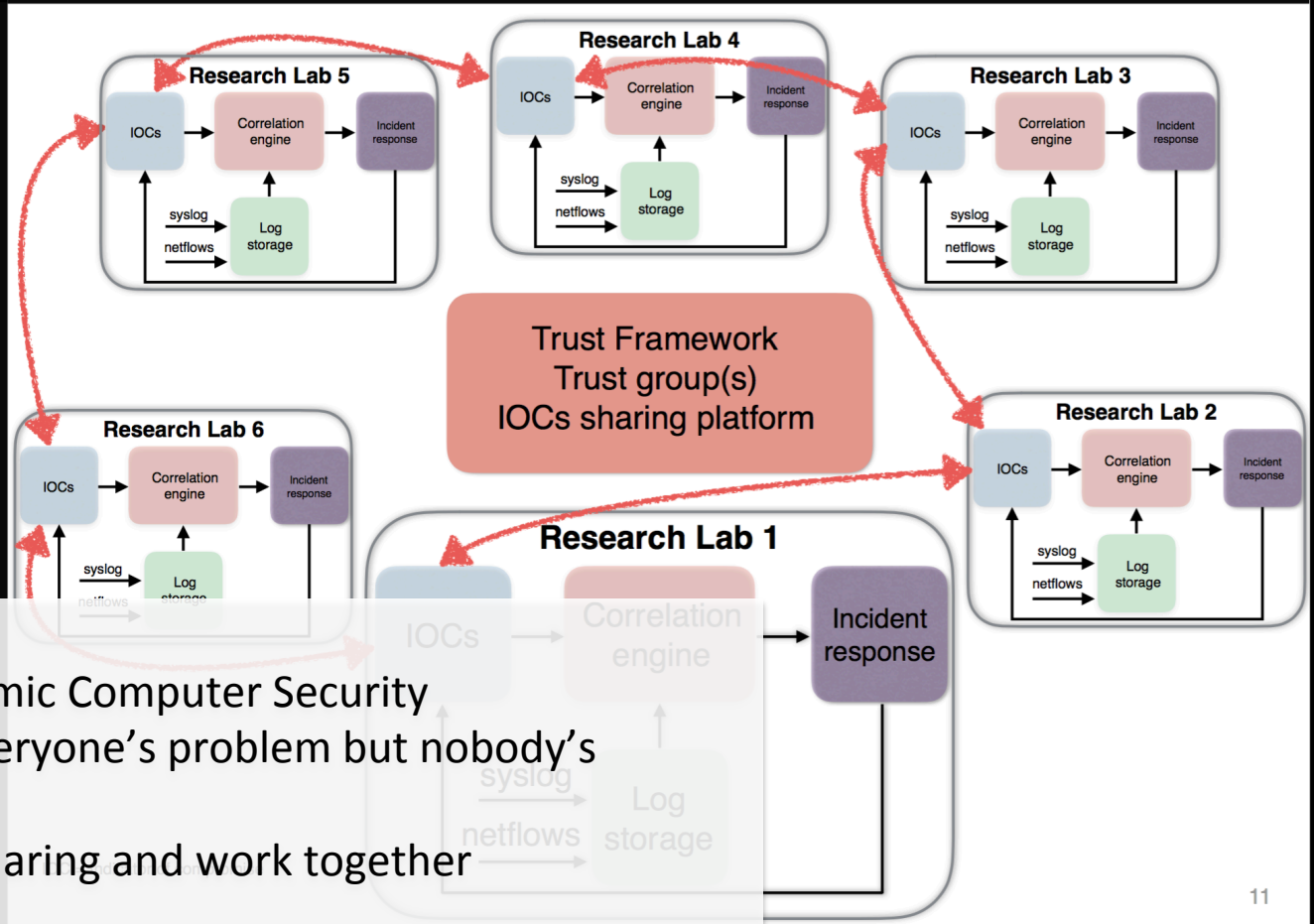


22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

Security Awareness, Future Landscape

SECURITY

A global response



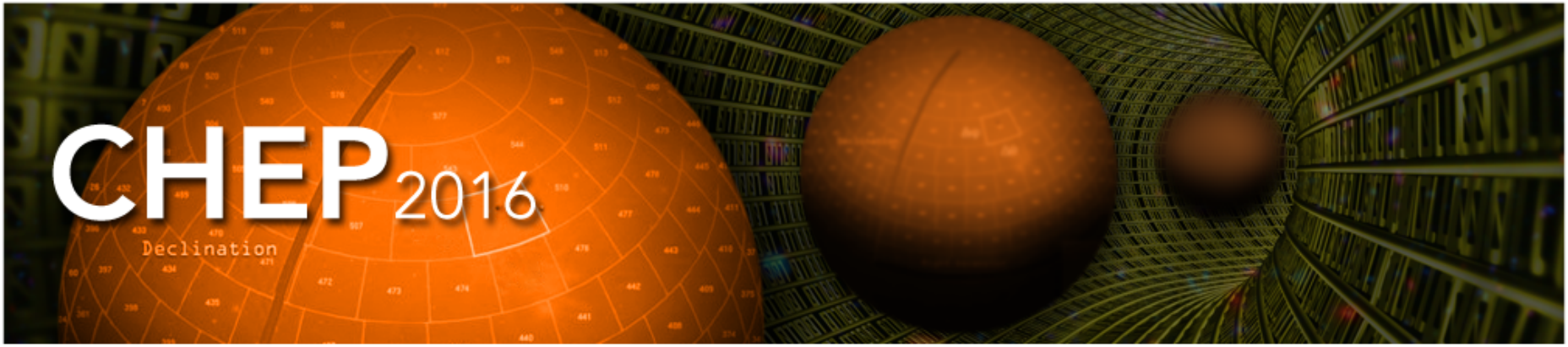
Future of Academic Computer Security

- Security is everyone's problem but nobody's responsibility
- Enable IOC sharing and work together

“We are there to help you”

CERN Security Consulting

- Improving communication between security and developers before poor software decisions are made
- Involving users by teaching them penetration testing, Whitehats

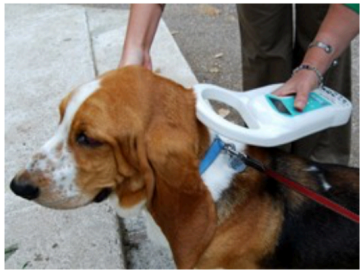


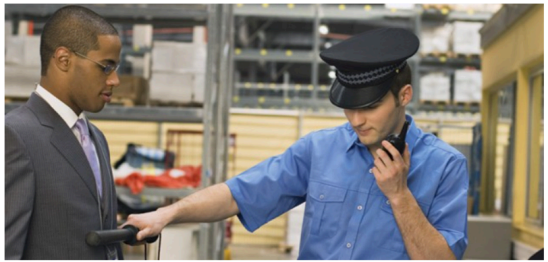
22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

Moving away from certificate based AUTHN

AUTHENTICATION AND AUTHORISATION

Quick recap

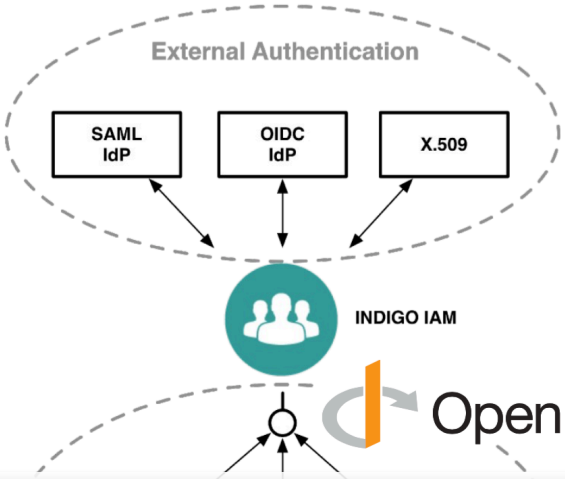


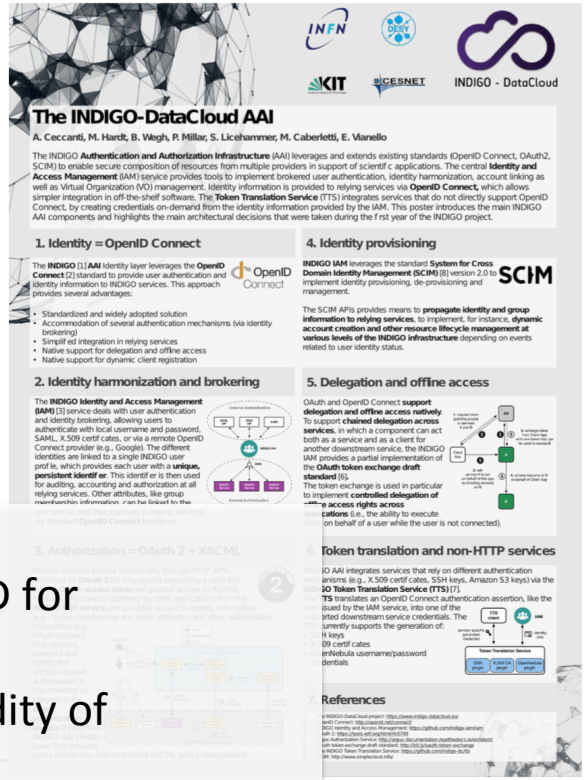


Authn

Authz

INDIGO-DataCloud AAI framework

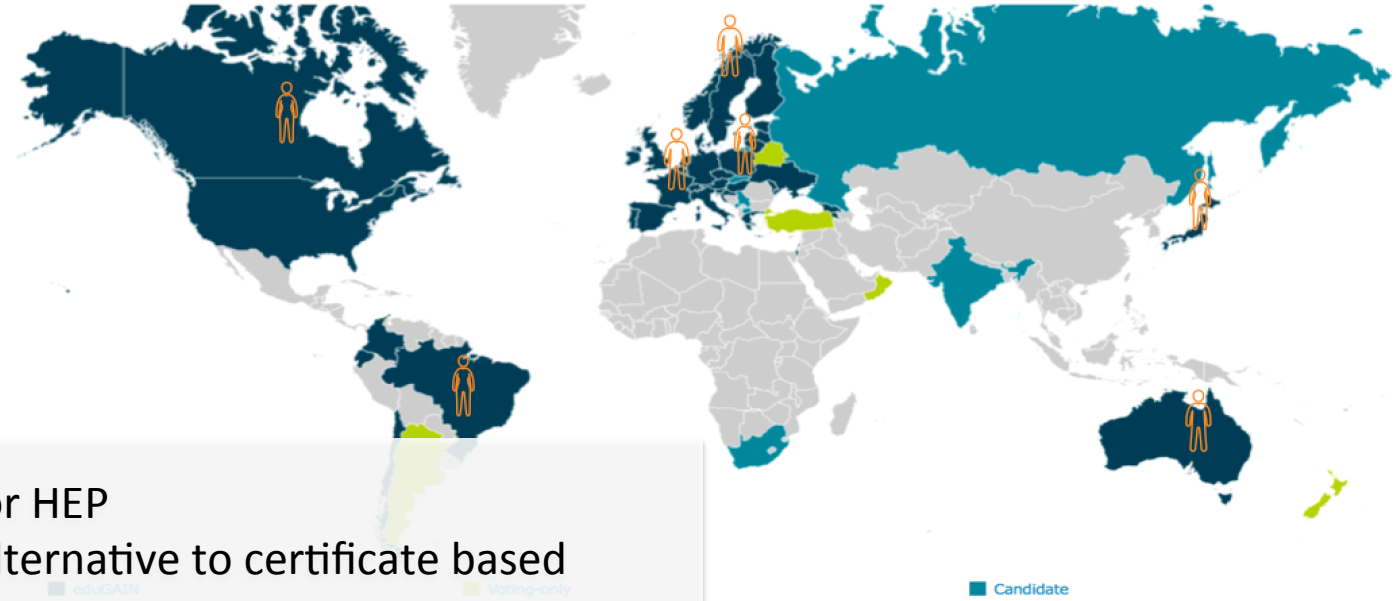




Federated Identities and Anonymized Delegation

- Funnel identities into a central IdP and use OpenID for authentication and delegation under the hood
- “Macaroons” (like cookies but better) control validity of authorisation tokens

An alternative source of identities, federated identity via eduGAIN

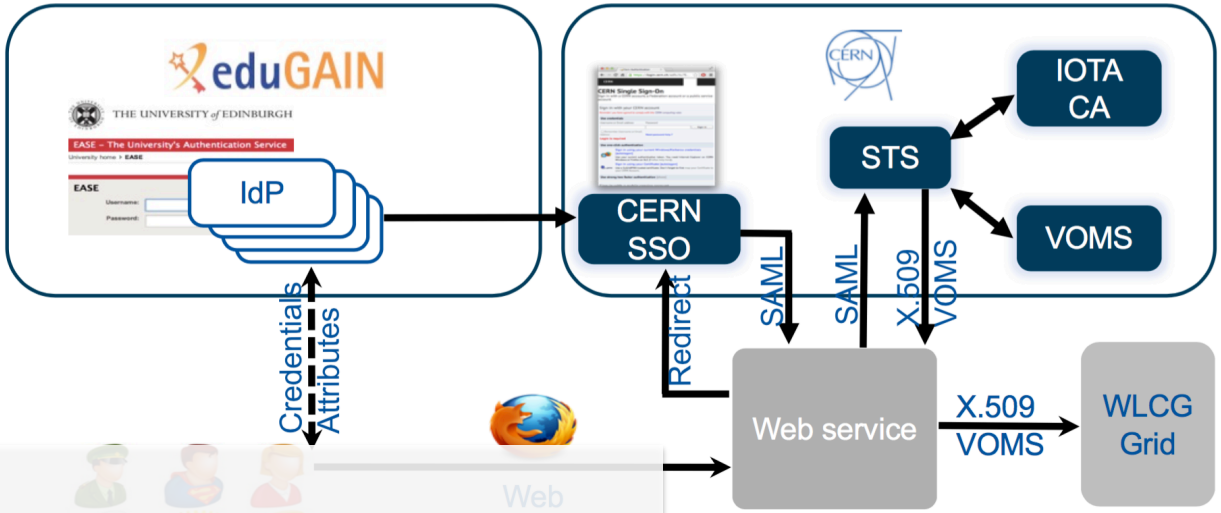


Source: eduGAIN, data taken March 2016

eduGAIN for HEP

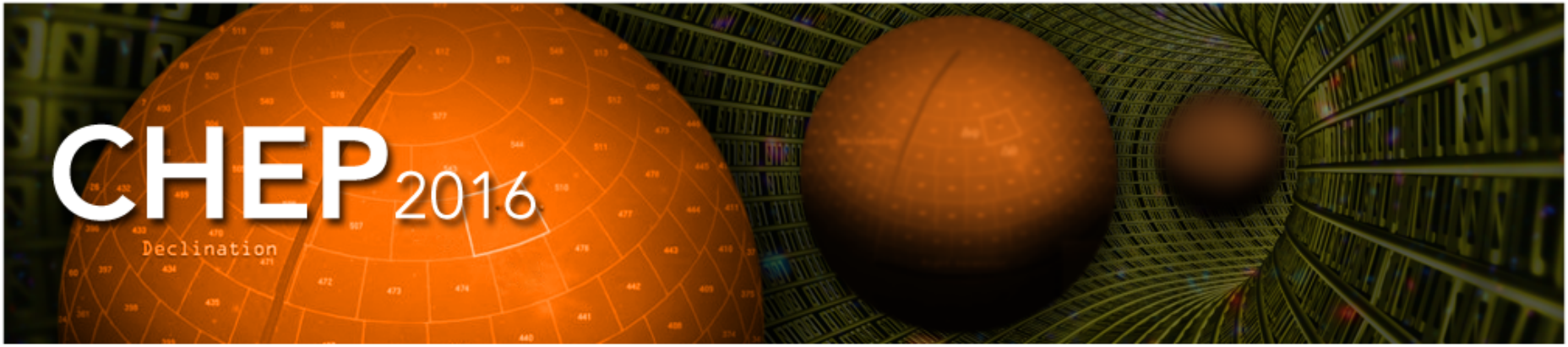
- Viable alternative to certificate based federation
- Work to ensure that eduGAIN is as secure as current access methods

How can I implement X509 free?



X509 Free Pilot

- Transparent eduGAIN access to grid services
- Ready to go for some Vos
- Currently being implemented for PanDA



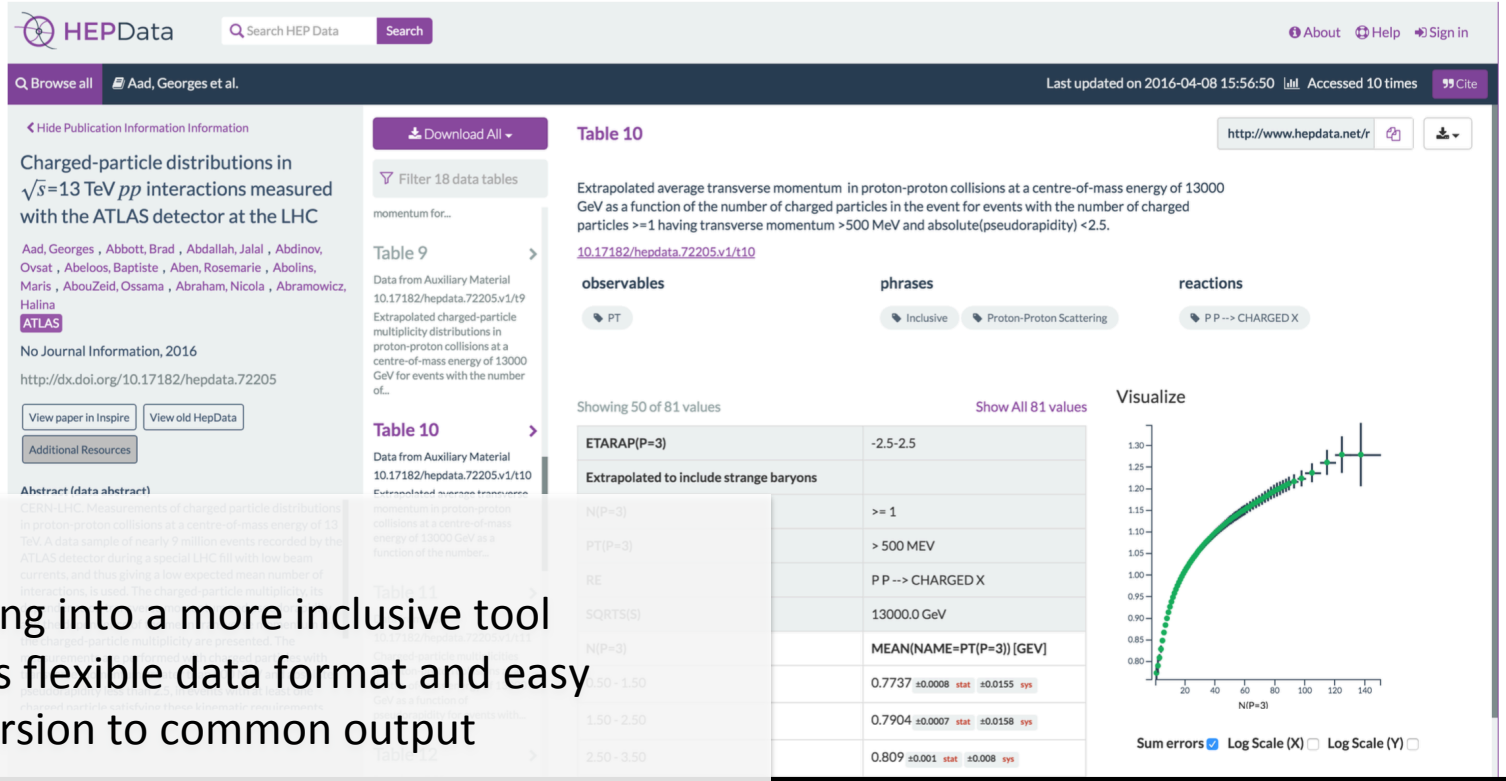
22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

How are we sharing knowledge, and enabling collaboration?

COLLABORATIVE TOOLS

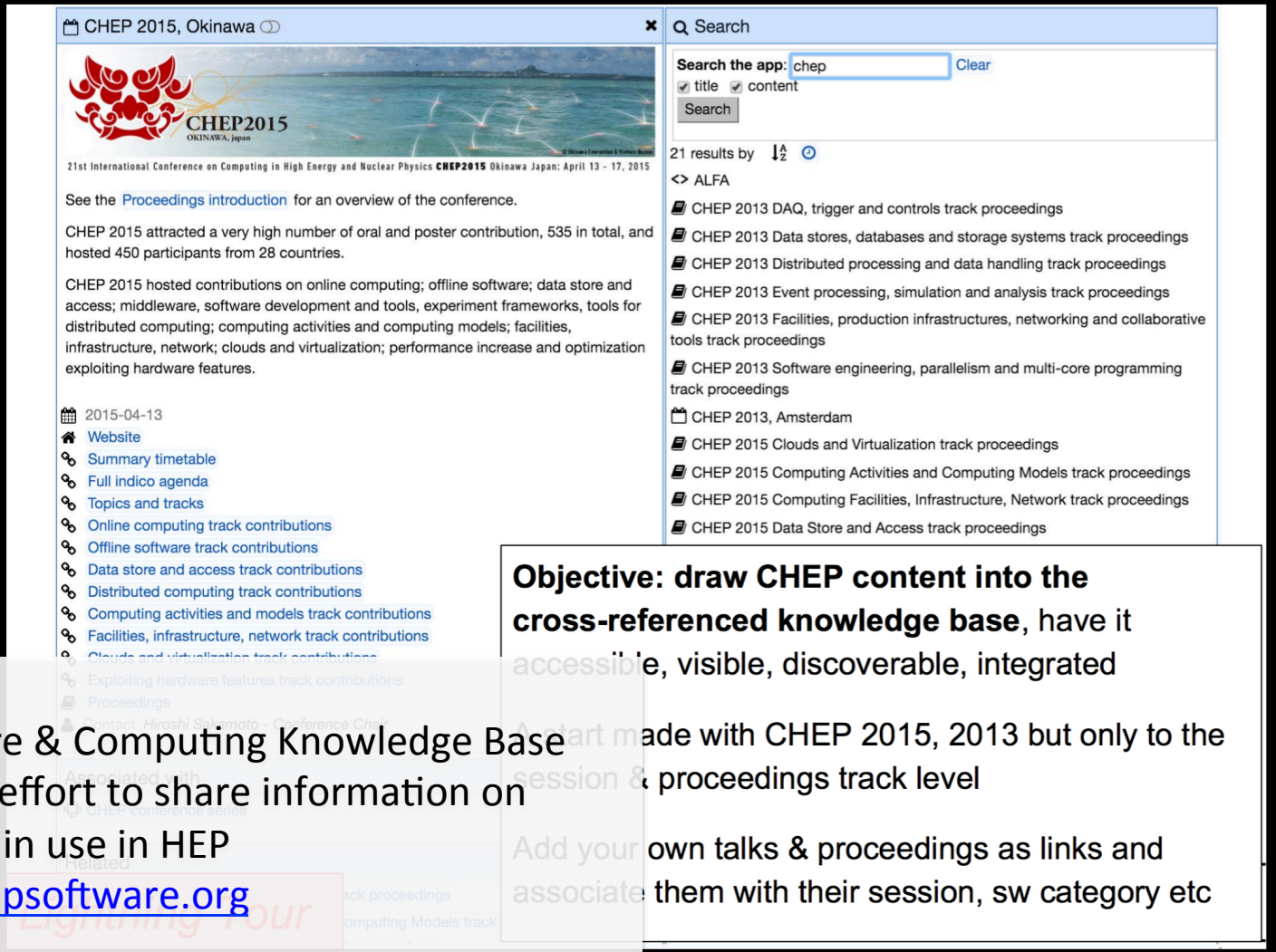
HEPData

Detailed view with quick access to all information



HEPData

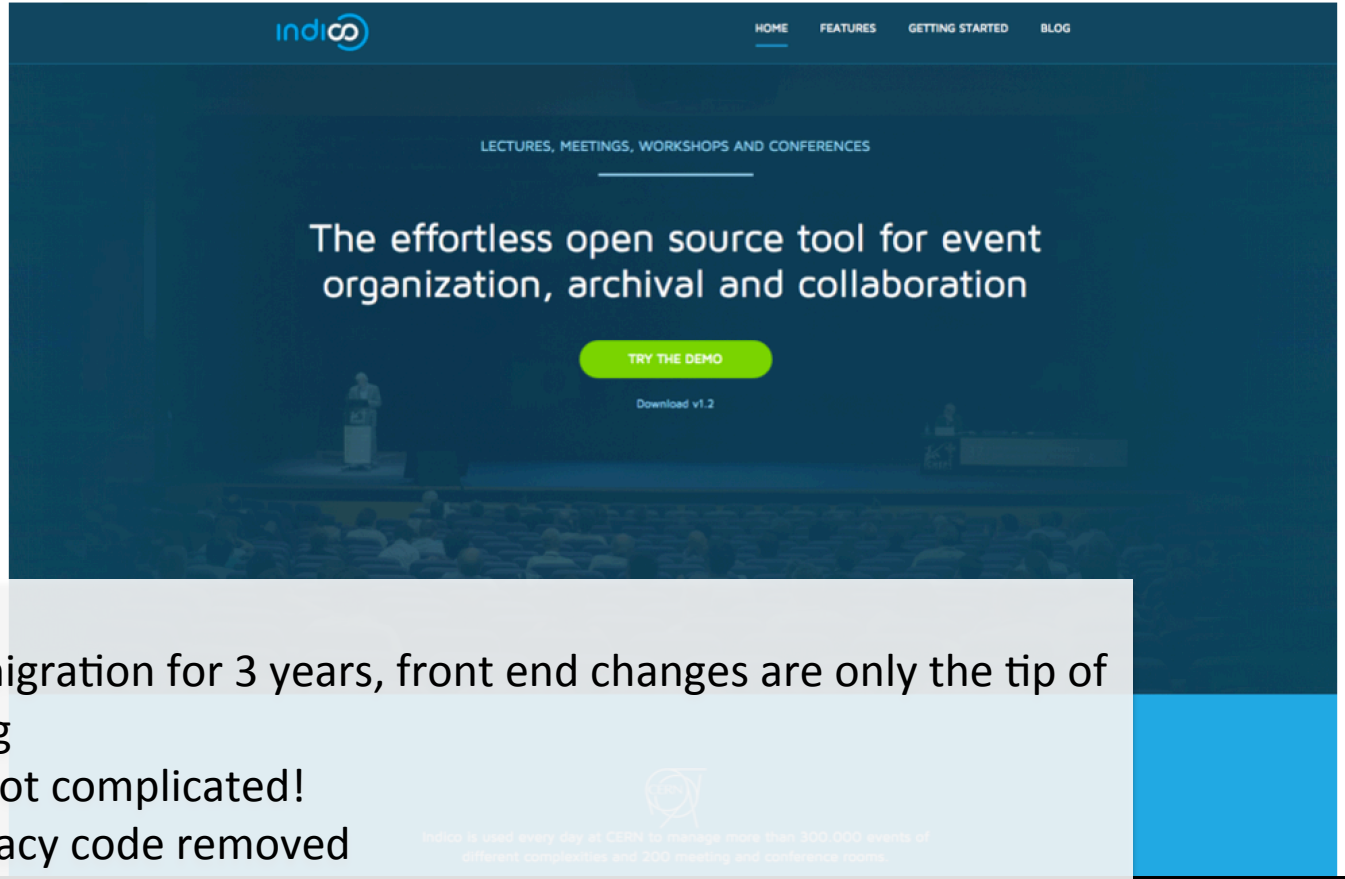
- Evolving into a more inclusive tool
- Allows flexible data format and easy conversion to common output



HEP Software & Computing Knowledge Base

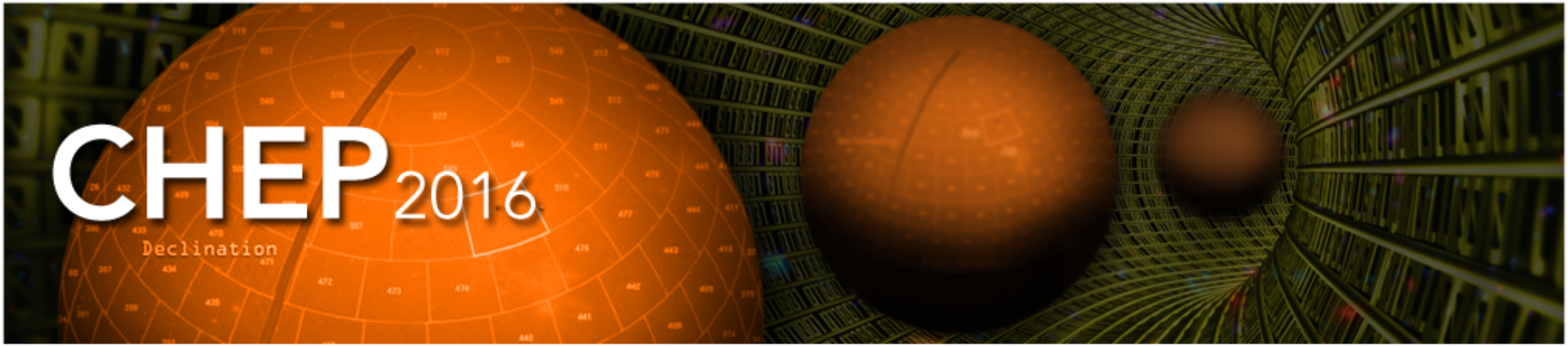
- Ongoing effort to share information on software in use in HEP
- http://hepsoftware.org

http://indico-software.org



Indico 2.0

- Ongoing migration for 3 years, front end changes are only the tip of the iceburg
- Complex not complicated!
- 78% of legacy code removed



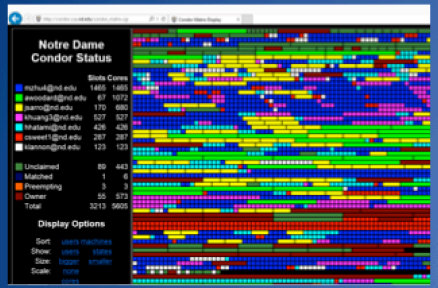
22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

Can we get the same result twice, 10 times, 100 times?

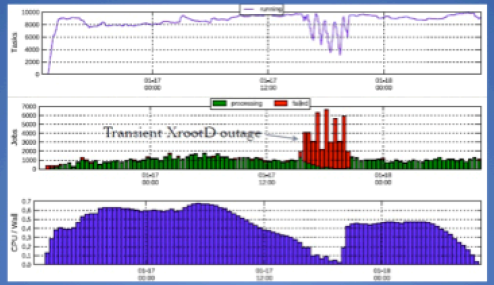
REPRODUCIBILITY

What we usually work on:

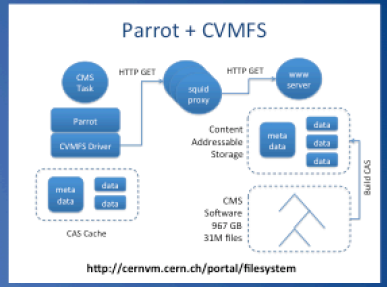
HTCondor



Lobster Data Analysis



Global Filesystems



Today, a different question:

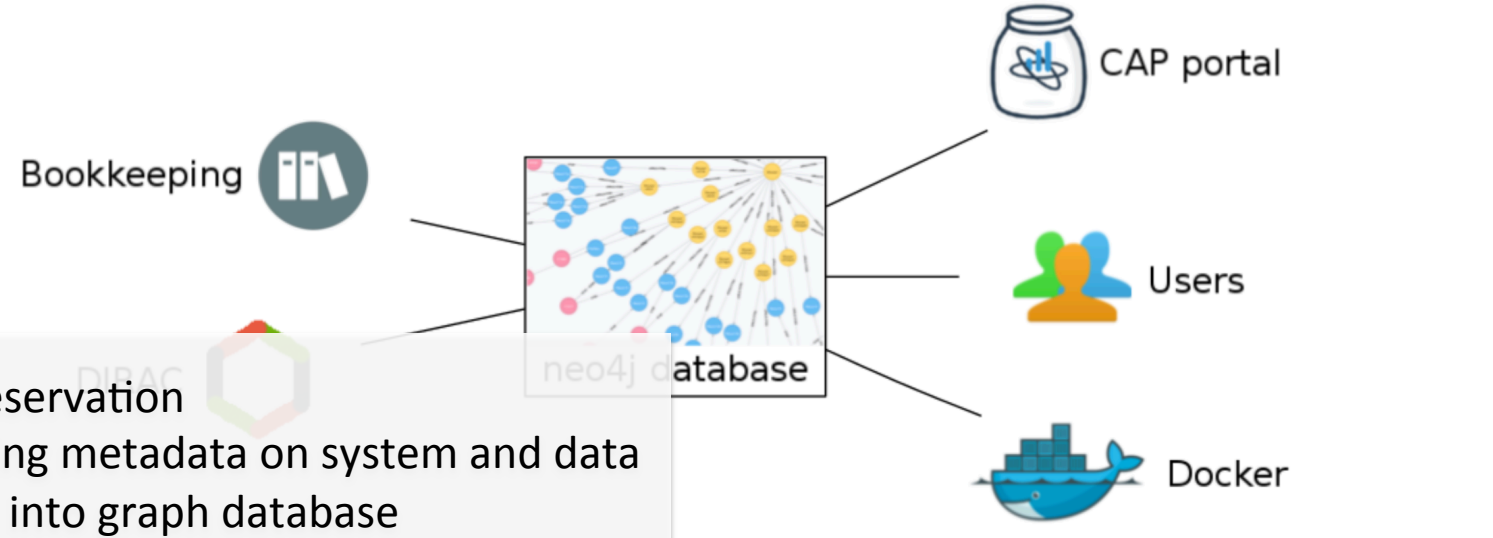
What happens if we attempt to run the same thing twice?

Reproducibility

- Where do differences come from? Non-deterministic code
- ROOT-diff tool developed to do bitwise comparison
- By artificially warping machine time, can get bitwise equality… to a point

The big picture

Providing the information on data production and the software, and linking the monitoring system Dirac and bookkeping to the CERN analysis preservation (CAP) portal, users and containerised software in Docker

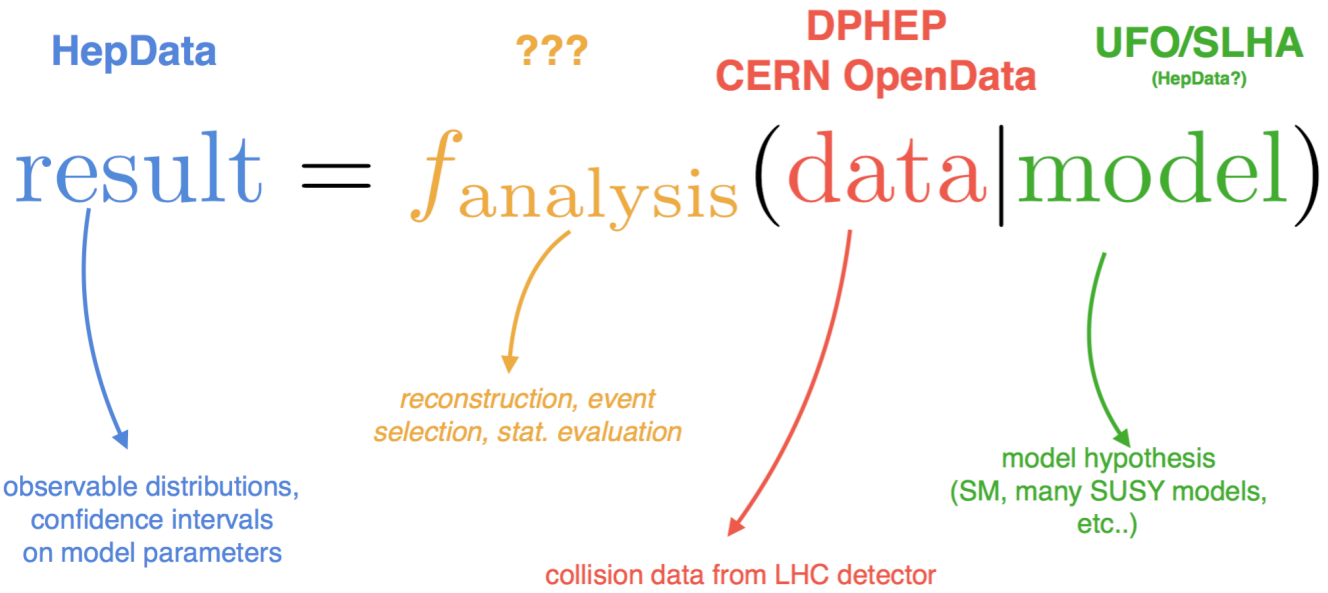


LHCb Data Preservation

- Consolidating metadata on system and data production into graph database
- Can mine graph for statistics on software usage and help when troubleshooting

Current Standards / Preservation / Archival efforts:

Efforts like DASPOS and CERN Analysis Preservation try to fill a gap in the preservation efforts.
Data only is not enough.We need software and environment as well.

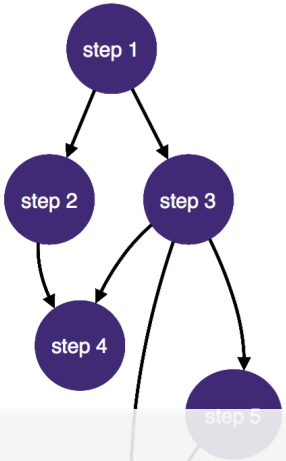


How to preserve $f_{\text{analysis}}(\cdot)$?

2. Problem: Preserve Parametrized Workflow

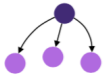
Therefore: Sequentially build up graph, as sufficient information becomes available, using a number of stages that add nodes and edges

To capture analysis workflow, capture the stages.

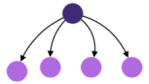


Example: Parametrized Map-Reduce

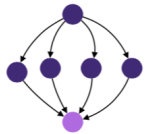
Stage 1:
unknown number of files. e.g. download & unpack archive with a priori unknown # of files



Stage 2:
for each file in the archive, add node to process it
(only possible after first node done)



Stage 3:
add a node that merges results of the map nodes
node/edge can be added before execution of map nodes



- To capture analysis workflow, capture the individual stages
- Preserve in a parameterized way; source code is not enough, need full environment


And the posters…

CERN Document Server

Towards the Next Generation CERN Institutional Repository

Retrieval

The new version of CDS significantly improves the information retrieval system. By switching to Elasticsearch as the underlying search engine and redesigning our complex data model, the new CDS provides up to 3 times faster search results and up to 50 times faster indexing. Another determining change is a completely new design that offers a clearer, more intuitive way of searching, and real time filtering and previewing of data.

Submission

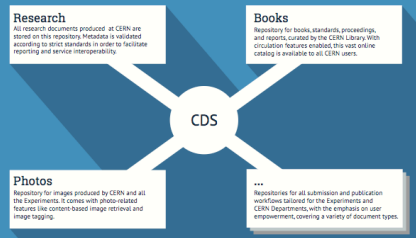
The CDS submission system has undergone a complete redesign based on our extensive experience working with different communities at CERN. The new version, based on a flexible data model, offers the possibility to create tailored content for different user communities. Together with customizable submission and publication workflows comes a new, richer user interface and significant speed improvements that enhance user experience.

Community driven

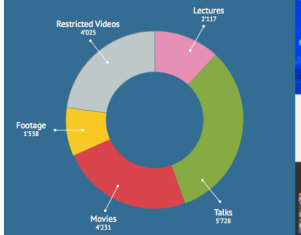
The new CDS aims to be the CERN's document hub, acting as an aggregator over specialized repositories, each having its own software stack, with features enabled based on the repository's content.

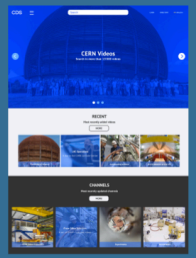
The aim is to enable each content producer community to have its own identity, both visually and functionally, as well as increased control on the data model and the submission, curation, management, and dissemination of the data.

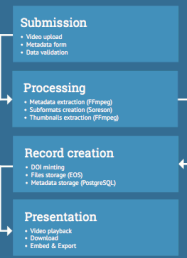
powered by INVENIO



Videos



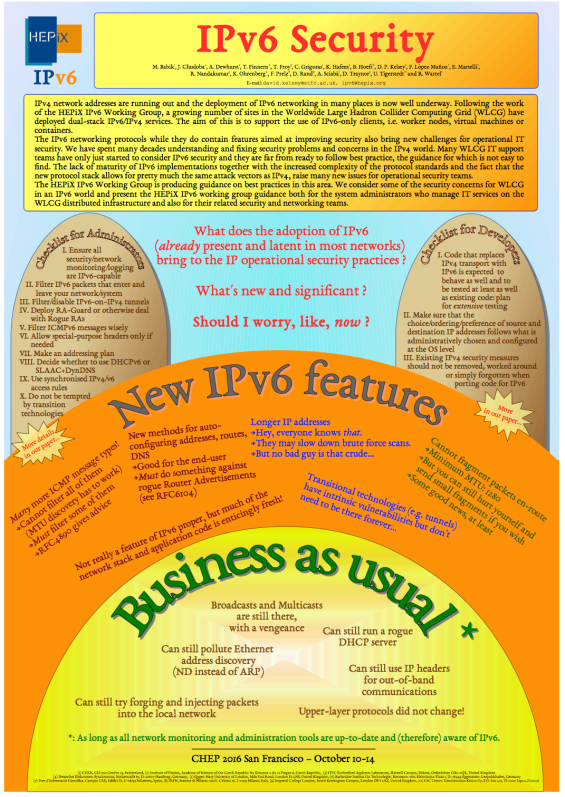




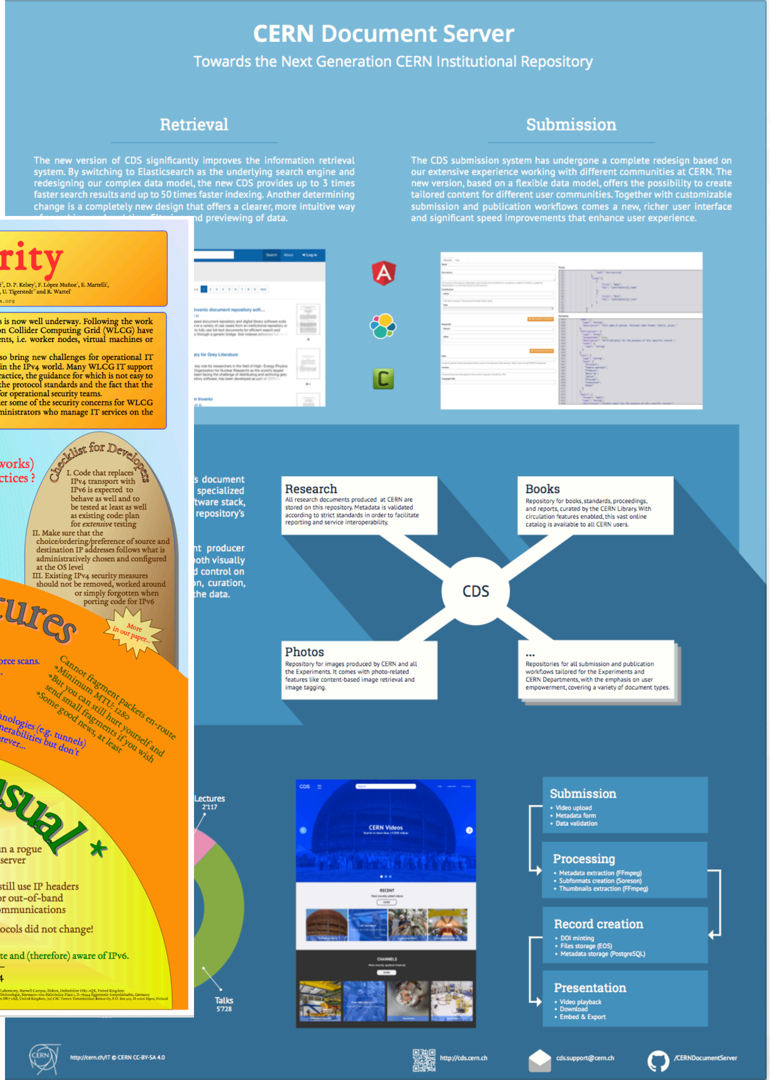
http://cds.cern.ch

cds.support@cern.ch

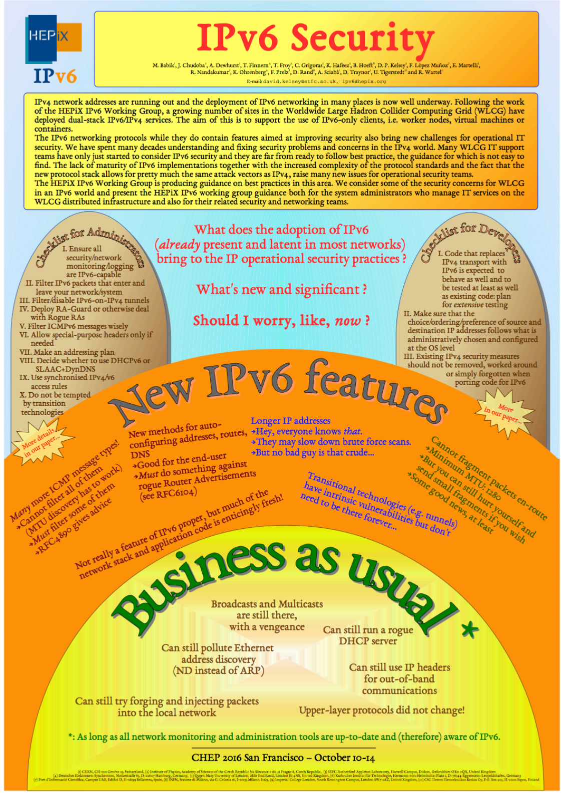
/CERNDocumentServer



HEPiX
IPv6
IPv6 Security
IPv4 network addresses are running out and the deployment of IPv6 networking in many places is now well underway. Following the work of the HEPiX IPv6 Working Group, a growing number of sites in the Worldwide Large Hadron Collider Computing Grid (WLCG) have deployed dual-stack IPv6/IPv4 services. The aim of this is to support the use of IPv6-only clients, i.e. worker nodes, virtual machines or containers.
The IPv6 networking protocols while they do contain features aimed at improving security also bring new challenges for operational IT security. We have spent many decades understanding and fixing security problems and concerns in the IPv4 world. Many WLCG IT support teams have only just started to consider IPv6 security and they are far from ready to follow best practice, the guidance for which is not easy to find. The lack of maturity of IPv6 implementations together with the increased complexity of the protocol standards and the fact that the new protocol stack allows for pretty much the same attack vectors as IPv4, raise many new issues for operational security teams.
The HEPiX IPv6 Working Group is producing guidance on best practices in this area. We consider some of the security concerns for WLCG in an IPv6 world and present the HEPiX IPv6 working group guidance both for the system administrators who manage IT services on the WLCG distributed infrastructure and also for their related security and networking teams.
Checklist for Administrators
I. Ensure all security/network monitoring/logging are IPv6-capable
II. Filter IPv6 packets that enter and leave your network/system
III. Filter/disable IPv6-on-IPv4 tunnels
IV. Deploy RA-Guard or otherwise deal with Rogue RAs
V. Filter ICMPv6 messages wisely
VI. Allow special-purpose headers only if needed
VII. Make an addressing plan
VIII. Decide whether to use DHCPv6 or SLAAC+DynDNS
IX. Use synchronised IPv4/v6 access rules
X. Do not be tempted by transition technologies
What does the adoption of IPv6 (already present and latent in most networks) bring to the IP operational security practices ?
What's new and significant ?
Should I worry, like, now ?
Checklist for Developers
I. Code that replaces IPv4 transport with IPv6 is expected to behave as well and to be tested at least as well as existing code: plan for extensive testing
II. Make sure that the choice/ordering/preference of source and destination IP addresses follows what is administratively chosen and configured at the OS level
III. Existing IPv4 security measures should not be removed, worked around or simply forgotten when porting code for IPv6
More in our paper...
New IPv6 features
More details in our paper...
New methods for auto-configuring addresses, routes, DNS
→Good for the end-user
→Must do something against rogue Router Advertisements (see RFC6104)
Longer IP addresses
→Hey, everyone knows that.
→They may slow down brute force scans.
→But no bad guy is that crude...
Many more ICMP message types!
→Cannot filter all of them (MTU discovery has to work)
→Must filter some of them
→RFC4890 gives advice
Transitional technologies (e.g. tunnels) have intrinsic vulnerabilities but don't need to be there forever...
Cannot fragment packets en-route
→Minimum MTU: 1280
→But you can still hurt yourself and send small fragments if you wish
→Some good news, at least
Not really a feature of IPv6 proper, but much of the network stack and application code is enticingly fresh!
Business as usual *
Broadcasts and Multicasts are still there, with a vengeance
Can still run a rogue DHCP server
Can still pollute Ethernet address discovery (ND instead of ARP)
Can still use IP headers for out-of-band communications
Can still try forging and injecting packets into the local network
Upper-layer protocols did not change!
*: As long as all network monitoring and administration tools are up-to-date and (therefore) aware of IPv6.
CHEP 2016 San Francisco – October 10-14



CERN Document Server
Towards the Next Generation CERN Institutional Repository
Retrieval
The new version of CDS significantly improves the information retrieval system. By switching to Elasticsearch as the underlying search engine and redesigning our complex data model, the new CDS provides up to 3 times faster search results and up to 50 times faster indexing. Another determining change is a completely new design that offers a clearer, more intuitive way
nd previewing of data.
Submission
The CDS submission system has undergone a complete redesign based on our extensive experience working with different communities at CERN. The new version, based on a flexible data model, offers the possibility to create tailored content for different user communities. Together with customizable submission and publication workflows comes a new, richer user interface and significant speed improvements that enhance user experience.
Research
All research documents produced at CERN are stored on this repository. Metadata is validated according to strict standards in order to facilitate reporting and service interoperability.
Books
Repository for books, standards, proceedings, and reports, curated by the CERN Library. With circulation features enabled, this vast online catalog is available to all CERN users.
CDS
Photos
Repository for images produced by CERN and all the Experiments. It comes with photo-related features like content-based image retrieval and image tagging.
...
Repositories for all submission and publication workflows tailored for the Experiments and CERN Departments, with the emphasis on user empowerment, covering a variety of document types.
Lectures
2'117
Talks
5'728
CERN Videos
Submission
• Video upload
• Metadata form
• Data validation
Processing
• Metadata extraction (FFmpeg)
• Subformats creation (Sorenson)
• Thumbnails extraction (FFmpeg)
Record creation
• DOI minting
• Files storage (EOS)
• Metadata storage (PostgreSQL)
Presentation
• Video playback
• Download
• Embed & Export
http://cern.ch/IT © CERN CC-BY-SA 4.0
http://cds.cern.ch
cds.support@cern.ch
/CERNDocumentServer


HEPiX
IPv6
IPv6 Security
M. Babik, J. Chudoba, A. Dewhurst, T. Finnern, T. Froy, C. Grigoras, K. Hafeez, B. Hoeft, D. P. Kelsey, F. López Muñoz, E. Martelli, R. Nandakumar, K. Ohrenberg, F. Prelz, D. Rand, A. Sciabà, D. Traynor, U. Tigerstedt and R. Wartel
IPv4 network addresses are running out and the deployment of IPv6 networking in many places is now well underway. Following the work of the HEPiX IPv6 Working Group, a growing number of sites in the Worldwide Large Hadron Collider Computing Grid (WLCG) have deployed dual-stack IPv6/IPv4 services. The aim of this is to support the use of IPv6-only clients, i.e. worker nodes, virtual machines or containers.
The IPv6 networking protocols while they do contain features aimed at improving security also bring new challenges for operational IT security. We have spent many decades understanding and fixing security problems and concerns in the IPv4 world. Many WLCG IT support teams have only just started to consider IPv6 security and they are far from ready to follow best practice, the guidance for which is not easy to find. The lack of maturity of IPv6 implementations together with the increased complexity of the protocol standards and the fact that the new protocol stack allows for pretty much the same attack vectors as IPv4, raise many new issues for operational security teams.
The HEPiX IPv6 Working Group is producing guidance on best practices in this area. We consider some of the security concerns for WLCG in an IPv6 world and present the HEPiX IPv6 working group guidance both for the system administrators who manage IT services on the WLCG distributed infrastructure and also for their related security and networking teams.
What does the adoption of IPv6 (already present and latent in most networks) bring to the IP operational security practices?
What's new and significant?
Should I worry, like, now?
Checklist for Administrators
Checklist for Developers
New IPv6 features
Business as usual*
Broadcasts and Multicasts are still there, with a vengeance
Can still run a rogue DHCP server
Can still pollute Ethernet address discovery (ND instead of ARP)
Can still use IP headers for out-of-band communications
Can still try forging and injecting packets into the local network
Upper-layer protocols did not change!
*: As long as all network monitoring and administration tools are up-to-date and (therefore) aware of IPv6.
CHEP 2016 San Francisco – October 10-14


CERN Document Server
Towards the Next Generation CERN Institutional Repo
Retrieval
The new version of CDS significantly improves the information retrieval system. By switching to Elasticsearch as the underlying search engine and redesigning our complex data model, the new CDS provides up to 3 times faster search results and up to 50 times faster indexing. Another determining change is a completely new design that offers a clearer, more intuitive way
Research
All research documents produced at CERN are stored on this repository. Metadata is validated according to strict standards in order to facilitate reporting and service interoperability.
Photos
Repository for images produced by CERN and all the Experiments. It comes with photo-related features like content-based image retrieval and image tagging.
CDS
Lectures 2'117
Talks 5'728
Record creation
DOI minting
Files storage (EOS)
Metadata storage (PostgreSQL)
Presentation
Video playback
Download
Embed & Export
http://cern.ch/IT © CERN CC-BY-SA 4.0
http://cds.cern.ch
cds.support@cern.ch
/CERNDocumentServer


CERN
Flash is Dead. Finally.
M Domaracky, N Juszka, M Peksa, T Baron
For almost 10 years, CERN has been providing live webcast of events using Adobe Flash technology. This year is finally the year when flash died at CERN! With Flash being slowly phased out on most streaming platforms, the CERN streaming service moved as well from Flash to HTTP streaming.
THEOplayer
The biggest challenge for providing pure HTML5 video streaming goes with the support of different streaming protocols across browsers and OS platforms. Thanks to THEOPLayer (www.theoplayer.com) we are able to stick with the HTTP Live Streaming (HLS) protocol from Apple and play it on all modern browsers on desktops and mobile devices. We are able to deliver the same experience as we did with Adobe Flash based players. Our users can still enjoy video of the speaker synchronised with video of the presentation, so they have the same experience as sitting in the auditoria.
DVR – Digital Video Recorder
We will soon introduce DVR functionality for all our live webcasts. Users that arrived late on the webcast website, will have a possibility to go back to the beginning of the webcast or if they missed something they can seek back to watch it again. With the DVR feature we will also be able to publish the recording right after the webcast is finished.
Micala
Michigan and CERN Automated Lecture Archiving (Micala) is an automated tool developed at CERN with a collaboration from the University of Michigan. We improved the workflow to faster publish the recorded lecture by integrating the Sorenson (http://www.sorensonmedia.com/squeeze/squeeze-server/) transcoding cluster to Micala. The communication with users is now done with the automatic creation of the ServiceNOW ticket for each recording.
New HTML5 viewer for recorded lectures
The CERN Recording Service provides recordings from 19 CERN rooms in a web lecture format (with synchronized speaker and slides). To give users a better control on the player side, we developed a new interactive HTML5 viewer with a built-in THEOPlayer.
Central Encoding System
With 19 CERN rooms capable of webcast and recording, about 300 live webcasts and 1200 lectures recorded every year, we needed a tool for our operators to easily start webcasts and recordings. We developed a Central Encoding Interface, from which our operators see all the planned events for a given day and with one click can start webcasting and/or recording. With this new interface we manage to almost eliminate issues where operators forget to start the webcast and with an automatic stop, we now support webcasts and recordings which finish out of standard working hours without additional manpower expenses.
Service statistics for last years
Recorded lectures
974
1185
1099
Live webcasts
260
268
271
http://cern.ch/IT ©CERN CC-BY-SA 4.0

Web Application Detection (WAD)

for asset inventory and vulnerability management

Sebastian Łopieński / CERN Computer Security Team
e-mail: Sebastian.Lopienski@cern.ch

Another usual day at CERN...

Aargh, WordPress is vulnerable again!

Fortunately, they've already published a security patch

OK! Let's tell WordPress admins at CERN to patch now!

But how do I know which CERN websites use WordPress?

Can I remotely detect technologies behind a website?

Wait! My browser plugin, *Wappalyzer*, can do it!

Let's see how *Wappalyzer* works

https://github.com/AliasIO/Wappalyzer

***Wappalyzer* detects web technologies by looking at:**

HTTP Response headers → `"Ubuntu": { "headers": { "X-Powered-By": "Ubuntu" [...]`
`"PHP": { "headers": { "Set-Cookie": "PHPSESSID" [...]`
`"Apache": { "headers": { "Server": "Apache(?:$|/([\d.]+))[^/~]*)\;version:\1" [...]`

HTML meta tags → `"Drupal": { "meta": { "Generator": "Drupal(?:\s([\d.]+))?\;version:\1" [...]`

Full HTML page content → `"JIRA": { "html": "Powered by Atlassian JIRA(?:v ((\d+\.\d+)?)\;version:\1" [...]`

Global JavaScript variables → `"jQuery": { "env": "^jQuery$" [...]`

`src` attribute of HTML `script` tags → `"AngularJS": { "script": "angular(?:\-|\.)([\d.]*\d)[^/]*\.js\;version:\1" [...]`

5 ways for detecting web technologies

~900 different technologies detected

Thousands of detection rules

Regular expressions help detecting software version

Nice plugin. Lots of detection rules! But I need a command-line tool

... after many hours of coding (and testing!)

https://github.com/CERN-CERT/WAD

```
$ pip install wad

$ wad -u http://chep2016.org
{ "http://chep2016.org": [
  { "type": "operating-systems",     "app": "Ubuntu", "ver": null   },
  { "type": "web-servers",          "app": "Apache", "ver": "2.4.7" },
  { "type": "programming-languages", "app": "PHP",    "ver": "5.5.9" },
  { "type": "cms",                  "app": "Drupal", "ver": "7"     },
  { "type": "javascript-frameworks", "app": "jQuery", "ver": "1.10"  }
] }
```

Let's add rules for CERN software

```
"Indico": {
  "cats": [ 1 ],
  "headers": {
    "Set-Cookie": "MAKACSESSION"
  },
  "html": "Powered by\s+(?:CERN )?
    <a href=\"http://(?:cdsware\.
    cern\.ch/indico/|indico-software
    \.org|cern\.ch/indico)\">(?:CDS )?
    Indico ( [\d\.]+)?\;version:\1",
  "icon": "Indico.png",
  "website": "indico-software.org"
}
```

Hurray! Now I can scan *all* CERN websites and webservers!

WAD results for CERN

Web sites scanned:	14330
Web servers scanned:	5850
Total technologies detected:	28045
Unique technologies detected:	140
[...]	
WordPress instances:	*10*

Stick Man Illustrations by Cath Noble (CERN). WAD contributors: Piotr Lizończyk, Vincent Brillault, Farzaneh Moghaddam, Antonio Perez Perez, Dame Jovanoski. Special thanks to Elbert Alias (Wappalyzer).

CERN Document Server

Towards the Next Generation CERN Institutional Repo

Retrieval

significantly improves the information retrieval
lasticsearch as the underlying search engine and
data model, the new CDS provides up to 3 times
p to 50 times faster indexing. Another determining
w design that offers a clearer, more intuitive way
d providing of data.

The CDS submission system has
our extensive experience working
new version, based on a flexible
tailored content for different user
submission and publication wor
and significant speed improveme

s document
specialized
ware stack,
repository's

t producer
oth visually
l control on
n, curation,
he data.

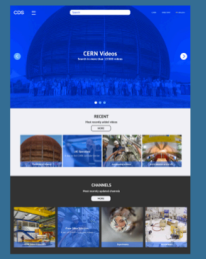
Research
All research documents produced at CERN are stored on this repository. Metadata is validated according to strict standards in order to facilitate reporting and service interoperability.

Book
Repositor
and repor
circulation
catalog is

CDS

Photos
Repository for images produced by CERN and all the Experiments. It comes with photo-related features like content-based image retrieval and image tagging.

...
Repositor
workflows
CERN Doc
empower



Record creation
- DOI minting
- Files storage (EOS)
- Metadata storage (PostgreSQL)

Presentation
- Video playback
- Download
- Embed & Export

http://cds.cern.ch
cds.support@cern.ch
/CERNDocumentServer

Flash is Dead. Finally.

M Domaracky, N Juszka, M Peksa, T Baron

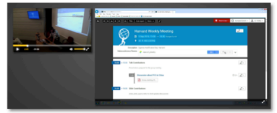
For almost 10 years, CERN has been providing live webcast of events using Adobe Flash technology. This year is finally the year when flash died at CERN! With Flash being slowly phased out on most streaming platforms, the CERN streaming service moved as well from Flash to HTTP streaming.

THEOplayer

The biggest challenge for providing pure HTML5 video streaming goes with the support of different streaming protocols across browsers and OS platforms. Thanks to THEOPlayer (www.theoplayer.com) we are able to stick with the HTTP Live Streaming (HLS) protocol from Apple and play it on all modern browsers on desktops and mobile devices. We are able to deliver the same experience as we did with Adobe Flash based players. Our users can still enjoy video of the speaker synchronised with video of the presentation, so they have the same experience as sitting in the auditoria.

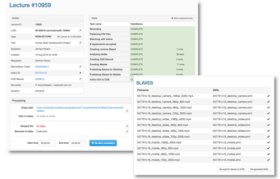
DVR – Digital Video Recorder

We will soon introduce DVR functionality for all our live webcasts. Users that arrived late on the webcast website, will have a possibility to go back to the beginning of the webcast or if they missed something they can seek back to watch it again. With the DVR feature we will also be able to publish the recording right after the webcast is finished.



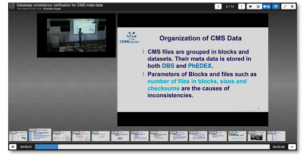
Micala

Michigan and CERN Automated Lecture Archiving (Micala) is an automated tool developed at CERN with a collaboration from the University of Michigan. We improved the workflow to faster publish the recorded lecture by integrating the Sorenson (http://www.sorensonmedia.com/squeeze/squeeze-server/) transcoding cluster to Micala. The communication with users is now done with the automatic creation of the ServiceNOW ticket for each recording.



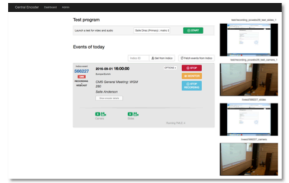
New HTML5 viewer for recorded lectures

The CERN Recording Service provides recordings from 19 CERN rooms in a web lecture format (with synchronized speaker and slides). To give users a better control on the player side, we developed a new interactive HTML5 viewer with a built-in THEOPlayer.

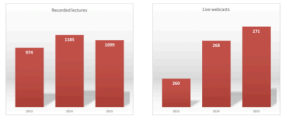


Central Encoding System

With 19 CERN rooms capable of webcast and recording, about 300 live webcasts and 1200 lectures recorded every year, we needed a tool for our operators to easily start webcasts and recordings. We developed a Central Encoding Interface, from which our operators see all the planned events for a given day and with one click can start webcasting and/or recording. With this new interface we manage to almost eliminate issues where operators forget to start the webcast and with an automatic stop, we now support webcasts and recordings which finish out of standard working hours without additional manpower expenses.



Service statistics for last years

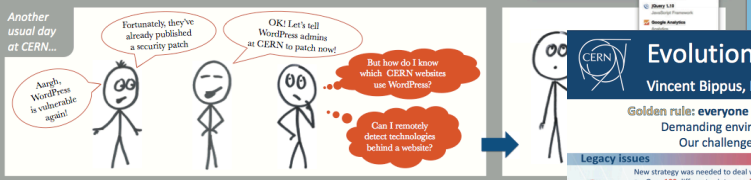


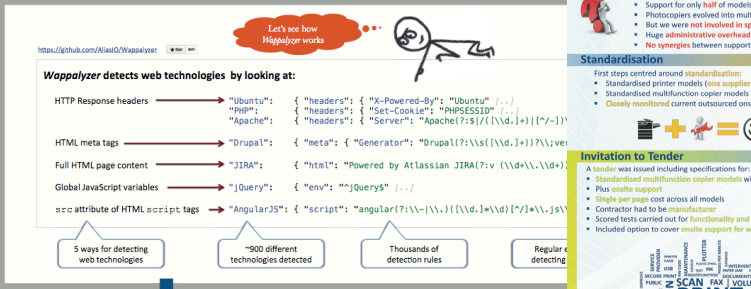
http://cern.ch/IT

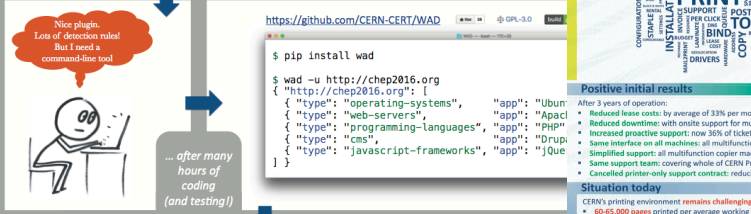
Web Application Detection (WAD)

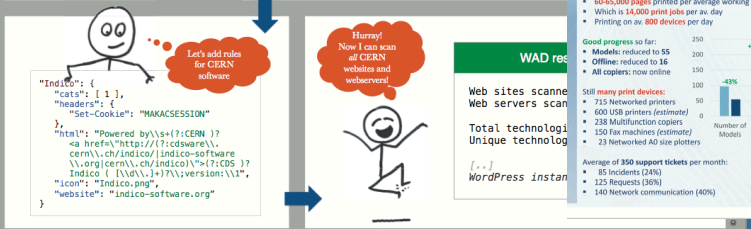
for asset inventory and vulnerability management

Sebastian Łopieński / CERN Computer Security Team
e-mail: Sebastian.Lopienski@cern.ch









Stick Man illustrations by Cath Noble (CERN). WAD contributors: Piotr Lizończyk, Vincent Brillault, Farzaneh Moghaddam, Antonio Perez Perez, Dame Jovanoski. Special thanks to Elbert Alias (Wappalyzer).

CERN Document Server

Towards the Next Generation CERN Institutional Repo

Retrieval Sub

Evolution of CERN Print Services

Vincent Bippus, Natalie Kane, Tim Smith (CERN, IT-CDA)

Golden rule: everyone at CERN can print – ALL staff, visitors, conference participants, anyone!
Demanding environment: **approx. 1000 networked print devices in 230 buildings**
Our challenge: evolve Print Services into a robust and **user-focused service**

Legacy Issues

New strategy was needed to deal with **legacy issues:**
- Over **100** different printer **models**
- Support for only **half** of models onsite
- Photocopiers evolved into multifunction copier machines
- But we were **not involved in specifications**
- Huge **administrative overhead costs**
- **No synergies** between support providers

Standardisation

First steps centred around **standardisation:**
- Standardised printer models (**one supplier** not two)
- Standardised multifunction copier models (**one supplier** not three)
- **Closely monitored** current outsourced onsite support

Invitation to Tender

A **tender** was issued including specifications for:
- **Standardised multifunction copier models** with **set prices** (just 4 models)
- Plus **onsite support**
- **Single per page** cost across all models
- Contractor had to be **manufacturer**
- Scored tests carried out for **functionality and usability**
- Included option to cover **onsite support for whole Printer Park**

Positive initial results

After 3 years of operation:
- **Reduced lease costs:** by average of 33% per model
- **Reduced downtime:** with onsite support for multifunction copier machines
- **Increased proactive support:** now 36% of tickets being dealt with proactively
- **Same interface on all machines:** all multifunction copier machines updated
- **Simplified support:** all multifunction copier machines connected to network
- **Same support team:** covering whole of CERN Printer Park
- **Cancelled printer-only support contract:** reducing support costs by 35%

Situation today

CERN's printing environment **remains challenging:**
- **60-65,000 pages** printed per average working day
- Which is **14,000 print jobs** per day
- Printing on av. **800 devices** per day

Good progress so far:
- **Models:** reduced to **55**
- **Offline:** reduced to **16**
- **All copiers:** now online

Still many print devices:
- 715 Networked printers
- 600 USB printers *(estimate)*
- 238 Multifunction copiers
- 150 Fax machines *(estimate)*
- 23 Networked A0 size plotters

Average of **350 support tickets** per month:
- 85 Incidents (24%)
- 125 Requests (36%)
- 140 Network communication (40%)

Improving user experience

We provide **search and filtering options** to organise list of printers:
- By **functionality:** ☐ Color ☐ A3 ☐ Multifunction Copier
- By **geolocation:** Printers available in your building / Printers available in buildings around your actual position

Easy printer installation

Printer Status page gives basic information including map
With easy **Web Add Printer** buttons:
- **Windows:**
  - For Windows domain members only

- **Mac:**
  - **Signed applescript application**
  - Launched through URL
  - Gets printer settings and information
  - Downloads appropriate drivers package (if not already present)
  - Configures CUPS (Common Unix Printer System) using lpadmin command

And also:
- **Self Service**
- User-friendly interface: based on JAMF Casper Suite
- Installs appropriate driver
- Multiple queue configuration in one go
- Proposes nearby printers, complete list also available

Simple print solution for visitors

Mail2Print Service: any device registered on CERN network can send email
Can print **email body and/or attachments**

Validation link emailed to user, opens website to select:
- **Printer** to use
- Print **options**, such as colour or double-sided

Future plans

- **Reduce printer numbers:** find solution for areas with multiple devices
- **Increase proactive support:** extend to printers? Identify owners & budget codes
- **Extend leasing to printers?** pay per usage but minimum charge, which would:
  - **Remove printers** from asset list & help projects with **time-limited budgets**
  - Reduce **admin overhead costs** as no need to order & hold stocks
  - Easier to **update models** when required
  - Automatically add to machines with **proactive support!**
- **Enable "my favourite printers"** on Windows?
- **New dedicated app:** enabling printer configuration on visitor devices

N. Kane, V. Bippus http://cern.ch/IT ©CERN CC-BY-SA 4.0

Flash is Dead. Finally.

M Domaracky, N Juszka, M Peksa, T Baron

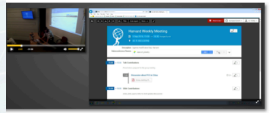
For almost 10 years, CERN has been providing live webcast of events using Adobe Flash technology. This year is finally the year when flash died at CERN! With Flash being slowly phased out on most streaming platforms, the CERN streaming service moved as well from Flash to HTTP streaming.

THEOplayer

The biggest challenge for providing pure HTML5 video streaming goes with the support of different streaming protocols across browsers and OS platforms. Thanks to THEOPLayer (www.theoplayer.com) we are able to stick with the HTTP Live Streaming (HLS) protocol from Apple and play it on all modern browsers on desktops and mobile devices. We are able to deliver the same experience as we did with Adobe Flash based players. Our users can still enjoy video of the speaker synchronised with video of the presentation, so they have the same experience as sitting in the auditoria.

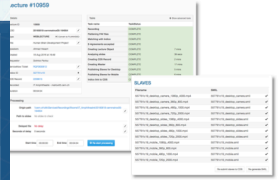
/R – Digital Video Recorder

will soon introduce DVR functionality for all our live webcasts. Users t arrived late on the webcast website, will have a possibility to go k to the beginning of the webcast or if they missed something they seek back to watch it again. With the DVR feature we will also be e to publish the recording right after the webcast is finished.



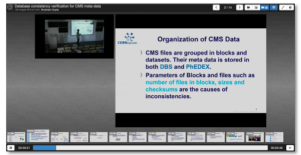
Micala

Michigan and CERN Automated Lecture Archiving (Micala) is an utomated tool developed at CERN with a collaboration from the University of Michigan. We improved the workflow to faster publish he recorded lecture by integrating the Sorenson http://www.sorensonmedia.com/squeeze/squeeze-server/) ranscoding cluster to Micala. The communication with users is now lone with the automatic creation of the ServiceNOW ticket for each ecording.



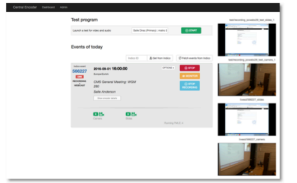
New HTML5 viewer for recorded lectures

The CERN Recording Service provides recordings from 19 CERN rooms in a web lecture format (with synchronized speaker and slides). To give users a better control on the player side, we developed a new interactive HTML5 viewer with a built-in THEOPlayer.



Central Encoding System

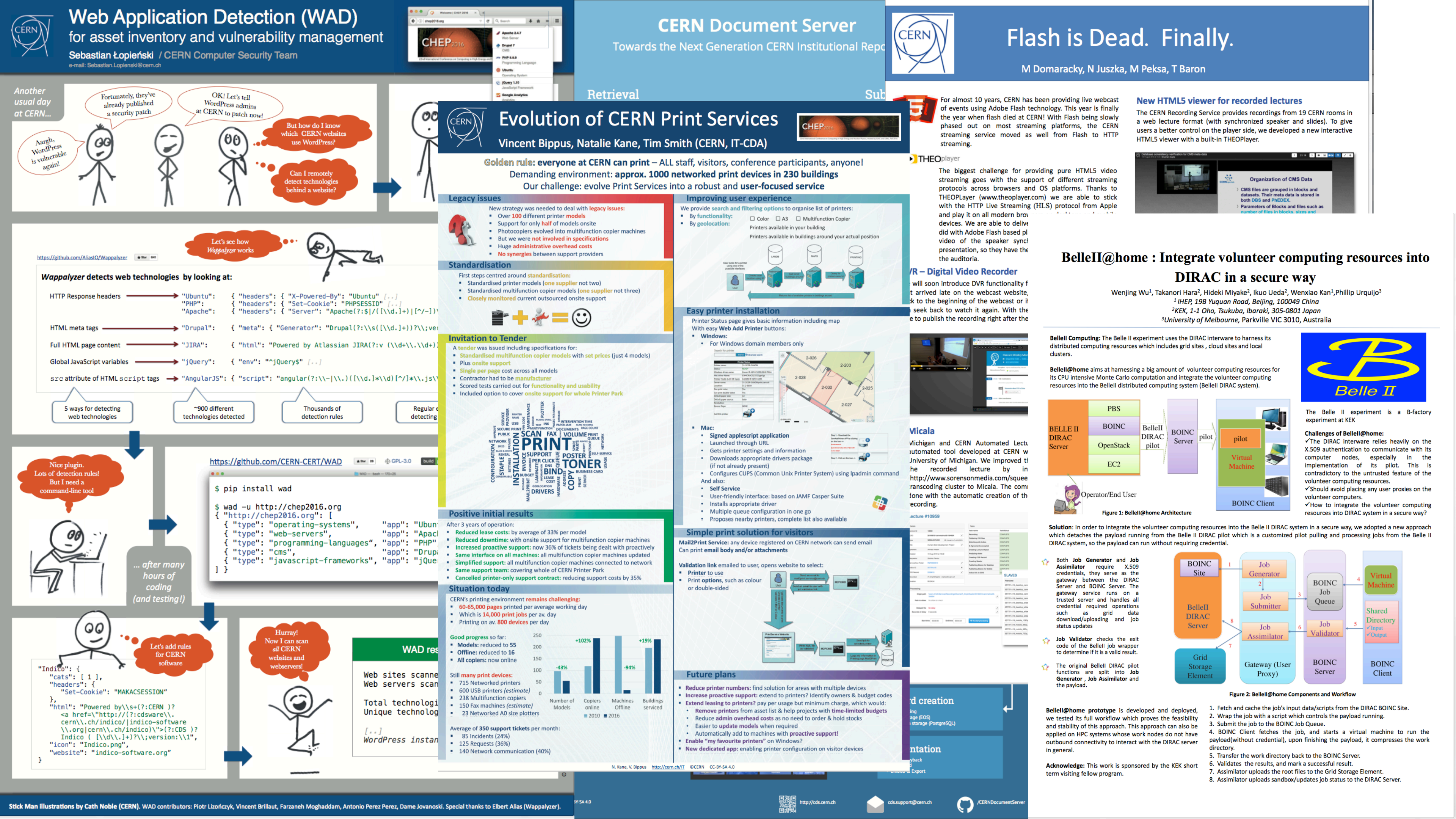
With 19 CERN rooms capable of webcast and recording, about 300 live webcasts and 1200 lectures recorded every year, we needed a tool for our operators to easily start webcasts and recordings. We developed a Central Encoding Interface, from which our operators see all the planned events for a given day and with one click can start webcasting and/or recording. With this new interface we manage to almost eliminate issues where operators forget to start the webcast and with an automatic stop, we now support webcasts and recordings which finish out of standard working hours without additional manpower expenses.



Service statistics for last years

d creation
ge (EOS)
storage (PostgreSQL)
ntation
Export

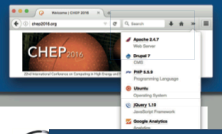
http://cds.cern.ch cds.support@cern.ch /CERNDocumentServer


Web Application Detection (WAD)
for asset inventory and vulnerability management
Sebastian Łopieński / CERN Computer Security Team
e-mail: Sebastian.Lopienski@cern.ch
Another usual day at CERN...
Fortunately, they've already published a security patch
OK! Let's tell WordPress admins at CERN to patch now!
Aargh, WordPress is vulnerable again!
But how do I know which CERN websites use WordPress?
Can I remotely detect technologies behind a website?
Let's see how Wappalyzer works
https://github.com/AliasIO/Wappalyzer
Wappalyzer detects web technologies by looking at:
HTTP Response headers
HTML meta tags
Full HTML page content
Global JavaScript variables
src attribute of HTML script tags
5 ways for detecting web technologies
~900 different technologies detected
Thousands of detection rules
Nice plugin. Lots of detection rules! But I need a command-line tool
... after many hours of coding (and testing!)
https://github.com/CERN-CERT/WAD
$ pip install wad
$ wad -u http://chep2016.org
Let's add rules for CERN software
Hurray! Now I can scan all CERN websites and webservers!
WAD res
Web sites scanne
Web servers scan
Total technologi
Unique technolog
WordPress instan
Stick Man illustrations by Cath Noble (CERN). WAD contributors: Piotr Lizończyk, Vincent Brillaut, Farzaneh Moghaddam, Antonio Perez Perez, Dame Jovanoski. Special thanks to Elbert Alias (Wappalyzer).
CERN Document Server
Towards the Next Generation CERN Institutional Repo
Retrieval
Evolution of CERN Print Services
Vincent Bippus, Natalie Kane, Tim Smith (CERN, IT-CDA)
Golden rule: everyone at CERN can print – ALL staff, visitors, conference participants, anyone!
Demanding environment: approx. 1000 networked print devices in 230 buildings
Our challenge: evolve Print Services into a robust and user-focused service
Legacy issues
Standardisation
Invitation to Tender
Positive initial results
Situation today
Improving user experience
Easy printer installation
Simple print solution for visitors
Future plans
N. Kane, V. Bippus http://cern.ch/IT ©CERN CC-BY-SA 4.0
Flash is Dead. Finally.
M Domaracky, N Juszka, M Peksa, T Baron
For almost 10 years, CERN has been providing live webcast of events using Adobe Flash technology. This year is finally the year when flash died at CERN! With Flash being slowly phased out on most streaming platforms, the CERN streaming service moved as well from Flash to HTTP streaming.
New HTML5 viewer for recorded lectures
The CERN Recording Service provides recordings from 19 CERN rooms in a web lecture format (with synchronized speaker and slides). To give users a better control on the player side, we developed a new interactive HTML5 viewer with a built-in THEOPlayer.
THEOplayer
DVR – Digital Video Recorder
Micala
BelleII@home : Integrate volunteer computing resources into DIRAC in a secure way
Wenjing Wu[1], Takanori Hara[2], Hideki Miyake[2], Ikuo Ueda[2], Wenxiao Kan[1], Phillip Urquijo[3]
[1] IHEP, 19B Yuquan Road, Beijing, 100049 China
[2] KEK, 1-1 Oho, Tsukuba, Ibaraki, 305-0801 Japan
[3] University of Melbourne, Parkville VIC 3010, Australia
BelleII Computing: The Belle II experiment uses the DIRAC interware to harness its distributed computing resources which includes grid sites, cloud sites and local clusters.
BelleII@home aims at harnessing a big amount of volunteer computing resources for its CPU intensive Monte Carlo computation and integrate the volunteer computing resources into the BelleII distributed computing system (BelleII DIRAC system).
Belle II
Figure 1: BelleII@home Architecture
Figure 2: BelleII@home Components and Workflow
Acknowledge: This work is sponsored by the KEK short term visiting fellow program.

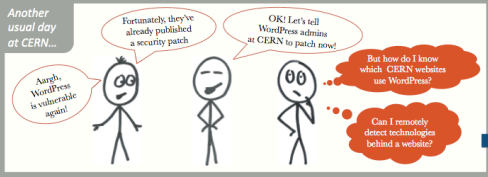
Web Application Detection (WAD)

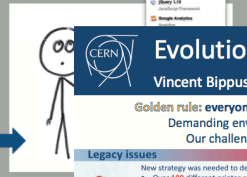
for asset inventory and vulnerability management

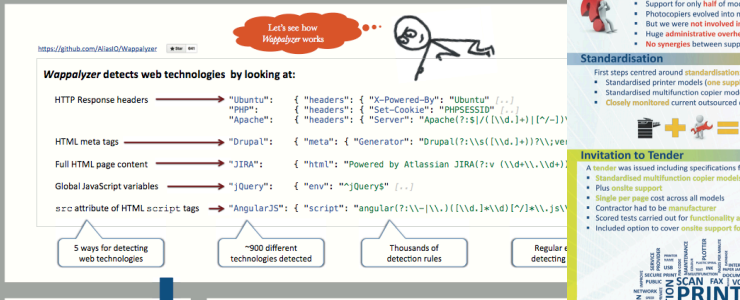
Sebastian Łopieński / CERN Computer Security Team
e-mail: Sebastian.Lopienski@cern.ch

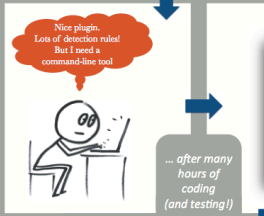


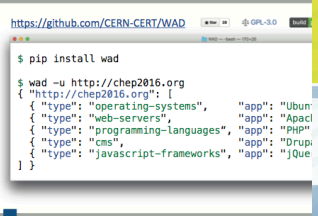
Another usual day at CERN...

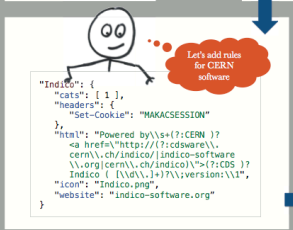


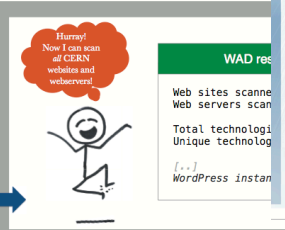












Stick Man illustrations by Cath Noble (CERN). WAD contributors: Piotr Lizończyk, Vincent Brillault, Farzaneh Moghaddam, Antonio Perez Perez, Dame Jovanoski. Special thanks to Elbert Alias (Wappalyzer).

CERN Document Server

Towards the Next Generation CERN Institutional Repo

Retrieval Sub

Evolution of CERN Print Services

Vincent Bippus, Natalie Kane, Tim Smith (CERN, IT-CDA)

Golden rule: everyone at CERN can print – ALL staff, visitors, conference participants, anyone!
Demanding environment: **approx. 1000 networked print devices in 230 buildings**
Our challenge: evolve Print Services into a robust and **user-focused service**

Legacy Issues

New strategy was needed to deal with **legacy issues:**
- Over **100** different printer **models**
- Support for only **half** of models onsite
- Photocopiers evolved into multifunction copier machines
- But we were **not involved in specifications**
- Huge **administrative overhead costs**
- **No synergies** between support providers

Standardisation

First steps centred around **standardisation:**
- Standardised printer models (**one supplier** not two)
- Standardised multifunction copier models (**one supplier** not three)
- **Closely monitored** current outsourced onsite support

Invitation to Tender

A **tender** was issued including specifications for:
- **Standardised multifunction copier models** with **set prices** (just 4 models)
- Plus **onsite support**
- **Single per page cost** across all models
- Contractor had to be **manufacturer**
- Scored tests carried out for **functionality and usability**
- Included option to cover **onsite support for whole Printer Park**

Positive initial results

After 3 years of operation:
- **Reduced lease costs:** by average of 33% per model
- **Reduced downtime:** with onsite support for multifunction copier machines
- **Increased proactive support:** now 36% of tickets being dealt with proactively
- **Same interface on all machines:** all multifunction copier machines updated
- **Simplified support:** all multifunction copier machines connected to network
- **Same support team:** covering whole of CERN Printer Park
- **Cancelled printer-only support contract:** reducing support costs by 35%

Situation today

CERN's printing environment **remains challenging:**
- **60-65,000 pages** printed per average working day
- **Which is 14,000 print jobs per** av. day
- Printing on av. **800 devices** per day

Good progress so far:
- **Models:** reduced to **55**
- **Offline:** reduced to **16**
- **All copiers:** now online

Still many print devices:
- 715 Networked printers
- 600 USB printers *(estimate)*
- 238 Multifunction copiers
- 150 Fax machines *(estimate)*
- 23 Networked A0 size plotters

Average of **350 support tickets** per month:
- 85 Incidents (24%)
- 125 Requests (36%)
- 140 Network communication (40%)

Improving user experience

We provide **search and filtering options** to organise list of printers:
- By **functionality:** ☐ Color ☐ A3 ☐ Multifunction Copier
- By **geolocation:** Printers available in your building / Printers available in buildings around your actual position

Easy printer installation

Printer Status page gives basic information including map
With easy **Web Add Printer** buttons:
- **Windows:**
  - For Windows domain members only

- **Mac:**
  - **Signed applescript application**
  - Launched through URL
  - Gets printer settings and information
  - Downloads appropriate drivers package (if not already present)
  - Configures CUPS (Common Unix Printer System) using lpadmin com

And also:
- **Self Service**
- User-friendly interface: based on JAMF Casper Suite
- Installs appropriate driver
- Multiple queue configuration in one go
- Proposes nearby printers, complete list also available

Simple print solution for visitors

Mail2Print Service: any device registered on CERN network can send email
Can print **email body and/or attachments**

Validation link emailed to user, opens website to select:
- **Printer** to use
- Print **options**, such as colour or double-sided

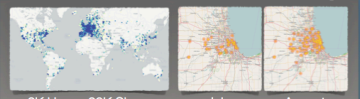
Future plans

- **Reduce printer numbers:** find solution for areas with multiple devices
- **Increase proactive support:** extend to printers? Identify owners & budget co
- **Extend leasing to printers?** pay per usage but minimum charge, which would
  - **Remove printers** from asset list & help projects with **time-limited budget**
  - Reduce **admin overhead costs** as no need to order & hold stocks
  - Easier to **update models** when required
  - Automatically add to machines with **proactive support!**
- **Enable "my favourite printers"** on Windows7
- **New dedicated app:** enabling printer configuration on visitor devices

http://cds.cern.ch

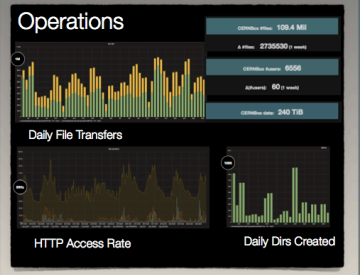
Users

ICHEP Chicago

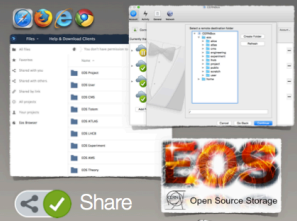


6K Users, 20K Shares — July — August

Operations

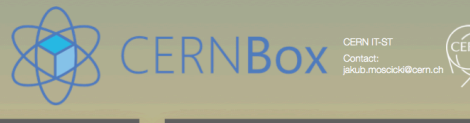


Big Data Repositories at the Fingertips



Shared File Spaces for All Platforms

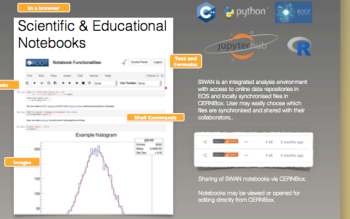
Data hub for analysis, shared file spaces and more



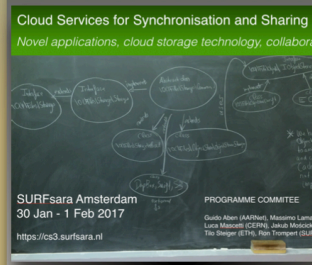
G. Adde, B. Chan, D. Duret, A. Fiorot, N. Guerrero, H.G. Labrador, J. Lopez, M. Lamanna, L. Mascetti, P. Mato, J.T. Mościcki, A.J. Peters, D. Piparo, E. Tejedor

CERN IT-ST
Contact: jakub.moscicki@cern.ch

Cloud data analysis storage backend



Education & Research Community



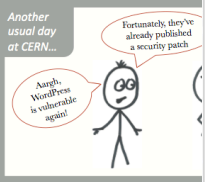
Federation
Multi-vendor API to connect on-premise clouds and sync/share solutions.

Web Application Detection (WAD)

for asset inventory and vulnerability management

Sebastian Łopieński / CERN Computer Security Team
e-mail: Sebastian.Lopienski@cern.ch

Another usual day at CERN...

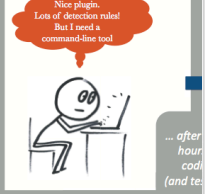


https://github.com/AliasIO/Wappalyzer

***Wappalyzer* detects web technol**

- HTTP Response headers
- HTML meta tags
- Full HTML page content
- Global JavaScript variables
- `src` attribute of HTML `script` tags

5 ways for detecting web technologies



Stick Man Illustrations by Cath Noble (CERN). WAD c

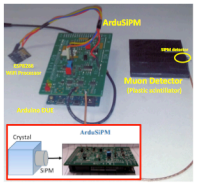
INFN Istituto Nazionale di Fisica Nucleare

An educational distributed Cosmic Ray detector network based on ArduSiPM using microcontrollers as data acquisition node NTP protocol as time distribution and IoT technology for data aggregation.

Valerio Bocci , Giacomo Chiodi, Paolo Fresch, Francesco Iacoangeli, Luigi Recchia

INFN Roma, Piazzale A.Moro, 2 - 00185 Roma
valerio.bocci@roma1.infn.it

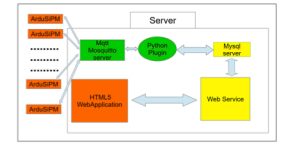
The advent of microcontrollers with enough CPU power and with analog and digital peripherals give the possibility to design a complete acquisition system in one chip. The existence of an world wide data infrastructure as internet allows to think at distributed network of detectors capable to elaborate and send data or respond to settings commands. The internet infrastructure allow us to do things unthinkable a few years ago, like to distribute the absolute time with tens of milliseconds precision to simple devices far apart from a few meters to thousands of kilometers and to create a Crowdsourcing experiment platform using simple detectors.



The ArduSiPM (1) is an easy hand-held battery operated data acquisition system based with an Arduino board, which is used to detect cosmic rays and nuclear radiation. The ArduSiPM uses an Arduino DUE (an open Software/Hardware board based on an ARM Cortex-M3 microcontroller) as processor board and a piggyback custom designed board (Shield), these are controlled by custom developed software and interface. The Shield contains different electronics features both to monitor, to set and to acquire the SiPM signal using the microcontroller board. The SiPM photon counting detector can be coupled to a cheap plastic scintillator to realize a cosmic ray detector (mainly muon particles). An ArduSiPM channel give informations about rate of events, arrival time and number of photons produced by muons, it contains all the feature from controls to data acquisition typical of High Energy Physics channel at a cost affordable for single user or school.

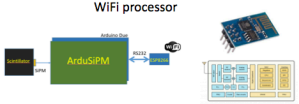
[1] The ArduSiPM a compact trasportable Software/Hardware Data Acquisition system for SiPM detector V. Bocci et al. IEEE NSS/MIC 2014 Pages: 1 - 5, DOI: 10.1109/NSSMIC.2014.7431252 arXiv:1411.7814

Using Network Time Protocol (NTP) the absolute time from the network, with a precision of tens millisecond. The network time can be used from a cloud of ArduSiPMs to detect offline coincidence events linked to Ultra High Energy Cosmic Ray



The terms of IoT (Internet of Things) define a set of data communication protocols and the capability of single embedded electronics objects to communicate using the internet .
The MQTT (Message Queue Telemetry Transport) is one of the main protocol used in IoT device for data transmission over TCP/IP, the client version can run easily in nowadays microcontrollers, the MQTT broker (the server version) can run also in credit card–sized single-board computers as well in big server.

ESP8266 MQTT, NTP and WiFi processor



The ArduSiPM sends data over rs232, the Wifi Processor elaborate the data and send them using the MQTT protocol to the server. We use as network processor the Espressif ESP8266 a low-cost Wi-Fi chip with full TCP/IP stack and a 32-bit RISC CPU running at 80 MHz. The ESP8266 can be used to send and configure MQTT packets,NTP request and configure ArduSiPM device.

ArduSiPM web configuration pages.

CERN Document Server

Towards the Next Generation CERN Institutional Repo

nt Services

h (CERN, IT-CDA)

ff, visitors, conference participants, anyone!
worked print devices in 230 buildings
a robust and **user-focused service**

Improving user experience

e provide search and filtering options to organise list of printers:
By functionality: ☐ Color ☐ A3 ☐ Multifunction Copier
By geolocation: Printers available in your building
Printers available in buildings around your actual position

Easy printer installation

Printer Status page gives basic information including map
With easy **Web Add Printer** buttons:

- **Windows:**
  - For Windows domain members only

- **Mac:**
  - **Signed applescript application**
  - Launched through URL
  - Gets printer settings and information
  - Downloads appropriate drivers package (if not already present)
  - Configures CUPS (Common Unix Printer System) using lpadmin com

And also:

- **Self Service**
  - User-friendly interface: based on JAMF Casper Suite
  - Installs appropriate driver
  - Multiple queue configuration in one go
  - Proposes nearby printers, complete list also available

Simple print solution for visitors

il2Print Service: any device registered on CERN network can send email
print **email body and/or attachments**

idation link emailed to user, opens website to select:
Printer to use
Print **options**, such as colour or double-sided

Future plans

educe printer numbers: find solution for areas with multiple devices
ncrease proactive support: extend to printers? Identify owners & budget co
Extend leasing to printers? pay per usage but minimum charge, which would

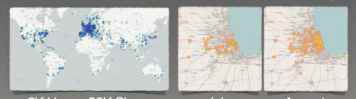
- **Remove printers** from asset list & help projects with **time-limited budget**
- Reduce **admin overhead costs** as no need to order & hold stocks
- Easier to **update models** when required
- Automatically add to machines with **proactive support!**

Enable "my favourite printers" on Windows?
New dedicated app: enabling printer configuration on visitor devices

http://cds.cern.ch

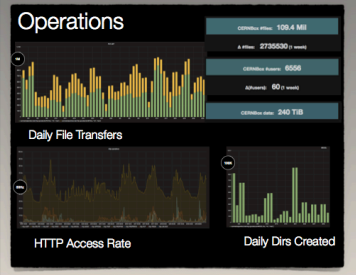
Users

ICHEP Chicago

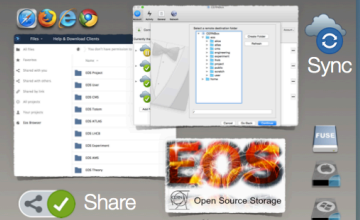


6K Users, 20K Shares — July — August

Operations



Big Data Repositories at the Fingertips



Shared File Spaces for All Platforms

Data hub for analysis, shared file spaces and more

G. Adde, B. Chan, D. Duret, A. Fiorot, N. Guerrero, H.G. Labrador, J. Lopez, M. Lamanna, L. Mascetti, P. Mato, J.T. Moscicki, A.J. Peters, D. Piparo, E. Tejedor

CERNBox

CERN IT-ST
Contact: jakub.moscicki@cern.ch

Cloud data analysis storage backend

Scientific & Educational Notebooks

SWAN is an integrated analysis environment with access to online data repositories in EOS and locally synchronised files in CERNBox. User may easily choose which files are synchronised and shared with their collaborators.

Sharing of SWAN notebooks via CERNBox.

Notebooks may be viewed or opened for editing directly from CERNBox.

Integration Architecture

Education & Research Community

Cloud Services for Synchronisation and Sharing (CS3)

Novel applications, cloud storage technology, collaborations

SURFsara Amsterdam
30 Jan - 1 Feb 2017
https://cs3.surfsara.nl

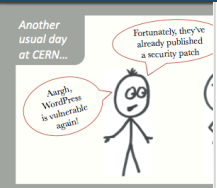
PROGRAMME COMMITEE
Guido Aben (AARNet), Massimo Lamanna (CERN), Luca Mascetti (CERN), Jakub Moscicki (CERN), Tilo Steiger (ETH), Ron Trompert (SURFsara)

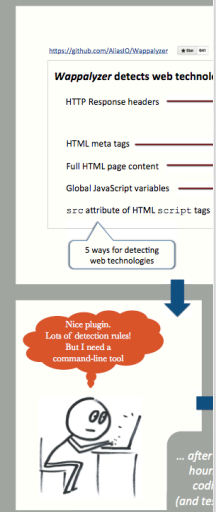
Federation
Multi-vendor API to connect on-premise clouds and sync/share solutions.

Web Application Detection (WAD)

for asset inventory and vulnerability management

Sebastian Łopieński / CERN Computer Security Team
e-mail: Sebastian.Lopienski@cern.ch





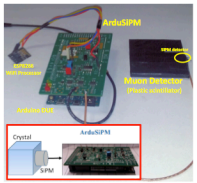
Stick Man illustrations by Cath Noble (CERN). WAD c

An educational distributed Cosmic Ray
based on ArduSiPM using microcor
acquisition node NTP protocol as time
technology for data aggre

Valerio Bocci , Giacomo Chiodi, Paolo Fresch, Francesco I

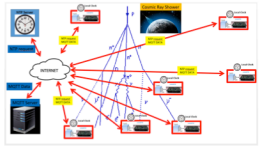
INFN Roma, Piazzale A.Moro, 2 - 00185 F
valerio.bocci@roma1.infn.it

The advent of microcontrollers with enough CPU power and with analog and digital peripherals give the poss
in one chip. The existence of an world wide data infrastructure as internet allows to think at distributed netw
data or respond to settings commands. The internet infrastructure allow us to do things unthinkable a few
with tens of milliseconds precision to simple devices far apart from a few meters to thousands of kilomete
platform using simple detectors.

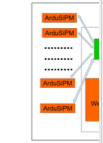


The ArduSiPM (1) is an easy hand-held battery operated data acquisition system based with an Arduino board, which is used to detect cosmic rays and nuclear radiation. The ArduSiPM uses an Arduino DUE (an open Software/Hardware board based on an ARM Cortex-M3 microcontroller) as processor board and a piggyback custom designed board (Shield), these are controlled by custom developed software and interface. The Shield contains different electronics features both to monitor, to set and to acquire the SiPM signal using the microcontroller board. The SiPM photon counting detector can be coupled to a cheap plastic scintillator to realize a cosmic ray detector (mainly muon particles). An ArduSiPM channel give informations about rate of events, arrival time and number of photons produced by muons, it contains all the feature from controls to data acquisition typical of High Energy Physics channel at a cost affordable for single user or school.

[1] The ArduSiPM a compact trasportable Software/Hardware Data Acquisition system for SiPM detector V. Bocci et al. IEEE NSS/MIC 2014 Pages: 1 - 5, DOI: 10.1109/NSSMIC.2014.7431252 arXiv:1411.7814



Using Network Time Protocol (NTP) the absolute time from the network, with a precision of tens millisecond. The network time can be used from a cloud of ArduSiPMs to detect offline coincidence events linked to Ultra High Energy Cosmic Ray



The terms of IoT (Inter
protocols and the capa
communicate using the
The MQTT (Message C
used in IoT device for C
easily in nowadays mic
run also in credit card–

ESP

The ArduSiPM sends data over Wi-Fi, the MQTT protocol to the server. We use as network processor the Espressif ESP8266 a low-cost Wi-Fi chip with full TCP/IP stack and a 32-bit RISC CPU running at 80 MHz. The ESP8266 can be used to send and configure MQTT packets,NTP request and configure ArduSiPM device.

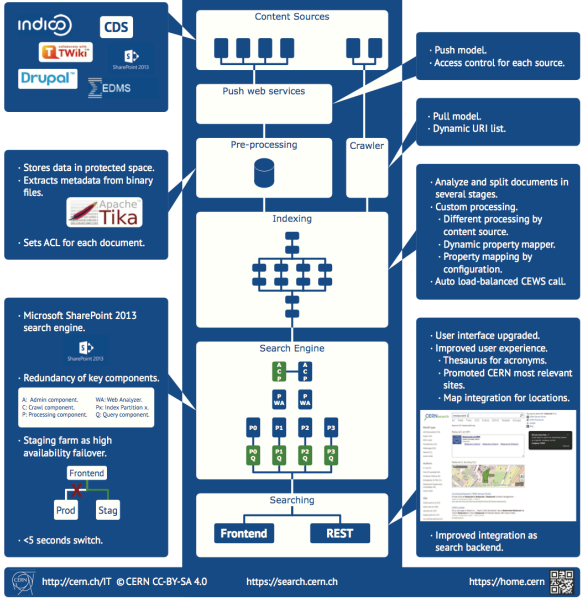
ArduSiPM web configuration pages.

Update on CERN Search: SharePoint 2013

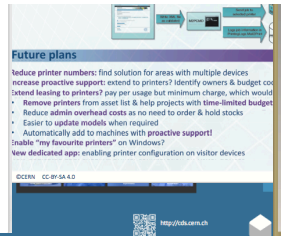
E. Alvarez, S. Fernandez, A. Lossent, I. Posada, A. Wagner [CERN]

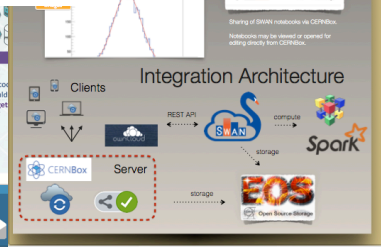
CERN's enterprise Search solution 'CERN Search' provides a central search solution for both users and CERN service providers. Public and protected documents from a wide range of documents are indexed and available for retrieval.

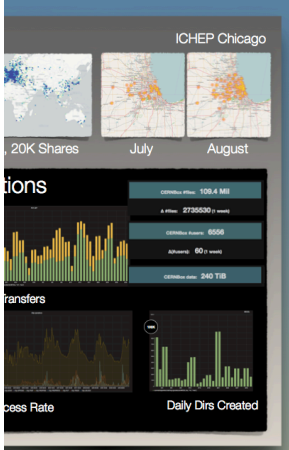
High Availability	Improved design	High Reliability
2000 daily queries	33 million documents	1100 daily unique visitors
29 servers	12000 daily page views	3 server farms

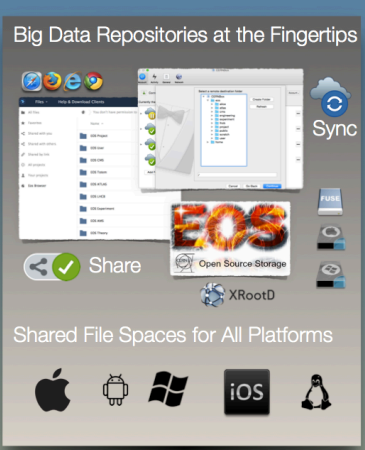


http://cern.ch/IT © CERN CC-BY-SA 4.0 https://search.cern.ch https://home.cern







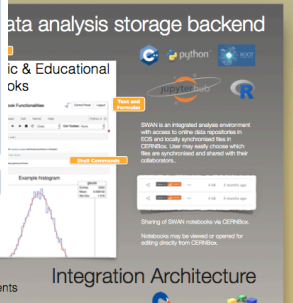


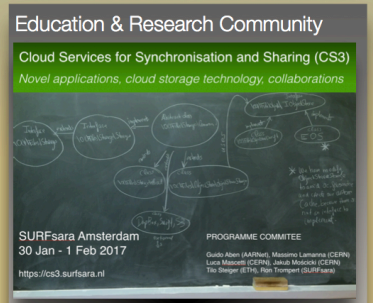
hub for analysis, shared file spaces and more

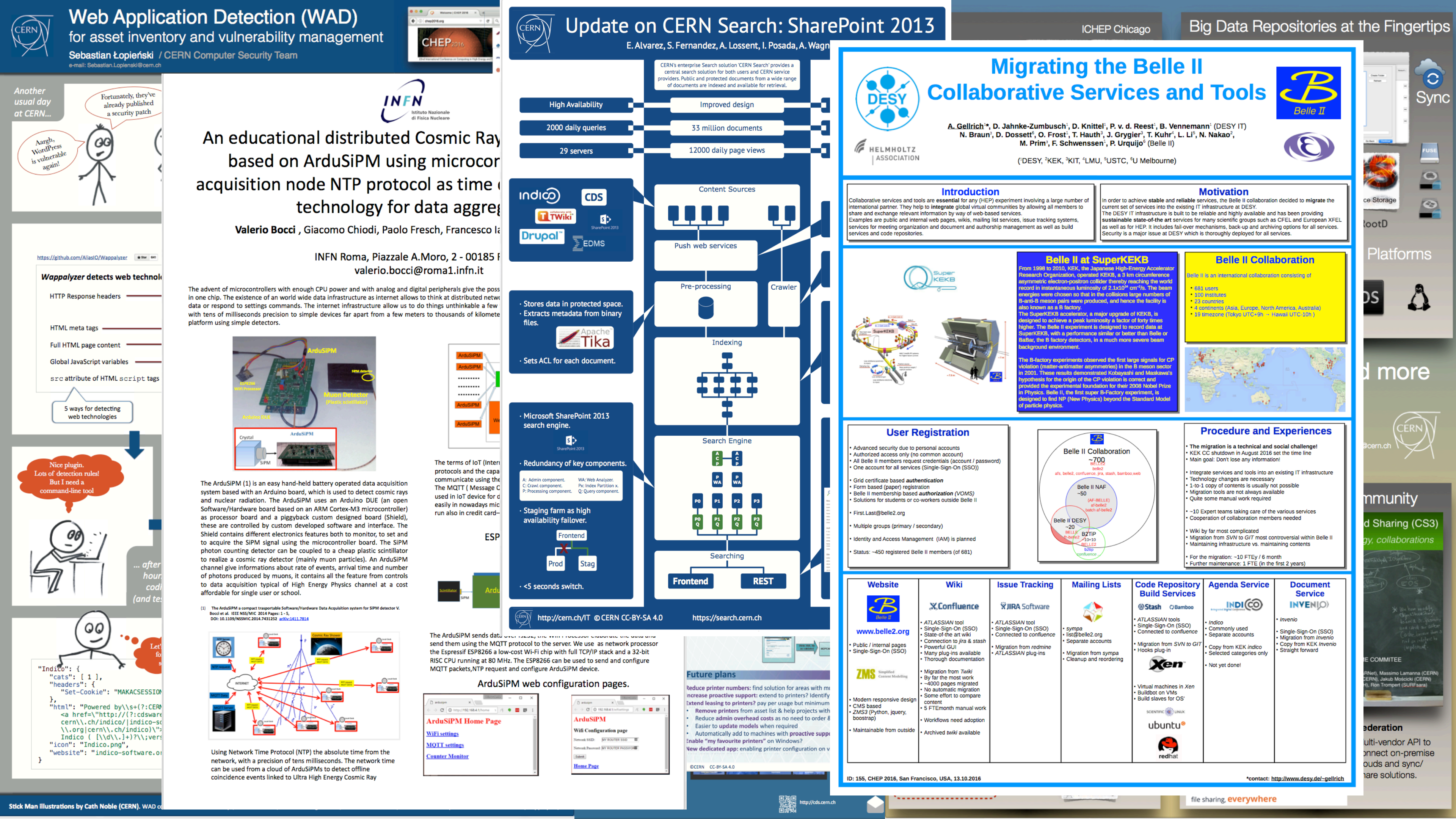
M. Lamanna
L. Mascetti
P. Mato
J.T. Mościcki
A.J. Peters
D. Piparo
E. Tejedor

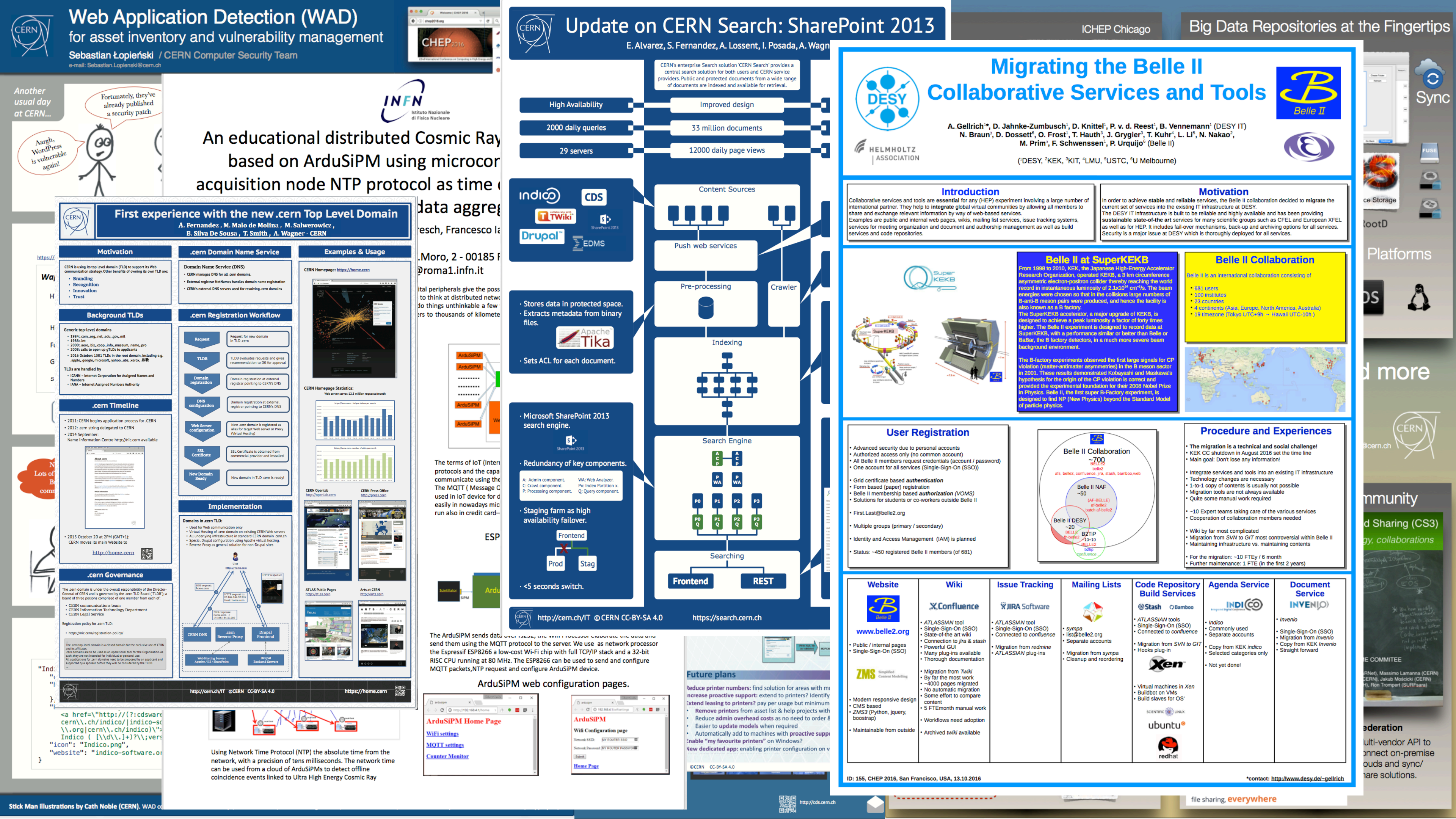
CERNBox

CERN IT-ST
Contact:
jakub.moscicki@cern.ch






Web Application Detection (WAD)
for asset inventory and vulnerability management
Sebastian Łopieński / CERN Computer Security Team
Another usual day at CERN...
Fortunately, they've already published a security patch
Aargh, WordPress is vulnerable again!
Wappalyzer detects web technol
HTTP Response headers
HTML meta tags
Full HTML page content
Global JavaScript variables
src attribute of HTML script tags
5 ways for detecting web technologies
Nice plugin. Lots of detection rules! But I need a command-line tool
Stick Man illustrations by Cath Noble (CERN).
INFN Istituto Nazionale di Fisica Nucleare
An educational distributed Cosmic Ray
based on ArduSiPM using microcon
acquisition node NTP protocol as time
technology for data aggreg
Valerio Bocci , Giacomo Chiodi, Paolo Fresch, Francesco I
INFN Roma, Piazzale A.Moro, 2 - 00185 R
valerio.bocci@roma1.infn.it
ArduSiPM web configuration pages.
Using Network Time Protocol (NTP) the absolute time from the network, with a precision of tens milliseconds. The network time can be used from a cloud of ArduSiPMs to detect offline coincidence events linked to Ultra High Energy Cosmic Ray
Update on CERN Search: SharePoint 2013
E. Alvarez, S. Fernandez, A. Lossent, I. Posada, A. Wagn
High Availability
2000 daily queries
29 servers
Improved design
33 million documents
12000 daily page views
Content Sources
Push web services
Pre-processing
Crawler
Indexing
Search Engine
Searching
Frontend
REST
Stores data in protected space.
Extracts metadata from binary files.
Sets ACL for each document.
Microsoft SharePoint 2013 search engine.
Redundancy of key components.
Staging farm as high availability failover.
<5 seconds switch.
http://cern.ch/IT © CERN CC-BY-SA 4.0 https://search.cern.ch
Future plans
ICHEP Chicago
Big Data Repositories at the Fingertips
Migrating the Belle II Collaborative Services and Tools
DESY
HELMHOLTZ ASSOCIATION
Belle II
Introduction
Motivation
Belle II at SuperKEKB
Belle II Collaboration
User Registration
Procedure and Experiences
Website
Wiki
Issue Tracking
Mailing Lists
Code Repository Build Services
Agenda Service
Document Service
www.belle2.org
ID: 155, CHEP 2016, San Francisco, USA, 13.10.2016
*contact: http://www.desy.de/~gellrich
Sync
Platforms
d more
mmunity
file sharing, everywhere



Web Application Detection (WAD)
for asset inventory and vulnerability management
Sebastian Łopieński / CERN Computer Security Team
Another usual day at CERN...
Fortunately, they've already published a security patch
Aargh, WordPress is vulnerable again!
An educational distributed Cosmic Ray
based on ArduSiPM using microcon
acquisition node NTP protocol as time
First experience with the new .cern Top Level Domain
A. Fernandez, M. Malo de Molina, M. Salwerowicz, B. Silva De Sousa, T. Smith, A. Wagner - CERN
Motivation
.cern Domain Name Service
Examples & Usage
Background TLDs
.cern Registration Workflow
.cern Timeline
Implementation
.cern Governance
http://cern.ch/IT ©CERN CC-BY-SA 4.0
https://home.cern
Update on CERN Search: SharePoint 2013
E. Alvarez, S. Fernandez, A. Lossent, I. Posada, A. Wagn
High Availability
2000 daily queries
29 servers
Improved design
33 million documents
12000 daily page views
Content Sources
Push web services
Pre-processing
Crawler
Indexing
Search Engine
Searching
Frontend
REST
Stores data in protected space.
Extracts metadata from binary files.
Sets ACL for each document.
Microsoft SharePoint 2013 search engine.
Redundancy of key components.
Staging farm as high availability failover.
<5 seconds switch.
http://cern.ch/IT ©CERN CC-BY-SA 4.0
https://search.cern.ch
ArduSiPM web configuration pages.
Future plans
ICHEP Chicago
Big Data Repositories at the Fingertips
Sync
Platforms
more
mmunity
d Sharing (CS3)
ederation
file sharing, everywhere
Migrating the Belle II Collaborative Services and Tools
A. Gellrich, D. Jahnke-Zumbusch, D. Knittel, P. v. d. Reest, B. Vennemann (DESY IT)
N. Braun, D. Dossett, O. Frost, T. Hauth, J. Grygier, T. Kuhr, L. Li, N. Nakao,
M. Prim, F. Schwennsen, P. Urquijo (Belle II)
(DESY, KEK, KIT, LMU, USTC, U Melbourne)
Introduction
Motivation
Belle II at SuperKEKB
Belle II Collaboration
User Registration
Procedure and Experiences
Website
Wiki
Issue Tracking
Mailing Lists
Code Repository Build Services
Agenda Service
Document Service
ID: 155, CHEP 2016, San Francisco, USA, 13.10.2016
Stick Man illustrations by Cath Noble (CERN).

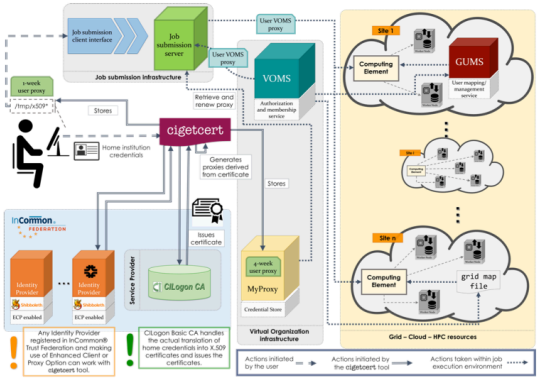
Grid Access with Federated Identities

Dave Dykstra, Mine Altunay, Jeny Teheran, Tanya Levshina,
Neha Sharma, Dennis Box, Kenneth Herner, Amanda Gao.
Scientific Computing Division, Fermilab, Batavia, IL

Key takeaways

- **cigetcert** is an open source, general purpose command line tool developed for this project. It enables federation-based authentication technologies for grid access. Any VO could use this tool to create certificates for their users with logins at institutions in the InCommon® Federation or any other federation supporting a SAML ECP X.509 certificate Service Provider.
- Our system isolates users from the certificate management process. A user only authenticates with his/her home organization in the same manner she/he always does (e.g. username/password or Kerberos). **cigetcert** facilitates this authentication and generates certificates transparent to the user with the help of CILogon Basic CA. The infrastructure manages certificates for the user.
- **cigetcert** stores a proxy certificate on the local disk and stores a longer-lived proxy in **MyProxy** for use by a job submission system to renew authentication so long-lived jobs can access storage.

Architecture



Motivation

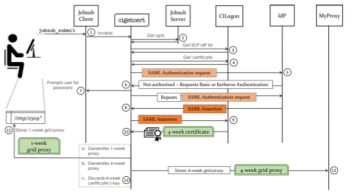
- Managing certificates by hand is often an impediment for grid users that are not tech-savvy.
- Fermilab had a grid job management system which managed certificates for users but was completely dependent on Kerberos tickets issued by a Fermilab Kerberos Certifying Authority (KCA).
  - This forced collaborating scientists to create local accounts at Fermilab or learn to configure Fermilab Kerberos authentication on their own machines, both of which proved difficult.
  - The KCA was expensive to maintain and losing software support.

Our goal

- To create a system where non-Fermilab collaborators can access our tools with low barriers using their own home institution credentials.

Design

- Make use of existing InCommon CILogon Basic CA and existing InCommon federated Identity Providers (IdPs).
  - Enable Enhanced Client or Proxy (ECP) option on the IdPs.
- Write new **cigetcert** command line tool to get certificates on users' behalf:
  - Generic open source tool, not Fermilab-specific.
  - Authenticate users with either Kerberos ticket or username/password.
  - Get 4-week certificate from CILogon Basic CA.
  - Generate and store 1-week proxy on local disk, and 4 week proxy in MyProxy, both unencrypted.
    - Complies with International Grid Trust Foundation (IGTF) rules.
  - Re-use proxy certificates that have enough time remaining, to reduce server load.



- Modify Fermilab job submission system to use **cigetcert** tool:
  - If a Fermilab user already has a Kerberos ticket, use it for authentication. If not, tell user to run **cigetcert** to enter his or her password.
- Modify job submission server to renew user proxies from MyProxy.

Status and Future plans

- Phase 1, using only the Fermilab IdP and automatically registering user DNs in VOMS, is complete and in production.
- Phase 2 is scheduled for next year. The goal is to configure **cigetcert** to use other InCommon IdPs and create a new web service with federated single-sign-on to register user DNs automatically in VOMS.
- **cigetcert** is available at https://github.com/fermitools/cigetcert

Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.
CHEP abstract ID: 183.

Migrating the Belle II Collaborative Services and Tools

A. Gellrich*, D. Jahnke-Zumbusch*, D. Knittel*, P. v. d. Reest*, B. Vennemann* (DESY IT)
N. Braun*, D. Dossett§, O. Frost*, T. Hauth*, J. Grygier*, T. Kuhr*, L. Li*, N. Nakao*,
M. Prim*, F. Schwenssen*, P. Urquijo§ (Belle II)
(*DESY, ²KEK, ³KIT, ⁴LMU, ⁵USTC, ⁶U Melbourne)

Introduction

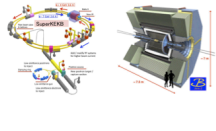
Collaborative services and tools are **essential** for any (HEP) experiment involving a large number of international partner. They help to **integrate** global virtual communities by allowing all members to share and exchange relevant information by way of web-based services.
Examples are public and internal web pages, wikis, mailing list services, issue tracking systems, services for meeting organization and document and authorship management as well as build services and code repositories.

Motivation

In order to achieve **stable** and **reliable** services, the Belle II collaboration decided to **migrate** the current set of services into the existing IT infrastructure at DESY.
The DESY IT infrastructure is built to be reliable and highly available and has been providing **sustainable state-of-the art** services for many scientific groups such as CFEL and European XFEL as well as for HEP. It includes fail-over mechanisms, back-up and archiving options for all services.
Security is a major issue at DESY which is thoroughly deployed for all services.

Belle II at SuperKEKB

From 1998 to 2010, KEK, the Japanese High-Energy Accelerator Research Organization, operated KEKB, a 3 km circumference asymmetric electron-positron collider thereby reaching the world record in instantaneous luminosity of 2.1×10^{34} cm^{-2}s^{-1}. The beam energies were chosen so that in the collisions large numbers of B-anti-B meson pairs were produced, and hence the facility is also known as a B factory.
The SuperKEKB accelerator, a major upgrade of KEKB, is designed to achieve a peak luminosity a factor of forty times higher. The Belle II experiment is designed to record data at SuperKEKB, with a performance similar or better than Belle or BaBar, the B factory detectors, in a much more severe beam background environment.
The B-factory experiments observed the first large signals for CP violation (matter-antimatter asymmetries) in the B meson sector in 2001. These results demonstrated Kobayashi and Maskawa's hypothesis for the origin of the CP violation is correct and provided the experimental foundation for their 2008 Nobel Prize in Physics. Belle II, the first super B-Factory experiment, is designed to find NP (New Physics) beyond the Standard Model of particle physics.



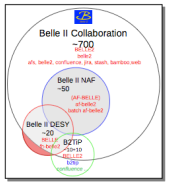
Belle II Collaboration

Belle II is an international collaboration consisting of

- 681 users
- 100 institutes
- 23 countries
- 4 continents (Asia, Europe, North America, Australia)
- 18 timezone (Tokyo UTC+9h – Hawaii UTC-10h)

User Registration

- Advanced security due to personal accounts
- Authorized access only (no common account)
- All Belle II members request credentials (account / password)
- One account for all services (Single-Sign-On (SSO))
- Grid certificate based **authentication**
- Form based (paper) registration
- Belle II membership based **authorization** (*VOMS*)
- Solutions for students or co-workers outside Belle II
- First.Last@belle2.org
- Multiple groups (primary / secondary)
- Identity and Access Management (IAM) is planned
- Status: ~450 registered Belle II members (of 681)



Procedure and Experiences

- **The migration is a technical and social challenge!**
- KEK CC shutdown in August 2016 set the time line
- Main goal: Don't lose any information!
- Integrate services and tools into an existing IT infrastructure
- Technology changes are necessary
- 1-to-1 copy of contents is usually not possible
- Migration tools are not always available
- Quite some manual work required
- ~10 Expert teams taking care of the various services
- Cooperation of collaboration members needed
- Wiki by far most complicated
- Migration from *SVN* to *GIT* most controversial within Belle II
- Maintaining infrastructure vs. maintaining contents
- For the migration: ~10 FTEy / 6 month
- Further maintenance: 1 FTE (in the first 2 years)

Website

www.belle2.org

- Public / internal pages
- Single-Sign-On (SSO)
- Modern responsive design
- CMS based
- *ZMS3* (Python, jquery, bootstrap)
- Maintainable from outside

Wiki

Confluence

- *ATLASSIAN* tool
- Single-Sign-On (SSO)
- State-of-the art wiki
- Connection to *jira & stash*
- Powerful GUI
- Many plug-ins available
- Thorough documentation
- Migration from *Twiki*
- By far the most work
- ~4000 pages migrated
- No automatic migration
- Some effort to compare content
- 5 FTE*month manual work
- Workflows need adoption
- Archived *twiki* available

Issue Tracking

JIRA Software

- *ATLASSIAN* tool
- Single-Sign-On (SSO)
- Connected to confluence
- Migration from redmine
- *ATLASSIAN* plug-ins

Mailing Lists

- *sympa*
- list@belle2.org
- Separate accounts
- Migration from sympa
- Cleanup and reordering

Code Repository Build Services

Stash Bamboo

- *ATLASSIAN* tools
- Single-Sign-On (SSO)
- Connected to confluence
- Migration from *SVN* to *GIT*
- Hooks plug-in

Xen

- Virtual machines in *Xen*
- Buildbot on VMs
- Build slaves for OS'

ubuntu redhat

Agenda Service

INDICO

- *Indico*
- Commonly used
- Separate accounts
- Copy from KEK *indico*
- Selected categories only
- Not yet done!

Document Service

INVENIO

- *invenio*
- Single-Sign-On (SSO)
- Migration from *invenio*
- Copy from KEK *invenio*
- Straight forward

ID: 155, CHEP 2016, San Francisco, USA, 13.10.2016

*contact: http://www.desy.de/~gellrich

Grid Access with Federated Identities

Dave Dykstra, Mine Altunay, Jeny Teheran, Tanya Levshina,
Neha Sharma, Dennis Box, Kenneth Herner, Amanda Gao.
Scientific Computing Division, Fermilab, Batavia, IL

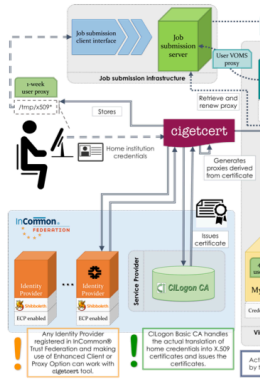
Fermilab

CILogon

Key takeaways

- **cigetcert** is an open source, general purpose command line tool developed for this project. It enables federation-based authentication technologies for grid access. Any VO could use this tool to create certificates for their users with logins at institutions in the InCommon® Federation or any other federation supporting a SAML ECP X.509 certificate Service Provider.
- Our system isolates users from the certificate management process. A user only authenticates with his/her home organization in the same manner she/he always does (e.g. username/password or Kerberos). **cigetcert** facilitates this authentication and generates certificates transparent to the user with the help of CILogon Basic CA. The infrastructure manages certificates for the user.
- **cigetcert** stores a proxy certificate on the local disk and stores a longer-lived proxy in **MyProxy** for use by a job submission system to renew authentication so long-lived jobs can access storage.

Architecture



Motivation

- Managing certificates by hand is often an impediment for grid users that are not tech-savvy.
- Fermilab had a grid job management system which managed certificates for users but was completely dependent on Kerberos tickets issued by a Fermilab Kerberos Certifying Authority (KCA).
  - This forced collaborating scientists to create local accounts at Fermilab or learn to configure Fermilab Kerberos authentication on their own machines, both of which proved difficult.
  - The KCA was expensive to maintain and losing software support.

Our goal

- To create a system where non-Fermilab collaborators can access our tools with low barriers using their own home institution credentials.

Design

- Make use o
  - InCommon
  - Enable
- Write new c
  users' beha
  - Generic c
  - Authentic
  - Get 4-we
  - Generate
    MyProxy
  - Compl
- Re-use p
  server lo

- Modify Ferm
  - If a Ferr
    authenti
    her pass
- Modify job submission server to renew user proxies from MyProxy.

Status and Future plans

- Phase 1, using only the Fermilab IdP and automatically registering user DNs in VOMS, is complete and in production.
- Phase 2 is scheduled for next year. The goal is to configure **cigetcert** to use other InCommon IdPs and create a new web service with federated single-sign-on to register user DNs automatically in VOMS.
- **cigetcert** is available at https://github.com/fermitools/cigetcert

Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.
CHEP abstract ID: 183.

CHEP 2016

http://atlascomputingagora.web.cern.ch/

The ATLAS Computing Agora (ACA): a resource web site for citizen science projects

The ATLAS collaboration has recently setup a number of citizen science projects which have a strong IT component: event simulation through *volunteer computing*, algorithm improvements via Machine Learning challenges, event scanning on citizen science platforms, use of open data, etc. Most of the interactions with volunteers are handled through message boards, but specific outreach material has also been developed, giving enhanced visibility to the ATLAS software and computing techniques, challenges and community. The need for a dedicated platform was thus identified and discussed within the ATLAS Computing group. In the proposed "ATLAS Computing Agora" web site, static descriptions of the software and computing domains are combined with links to outreach activities and news pieces, in order to reach and engage communities beyond the usual circles of followers. Analysis of the number of visitors and evaluation surveys will help to better identify the origin, needs, wishes and motivation of our visitors.

"While running ATLAS simulation, please have a look at… and discover…"

aca ATLAS

Data Flow — Simulation, Distributed Production, Reconstruction, Calibration, Analysis, New ideas and models

Projects — Citizen science, Machine Learning, Open data, Visualising the Invisible

Updates — Outreach, Lectures and Talks, CHEP Briefings

Data Flow

Most of the documents prepared for the media and the general public describe detectors and physics results without detailing the data flow. This weakness was partially addressed in 2016 by a "CERN backgrounder" (short document posted on the web and social media) describing the 12 steps from ideas to discoveries:

1. Ideas and models
2. Simulation
3. Construction
4. Collisions
5. Detection
6. Event selection
7. Reconstruction
8. Calibration
9. Data analysis
10. Comparison with simulation
11. Bumps and statistics
12. Deeper exploration

Each of these items involves software and computing activities. The 12 steps were adapted in ACA to highlight specific domains. Those describing computing and offline software are indicated in blue.

The theme "data flow" allows for the smooth integration of outreach material developed by other actors, such as the IT department, as well as the integration of plots, maps & numbers extracted from the ATLAS computing dashboards and monitoring tools used by shifters and experts.

Data and data flow have proven to be a strong source of inspiration for artists, educators and the public. A colourful and up-to-date page provides an overview

ACA visitors can either click on word bubbles or use keywords to access a set of project specific web sites.

Projects

ATLAS outreach involves communication platforms and content, educational tools, local events and visits addressing the general public, policy makers, students and teachers.

Projects are complementary: proposed by a small group of collaboration members, funded by their institutes, they reach specific audiences and are often set up in partnership with educators, artists or experts from other disciplines. Several projects with large outreach potential have a strong software and computing component:

❑ ***Citizen science projects***: volunteers celebrate the Higgs anniversary by looking for secondary vertices on events

❑ ***Sonification***: data transformed into sounds are a source of inspiration for musicians.

❑ ***Virtual reality:*** the ATLASrift project takes Oculus fans down to the ATLAS cavern.

❑ ***Open data:*** the portal includes interactive tools and Virtual Machines

Updates

Agora*: central spot of ancient Greek cities. By extension a "gathering place" or "assembly".*

Interactions between the projects developers and participants occur primarily through message boards, but resource links are a source of inspiration and motivation. The ATLAS outreach group has developed over the years a strong reputation for communicating UPDATES to the general public:

❑ ***News****: describe specific events chosen by the communication team. This model was used on a few occasions for software and computing tasks*

❑ ***Blogs****: written and signed by collaboration members, they have a more personal style and describe every day life and work.*

❑ ***Briefings****: describe and summarise of the content of a publication to the public. This model, originally developed for physics publications, can be used for papers such as CHEP proceedings.*

❑ ***Links to talks*** and seminars given as part of CERN schools and students programs. They provide a deeper understanding.

❑ ***Virtual visits:*** are organised on request. They allow an interactive connection from the ATLAS control room. ATLAS activities are introduced, followed by a Q&A session.

❑ ***Evaluation:*** goes through surveys, which help to better understand the audience profile, needs and inputs. Some outreach projects have successfully pioneered this approach.

Claire Adam (LAL, Université Paris Sud, France) and Riccardo Bianchi (University of Pittsburg, US)
On behalf of the ATLAS collaboration.

ATLAS EXPERIMENT

Migrating the Belle II Collaborative Services and Tools

A. Gellrich[1], D. Jahnke-Zumbusch[1], D. Knittel[1], P. v. d. Reest[1], B. Vennemann[1] (DESY IT)
T. Braun[2], D. Dossett[3], O. Frost[1], T. Hauth[2], J. Grygier[2], T. Kuhr[2], L. Li[2], N. Nakao[2],
M. Prim[2], F. Schwenssen[1], P. Urquijo[3] (Belle II)

([1]DESY, [2]KEK, [3]KIT, [4]LMU, [5]USTC, [6]U Melbourne)

Motivation

In order to achieve **stable** and **reliable** services, the Belle II collaboration decided to **migrate** the current set of services into the existing IT infrastructure at DESY.
The DESY IT infrastructure is built to be reliable and highly available and has been providing **sustainable state-of-the art** services for many scientific groups such as CFEL and European XFEL as well as for HEP. It includes fail-over mechanisms, back-up and archiving options for all services. Security is a major issue at DESY which is thoroughly deployed for all services.

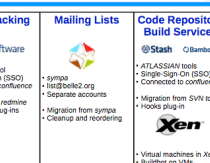
Belle II at SuperKEKB

From 1998 to 2010, KEK, the Japanese High-Energy Accelerator Research Organization, operated KEKB, a 3 km circumference asymmetric electron-positron collider thereby reaching the world record in instantaneous luminosity of 2.1×10^{34} cm^{-2}s^{-1}. The beam energies were chosen so that in the collisions large numbers of B-anti-B meson pairs were produced, and hence the facility is also known as a B factory.
The SuperKEKB accelerator, a major upgrade of KEKB, is designed to achieve a peak luminosity a factor of forty times higher. The Belle II experiment is designed to record data at SuperKEKB, with a performance similar or better than Belle or Babar, the B factory detectors, in a much more severe beam background environment.
The B-factory experiments observed the first large signals for CP violation (matter-antimatter asymmetries) in the B meson sector in 2001. These results demonstrated Kobayashi and Maskawa's hypothesis for the origin of the CP violation is correct and provided the experimental foundation for their 2008 Nobel Prize in Physics. Belle II, the first super B-Factory experiment, is designed to find NP (New Physics) beyond the Standard Model of particle physics.

Belle II Collaboration

Belle II is an international collaboration consisting of

- 681 users
- 100 institutes
- 23 countries
- 4 continents (Asia, Europe, North America, Australia)
- 18 timezone (Tokyo UTC+9h – Hawaii UTC-10h)



Procedure and Experiences

- **The migration is a technical and social challenge!**
- KEK CC shutdown in August 2016 set the time line
- Main goal: Don't lose any information!
- Integrate services and tools into an existing IT infrastructure
- Technology changes are necessary
- 1-to-1 copy of contents is usually not possible
- Migration tools are not always available
- Quite some manual work required
- ~10 Expert teams taking care of the various services
- Cooperation of collaboration members needed
- Wiki by far most complicated
- Migration from SVN to *GIT* most controversial within Belle II
- Maintaining infrastructure vs. maintaining contents
- For the migration: ~10 FTEy / 6 month
- Further maintenance: 1 FTE (in the first 2 years)

Issue Tracking

JIRA Software

- ATLASSIAN tool
- Single-Sign-On (SSO)
- Connected to confluence
- Migration from redmine
- ATLASSIAN plug-ins

Mailing Lists

- sympa
- list@belle2.org
- Separate accounts
- Migration from sympa
- Cleanup and reordering

Code Repository Build Services

Stash Bamboo

- ATLASSIAN tools
- Single-Sign-On (SSO)
- Connected to confluence
- Migration from SVN to *GIT*
- Hooks plug-in

Xen

- Virtual machines in *Xen*
- Buildbot on VMs
- Build slaves for OS'

ubuntu redhat

Agenda Service

INDICO

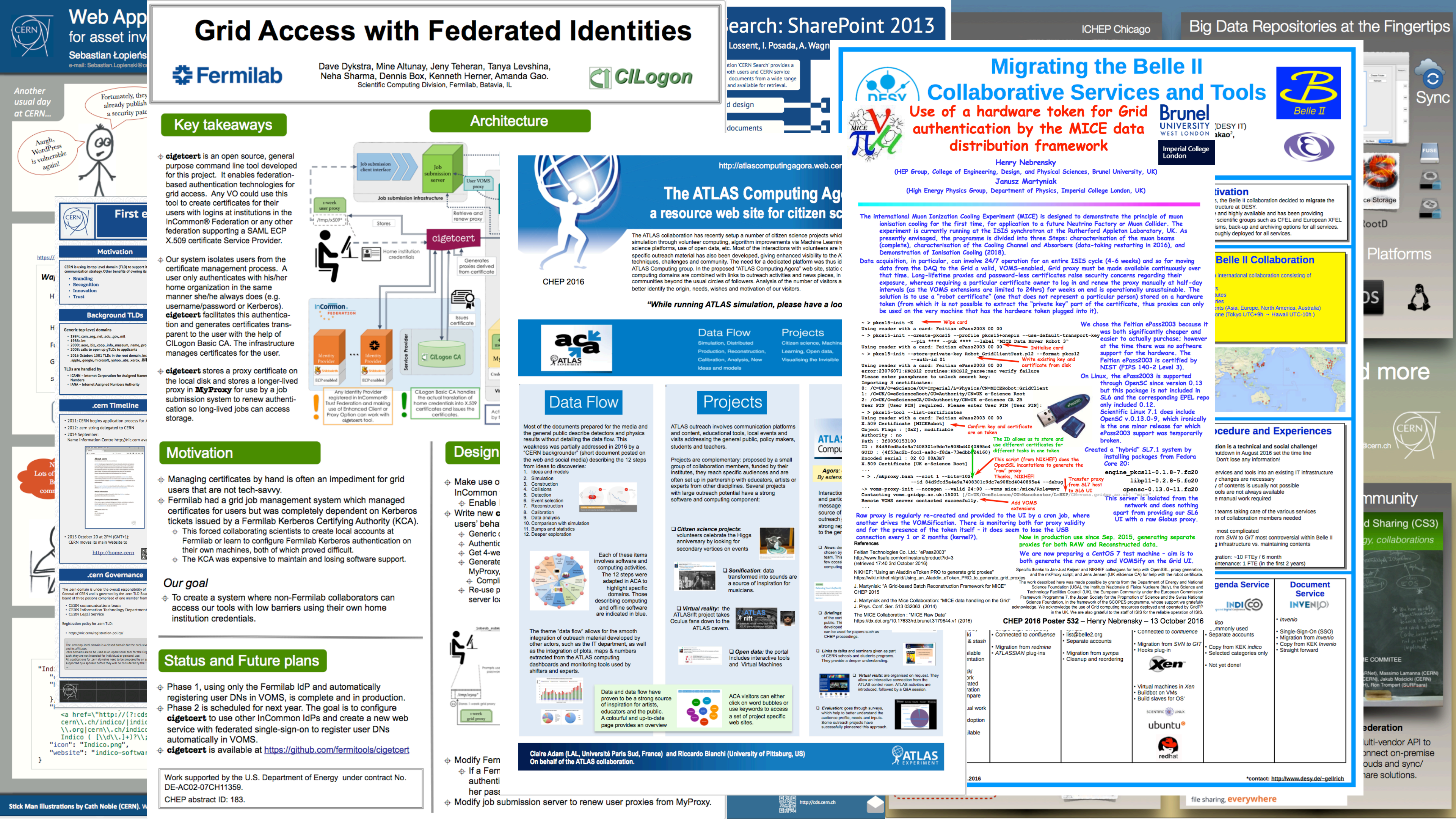
- *Indico*
- Commonly used
- Separate accounts
- Copy from KEK *indico*
- Selected categories only
- Not yet done!

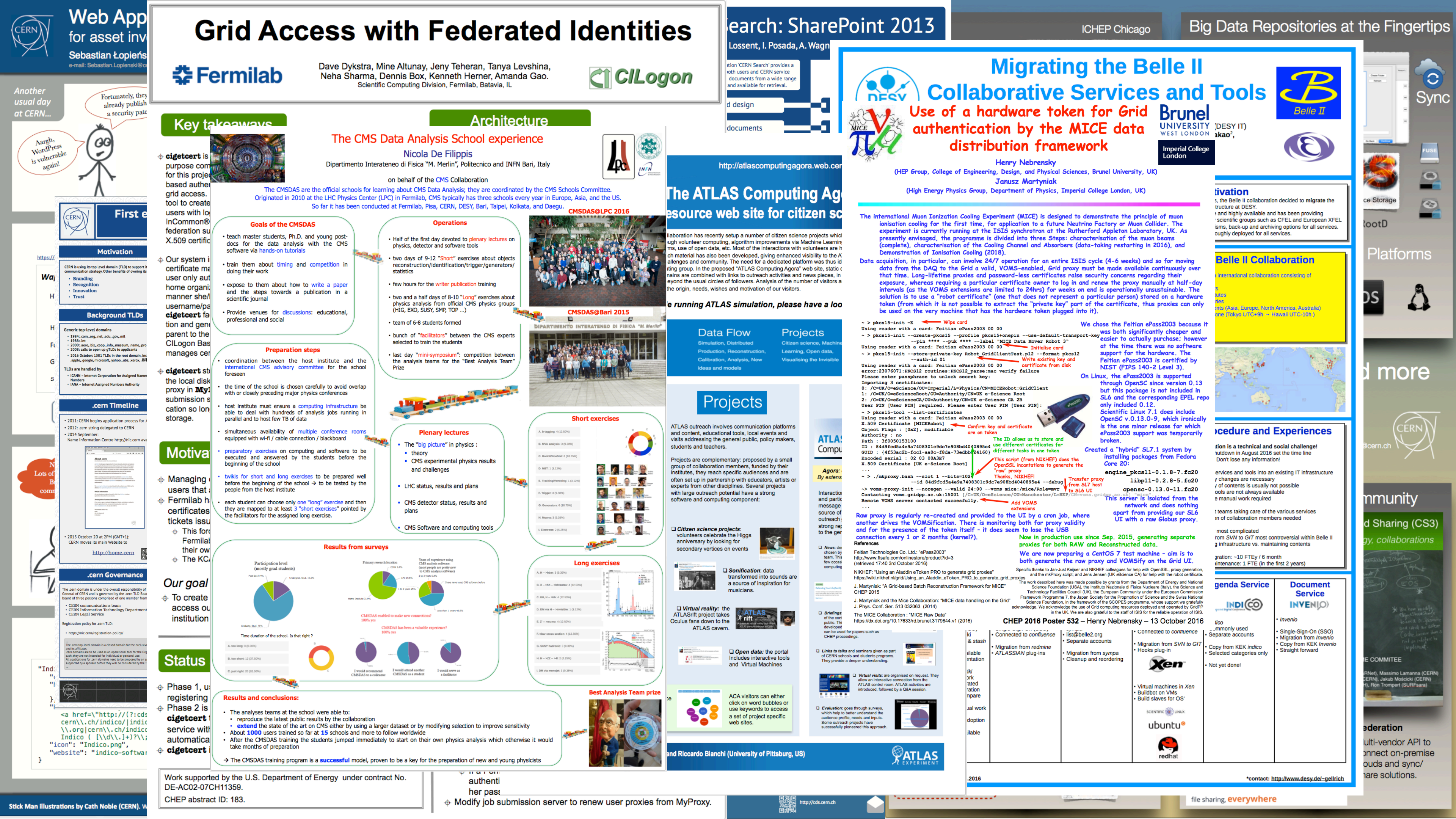
Document Service

INVENIO

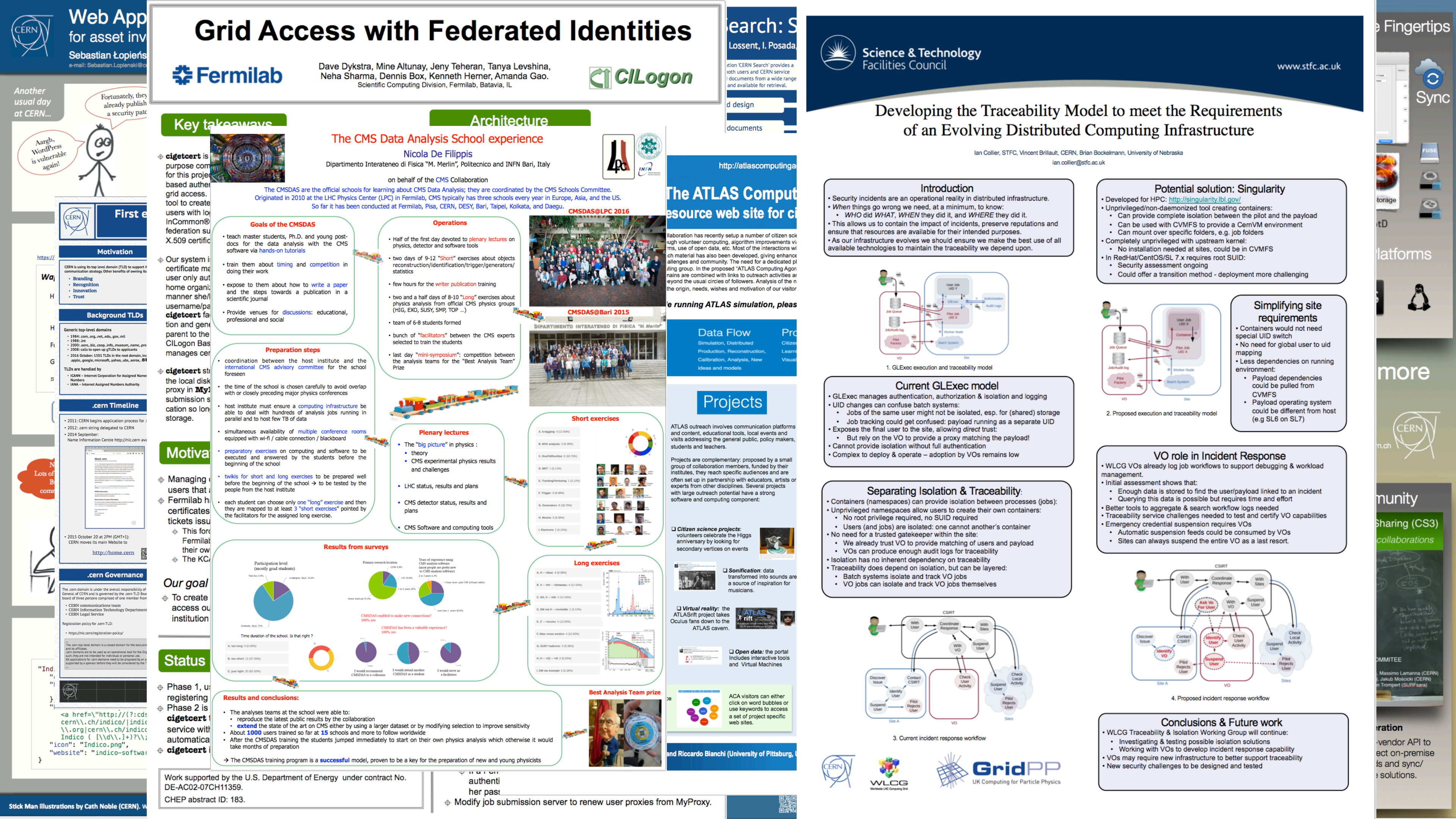
- *invenio*
- Single-Sign-On (SSO)
- Migration from *invenio*
- Copy from KEK *invenio*
- Straight forward

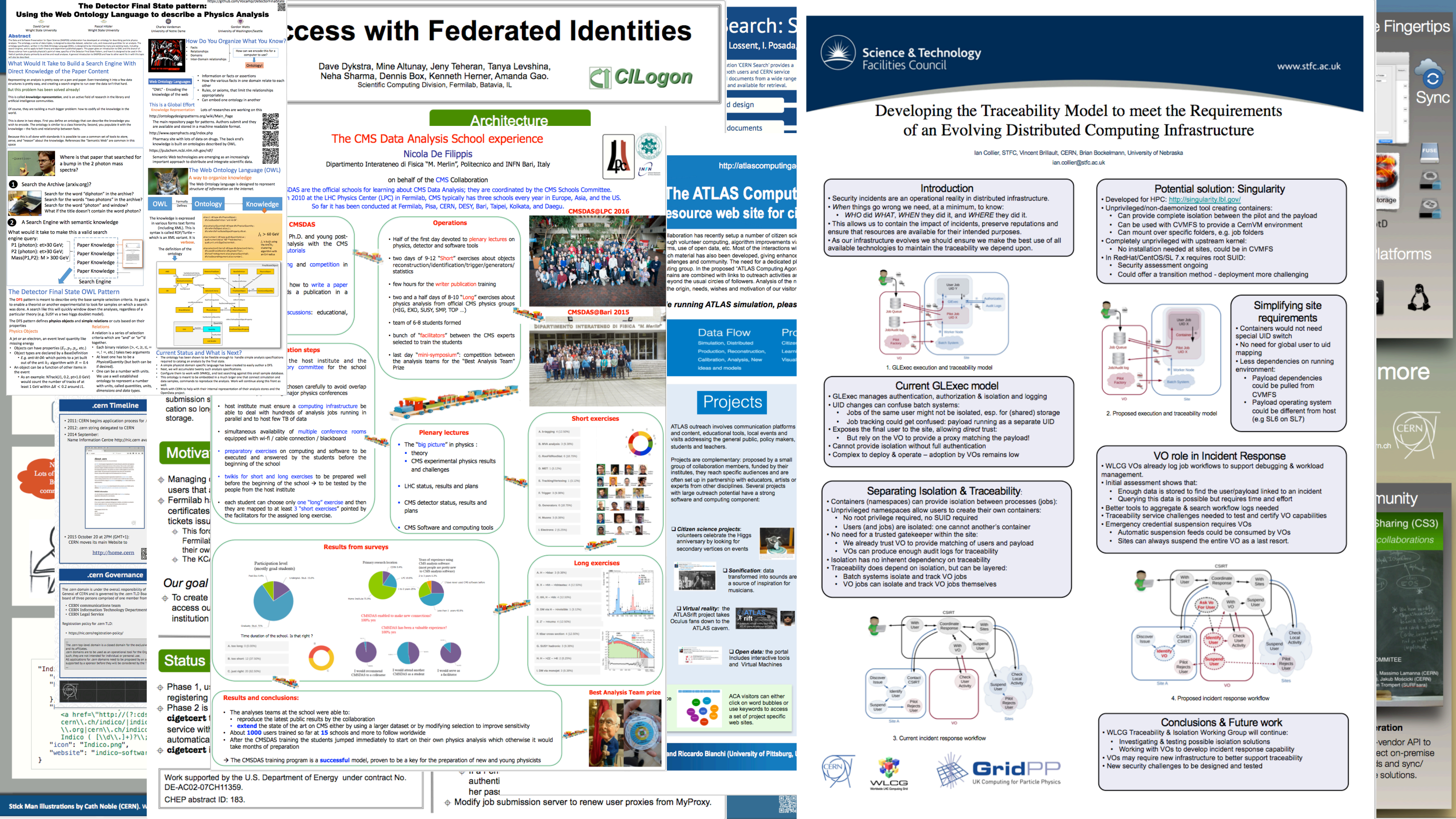
*contact: http://www.desy.de/~gellrich


Web App
for asset inv
Sebastian Łopieńs
Another usual day at CERN...
Grid Access with Federated Identities
Fermilab
Dave Dykstra, Mine Altunay, Jeny Teheran, Tanya Levshina,
Neha Sharma, Dennis Box, Kenneth Herner, Amanda Gao.
Scientific Computing Division, Fermilab, Batavia, IL
CILogon
Key takeaways
cigetcert is an open source, general purpose command line tool developed for this project. It enables federation-based authentication technologies for grid access. Any VO could use this tool to create certificates for their users with logins at institutions in the InCommon® Federation or any other federation supporting a SAML ECP X.509 certificate Service Provider.
Our system isolates users from the certificate management process. A user only authenticates with his/her home organization in the same manner she/he always does (e.g. username/password or Kerberos). cigetcert facilitates this authentication and generates certificates transparent to the user with the help of CILogon Basic CA. The infrastructure manages certificates for the user.
cigetcert stores a proxy certificate on the local disk and stores a longer-lived proxy in MyProxy for use by a job submission system to renew authentication so long-lived jobs can access storage.
Architecture
cigetcert
Motivation
Managing certificates by hand is often an impediment for grid users that are not tech-savvy.
Fermilab had a grid job management system which managed certificates for users but was completely dependent on Kerberos tickets issued by a Fermilab Kerberos Certifying Authority (KCA).
This forced collaborating scientists to create local accounts at Fermilab or learn to configure Fermilab Kerberos authentication on their own machines, both of which proved difficult.
The KCA was expensive to maintain and losing software support.
Our goal
To create a system where non-Fermilab collaborators can access our tools with low barriers using their own home institution credentials.
Status and Future plans
Phase 1, using only the Fermilab IdP and automatically registering user DNs in VOMS, is complete and in production.
Phase 2 is scheduled for next year. The goal is to configure cigetcert to use other InCommon IdPs and create a new web service with federated single-sign-on to register user DNs automatically in VOMS.
cigetcert is available at https://github.com/fermitools/cigetcert
Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.
CHEP abstract ID: 183.
Design
Modify job submission server to renew user proxies from MyProxy.
earch: SharePoint 2013
ICHEP Chicago
Big Data Repositories at the Fingertips
http://atlascomputingagora.web.cern.ch
The ATLAS Computing Agora
a resource web site for citizen science
CHEP 2016
Data Flow
Projects
Claire Adam (LAL, Université Paris Sud, France) and Riccardo Bianchi (University of Pittsburg, US)
On behalf of the ATLAS collaboration.
Migrating the Belle II Collaborative Services and Tools
Use of a hardware token for Grid authentication by the MICE data distribution framework
Henry Nebrensky
(HEP Group, College of Engineering, Design, and Physical Sciences, Brunel University, UK)
Janusz Martyniak
(High Energy Physics Group, Department of Physics, Imperial College London, UK)
CHEP 2016 Poster 532 – Henry Nebrensky – 13 October 2016
Belle II Collaboration
Agenda Service
Document Service
Sync
Platforms
Stick Man illustrations by Cath Noble (CERN).
file sharing everywhere



Grid Access with Federated Identities
Fermilab
Dave Dykstra, Mine Altunay, Jeny Teheran, Tanya Levshina, Neha Sharma, Dennis Box, Kenneth Herner, Amanda Gao.
Scientific Computing Division, Fermilab, Batavia, IL
CILogon
Key takeaways
Architecture
The CMS Data Analysis School experience
Nicola De Filippis
Dipartimento Interateneo di Fisica "M. Merlin", Politecnico and INFN Bari, Italy
on behalf of the CMS Collaboration
Goals of the CMSDAS
Operations
Preparation steps
Plenary lectures
Short exercises
Long exercises
Results from surveys
Results and conclusions:
Best Analysis Team prize
CMSDAS@LPC 2016
CMSDAS@Bari 2015
Web App for asset inv
Sebastian Łopieński
Motivation
Background TLDs
.cern Timeline
.cern Governance
Motiva
Our goal
Status
Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.
CHEP abstract ID: 183.
Modify job submission server to renew user proxies from MyProxy.
Stick Man illustrations by Cath Noble (CERN).
earch: SharePoint 2013
http://atlascomputingagora.web.cern
The ATLAS Computing Ag
esource web site for citizen sc
Data Flow
Projects
Projects
and Riccardo Bianchi (University of Pittsburg, US)
ATLAS EXPERIMENT
ICHEP Chicago
Big Data Repositories at the Fingertips
Migrating the Belle II Collaborative Services and Tools
Belle II
Use of a hardware token for Grid authentication by the MICE data distribution framework
Brunel UNIVERSITY WEST LONDON
Imperial College London
Henry Nebrensky
(HEP Group, College of Engineering, Design, and Physical Sciences, Brunel University, UK)
Janusz Martyniak
(High Energy Physics Group, Department of Physics, Imperial College London, UK)
CHEP 2016 Poster 532 – Henry Nebrensky – 13 October 2016
Belle II Collaboration
Agenda Service
Document Service
Sync
Platforms
d more
mmunity
file sharing, everywhere


Web App
for asset inv
Sebastian Łopień
Another usual day at CERN...
Fortunately, they already publish a security patc
Aargh, WordPress is vulnerable again!
Grid Access with Federated Identities
Fermilab
Dave Dykstra, Mine Altunay, Jeny Teheran, Tanya Levshina, Neha Sharma, Dennis Box, Kenneth Herner, Amanda Gao.
Scientific Computing Division, Fermilab, Batavia, IL
CILogon
Key takeaways
Architecture
The CMS Data Analysis School experience
Nicola De Filippis
Dipartimento Interateneo di Fisica "M. Merlin", Politecnico and INFN Bari, Italy
on behalf of the CMS Collaboration
The CMSDAS are the official schools for learning about CMS Data Analysis; they are coordinated by the CMS Schools Committee.
Originated in 2010 at the LHC Physics Center (LPC) in Fermilab, CMS typically has three schools every year in Europe, Asia, and the US.
So far it has been conducted at Fermilab, Pisa, CERN, DESY, Bari, Taipei, Kolkata, and Daegu.
CMSDAS@LPC 2016
Goals of the CMSDAS
Operations
CMSDAS@Bari 2015
Preparation steps
Plenary lectures
Short exercises
Long exercises
Results from surveys
Results and conclusions:
Best Analysis Team prize
The CMSDAS training program is a successful model, proven to be a key for the preparation of new and young physicists
Motivation
Background TLDs
.cern Timeline
.cern Governance
Our goal
Status
Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.
CHEP abstract ID: 183.
Modify job submission server to renew user proxies from MyProxy.
Stick Man illustrations by Cath Noble (CERN).
earch: S
Lossent, I. Posada
The ATLAS Comput
esource web site for ci
Data Flow
Projects
Science & Technology Facilities Council
www.stfc.ac.uk
Developing the Traceability Model to meet the Requirements of an Evolving Distributed Computing Infrastructure
Ian Collier, STFC, Vincent Brillault, CERN, Brian Bockelmann, University of Nebraska
ian.collier@stfc.ac.uk
Introduction
Potential solution: Singularity
1. GLExec execution and traceability model
Current GLExec model
2. Proposed execution and traceability model
Simplifying site requirements
Separating Isolation & Traceability
VO role in Incident Response
3. Current incident response workflow
4. Proposed incident response workflow
Conclusions & Future work
WLCG
GridPP
UK Computing for Particle Physics
Fingertips
Sync
more
munity
Sharing (CS3)
collaborations



The Detector Final State pattern:
Using the Web Ontology Language to describe a Physics Analysis
What Would It Take to Build a Search Engine With Direct Knowledge of the Paper Content
How Do You Organize What You Know?
The Web Ontology Language (OWL)
OWL
Ontology
Knowledge
Where is that paper that searched for a bump in the 2 photon mass spectra?
Search the Archive (arxiv.org)?
A Search Engine with semantic knowledge
The Detector Final State OWL Pattern
Current Status and What is Next?
.cern Timeline
.cern Governance
ccess with Federated Identities
Dave Dykstra, Mine Altunay, Jeny Teheran, Tanya Levshina,
Neha Sharma, Dennis Box, Kenneth Herner, Amanda Gao.
Scientific Computing Division, Fermilab, Batavia, IL
CILogon
Architecture
Motiva
Our goal
Status
Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.
CHEP abstract ID: 183.
Modify job submission server to renew user proxies from MyProxy.
The CMS Data Analysis School experience
Nicola De Filippis
Dipartimento Interateneo di Fisica "M. Merlin", Politecnico and INFN Bari, Italy
on behalf of the CMS Collaboration
CMSDAS
Operations
CMSDAS@LPC 2016
CMSDAS@Bari 2015
Plenary lectures
Short exercises
Long exercises
Results from surveys
Results and conclusions:
Best Analysis Team prize
The CMSDAS training program is a successful model, proven to be a key for the preparation of new and young physicists
earch: S
The ATLAS Comput
esource web site for ci
Data Flow
Projects
Citizen science projects
Sonification
Virtual reality
Open data
Stick Man illustrations by Cath Noble (CERN).
Science & Technology Facilities Council
www.stfc.ac.uk
Developing the Traceability Model to meet the Requirements of an Evolving Distributed Computing Infrastructure
Ian Collier, STFC, Vincent Brillault, CERN, Brian Bockelmann, University of Nebraska
ian.collier@stfc.ac.uk
Introduction
• Security incidents are an operational reality in distributed infrastructure.
• When things go wrong we need, at a minimum, to know:
• WHO did WHAT, WHEN they did it, and WHERE they did it.
• This allows us to contain the impact of incidents, preserve reputations and ensure that resources are available for their intended purposes.
• As our infrastructure evolves we should ensure we make the best use of all available technologies to maintain the traceability we depend upon.
1. GLExec execution and traceability model
Current GLExec model
• GLExec manages authentication, authorization & isolation and logging
• UID changes can confuse batch systems:
• Jobs of the same user might not be isolated, esp. for (shared) storage
• Job tracking could get confused: payload running as a separate UID
• Exposes the final user to the site, allowing direct trust:
• But rely on the VO to provide a proxy matching the payload!
• Cannot provide isolation without full authentication
• Complex to deploy & operate – adoption by VOs remains low
Separating Isolation & Traceability:
• Containers (namespaces) can provide isolation between processes (jobs):
• Unprivileged namespaces allow users to create their own containers:
• No root privilege required, no SUID required
• Users (and jobs) are isolated: one cannot another's container
• No need for a trusted gatekeeper within the site:
• We already trust VO to provide matching of users and payload
• VOs can produce enough audit logs for traceability
• Isolation has no inherent dependency on traceability
• Traceability does depend on isolation, but can be layered:
• Batch systems isolate and track VO jobs
• VO jobs can isolate and track VO jobs themselves
3. Current incident response workflow
Potential solution: Singularity
• Developed for HPC: http://singularity.lbl.gov/
• Unprivileged/non-daemonized tool creating containers:
• Can provide complete isolation between the pilot and the payload
• Can be used with CVMFS to provide a CernVM environment
• Can mount over specific folders, e.g. job folders
• Completely unprivileged with upstream kernel:
• No installation needed at sites, could be in CVMFS
• In RedHat/CentOS/SL 7.x requires root SUID:
• Security assessment ongoing
• Could offer a transition method - deployment more challenging
2. Proposed execution and traceability model
Simplifying site requirements
• Containers would not need special UID switch
• No need for global user to uid mapping
• Less dependencies on running environment:
• Payload dependencies could be pulled from CVMFS
• Payload operating system could be different from host (e.g SL6 on SL7)
VO role in Incident Response
• WLCG VOs already log job workflows to support debugging & workload management.
• Initial assessment shows that:
• Enough data is stored to find the user/payload linked to an incident
• Querying this data is possible but requires time and effort
• Better tools to aggregate & search workflow logs needed
• Traceability service challenges needed to test and certify VO capabilities
• Emergency credential suspension requires VOs
• Automatic suspension feeds could be consumed by VOs
• Sites can always suspend the entire VO as a last resort.
4. Proposed incident response workflow
Conclusions & Future work
• WLCG Traceability & Isolation Working Group will continue:
• Investigating & testing possible isolation solutions
• Working with VOs to develop incident response capability
• VOs may require new infrastructure to better support traceability
• New security challenges to be designed and tested
GridPP
UK Computing for Particle Physics
WLCG
e Fingertips
Sync
latforms
more
munity
Sharing (CS3)
collaborations
ration

The Detector Fina[l State Pattern]
Using the Web Ontology Language

David Carral
Wright State University

Pascal Hitzler
Wright State University

Abstract

The Data and Software Preservation for Open Science (DASPOS) collaboration has developed an ontology for describing particle physics analyses. The ontology, a series of data triples, is designed to describe dataset, selection cuts, and measured quantities for an analysis. The ontology specification, written in the Web Ontology Language (OWL), is designed to be interpreted by many pre-existing tools, including search engines, and to apply to both theory and experiment published papers. This paper gives an introduction to OWL and this branch of library science from a particle physicist's point of view, specifics of the Detector Final State Pattern, and how it is designed to be used in the field of particle physics primarily to archive and recall analyses. A general introduction to DASPOS and how its other work fits in with this topic will also be described.

What Would It Take to Build a Search Engine With Direct Knowledge of the Paper Content

Representing an analysis is pretty easy on a pen and paper. Even translating it into a few data structures is pretty easy, and creating a search engine to run over the data isn't that hard.

But this problem has been solved already!

This is called *knowledge representation*, and is an active field of research in the library and artificial intelligence communities.

Of course, they are tackling a much bigger problem: how to codify all the knowledge in the world.

This is done in two steps. First you define an ontology that can describe the knowledge you wish to encode. The ontology is similar to a class hierarchy. Second, you populate it with the knowledge – the facts and relationship between facts.

Because this is all done with standards it is possible to use a common set of tools to *store, serve, and "reason"* about the knowledge. References like "Semantic Web" are common in this space

Where is that paper that searched for a bump in the 2 photon mass spectra?

1. Search the Archive (arxiv.org)?

Search for the word "diphoton" in the archive?
Search for the words "two photons" in the archive?
Search for the word "photon" and window?
What if the title doesn't contain the word photon?

2. A Search Engine with semantic knowledge

What would it take to make this a valid search engine query:
P1 (photon): et>30 GeV;
P2 (photon): et>30 GeV;
Mass(P1,P2): M > 300 GeV

Paper Knowledge
Paper Knowledge
Paper Knowledge
Paper Knowledge
Search Engine

The Detector Final State OWL Pattern

The DFS pattern is meant to describe only the base sample selection criteria. Its goal i[s] to enable a theorist or another experimentalist to look for samples on which a search was done. A search like this will quickly window down the analyses, regardless of a particular theory (e.g. SUSY vs a two higgs doublet model).

The DFS pattern defines **physics objects** and **simple relations** or cuts based on their properties

Physics Objects

A jet or an electron, an event level quantity like missing energy
- Objects can have properties (E_T, p_T, p_x, etc.)
- Object types are declared by a BaseDefinition
  - *E.g. anti-kt-04*: which points to a jet built out of the anti k_T algorithm with $R = 0.4$
- An object can be a function of other items in the event
  - As an example: NTrack(J1, 0.2, pt>1.0 GeV) would count the number of tracks of at least 1 GeV within $\Delta R < 0.2$ around J1.

Relations

A relation is a series of selection criteria which are "and" or "or"'d together.
- Each binary relation (>, <, ≥, ≤, =, ! =, etc.) takes two argumen[ts]
- At least one has to be a *PhysicalQuantity* (but both can if desired).
- One can be a number with unit[s]. We use a well established ontology to represent a numbe[r] with units, called *quantities, un[its], dimensions and data types.*

.cern Timeline

- 2011: CERN begins application process for .[cern]
- 2012: .cern string delegated to CERN
- 2014 September: Name Information Centre http://nic.cern ava[ilable]
- 2015 October 20 at 2PM (GMT+1): CERN moves its main Website to

http://home.cern

.cern Governance

The .cern domain is under the overall responsibility of [the Director-]General of CERN and is governed by the .cern TLD Boa[rd, a] board of three persons comprised of one member from
- CERN communications team
- CERN Information Technology Department
- CERN Legal Service

Registration policy for .cern TLD:
- https://nic.cern/registration-policy/

```
"Ind[ico]": {
  ...
  "<a href=\"http://(?:cds
  cern\\.ch/indico/|indi
  \\.org|cern\\.ch/indic
  Indico ( [\\d\\.]+)?\\;
  "icon": "Indico.png",
  "website": "indico-softwa
}
```

Stick Man illustrations by Cath Noble (CERN). W

INFN DESY KIT CESNET INDIGO - DataCloud

The INDIGO-DataCloud AAI

A. Ceccanti, M. Hardt, B. Wegh, P. Millar, S. Licehammer, M. Caberletti, E. Vianello

The INDIGO **Authentication and Authorization Infrastructure** (AAI) leverages and extends existing standards (OpenID Connect, OAuth2, SCIM) to enable secure composition of resources from multiple providers in support of scientific applications. The central **Identity and Access Management** (IAM) service provides tools to implement brokered user authentication, identity harmonization, account linking as well as Virtual Organization (VO) management. Identity information is provided to relying services via **OpenID Connect,** which allows simpler integration in off-the-shelf software. The **Token Translation Service** (TTS) integrates services that do not directly support OpenID Connect, by creating credentials on-demand from the identity information provided by the IAM. This poster introduces the main INDIGO AAI components and highlights the main architectural decisions that were taken during the first year of the INDIGO project.

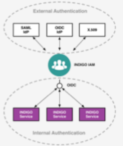
1. Identity = OpenID Connect

The **INDIGO** [1] **AAI** Identity layer leverages the **OpenID Connect** [2] standard to provide user authentication and identity information to INDIGO services. This approach provides several advantages:

- Standardized and widely adopted solution
- Accommodation of several authentication mechanisms (via identity brokering)
- Simplified integration in relying services
- Native support for delegation and offline access
- Native support for dynamic client registration

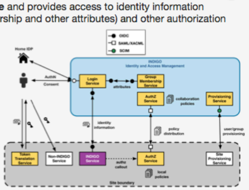
2. Identity harmonization and brokering

The **INDIGO Identity and Access Management (IAM)** [3] service deals with user authentication and identity brokering, allowing users to authenticate with local username and password, SAML, X.509 certificates, or via a remote OpenID Connect provider (e.g., Google). The different identities are linked to a single INDIGO user profile, which provides each user with a **unique, persistent identifier**. This identifier is then used for auditing, accounting and authorization at all relying services. Other attributes, like group membership information, can be linked to the user profile, and then exposed to relying services via standard **OpenID Connect i**nterfaces.



3. Authorization = OAuth 2 + XACML

INDIGO services expose functionality through HTTP APIs protected by **OAuth 2** [4]: only agents presenting a valid and trusted **OAuth access token** are granted access to INDIGO services. This token is obtained by client applications from the **INDIGO IAM service** and provides access to identity information (e.g., group membership and other attributes) and other authorization information (e.g., OAuth scopes). Fine-grained, powerful and distributed attribute-based authorization is implemented by integrating the **Argus authorization service** [5] with the INDIGO AAI identity layer. This provides policy distribution and centralized XACML policy management.



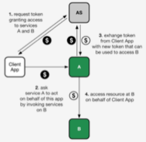
4. Identity provisioning

INDIGO IAM leverages the standard **System for Cross Domain Identity Management (SCIM)** [8] version 2.0 to implement identity provisioning, de-provisioning and management.

The SCIM APIs provides means to **propagate identity and group information to relying services**, to implement, for instance, **dynamic account creation and other resource lifecycle management at various levels of the INDIGO infrastructure** depending on events related to user identity status.

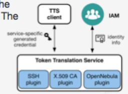
5. Delegation and offline access

OAuth and OpenID Connect **support delegation and offline access natively**. To support **chained delegation across services**, in which a component can act both as a service and as a client for another downstream service, the INDIGO IAM provides a partial implementation of the **OAuth token exchange draft standard** [6]. The token exchange is used in particular to implement **controlled delegation of offline access rights across applications** (i.e., the ability to execute tasks on behalf of a user while the user is not connected).



6. Token translation and non-HTTP services

INDIGO AAI integrates services that rely on different authentication mechanisms (e.g., X.509 certificates, SSH keys, Amazon S3 keys) via the **INDIGO Token Translation Service (TTS)** [7]. The **TTS** translates an OpenID Connect authentication assertion, like the one issued by the IAM service, into one of the supported downstream service credentials. The TTS currently supports the generation of:
- SSH keys
- X.509 certificates
- OpenNebula username/password credentials



7. References

1. The INDIGO-DataCloud project: https://www.indigo-datacloud.eu/
2. OpenID Connect: http://openid.net/connect/
3. INDIGO Identity and Access Management: https://github.com/indigo-iam/iam
4. OAuth 2: https://tools.ietf.org/html/rfc6749
5. Argus Authorization Service: http://argus-documentation.readthedocs.io/en/latest/
6. OAuth token exchange draft standard: http://bit.ly/oauth-token-exchange
7. The INDIGO Token Translation Service: https://github.com/indigo-dc/tts
8. SCIM: http://www.simplecloud.info/

INDIGO - DataCloud receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement RIA 653549

[CERN S]earch: S[...]
Lossent, I. Posada, [...]

[...]tion 'CERN Search' provides a [...]oth users and CERN service [...] documents from a wide range [...]and available for retrieval.

[...]d design

documents

[ht]tp://atlascomputinga[...]

[ATL]AS Comput[ing] [W]eb site for ci[...]

[...]setup a number of citizen scie[nce ...] algorithm improvements vi[a ...]tc. Most of the interactions wi[...]n developed, giving enhance[...]. The need for a dedicated pl[...]sed "ATLAS Computing Agor[a ...] links to outreach activities ar[...]f followers. Analysis of the n[...] and motivation of our visitor[s]

[ATLA]S simulation, pleas[e]

Flow: Distributed Reconstruction, Analysis, New [m]odels

Pro[...]: Citize[n ...] Learni[ng ...] Visual[...]

[Pro]jects

[...]es communication platforms [...]al tools, local events and [g]eneral public, policy makers,

[...]ntary: proposed by a small members, funded by their [s]pecific audiences and are [...]ship with educators, artists or [dis]ciplines. Several projects [...]ential have a strong [...]g component:

[Pro]jects: [...] the Higgs [...]ng for [...]n events

[...] *Sonification*: data transformed into sounds are a source of inspiration for musicians.

❑ *Open data:* the portal Includes interactive tools and Virtual Machines

ACA visitors can either click on word bubbles or use keywords to access a set of project specific web sites.

(University of Pittsburg, [...]

Science & Technology Facilities Council

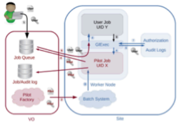
www.stfc.ac.uk

Developing the Traceability Model to meet the Requirements of an Evolving Distributed Computing Infrastructure

Ian Collier, STFC, Vincent Brillault, CERN, Brian Bockelmann, University of Nebraska
ian.collier@stfc.ac.uk

Introduction

- Security incidents are an operational reality in distributed infrastructure.
- *When* things go wrong we need, at a minimum, to know:
  - *WHO* did *WHAT*, *WHEN* they did it, and *WHERE* they did it.
- This allows us to contain the impact of incidents, preserve reputations and ensure that resources are available for their intended purposes.
- As our infrastructure evolves we should ensure we make the best use of all available technologies to maintain the traceability we depend upon.



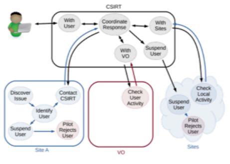
1. GLExec execution and traceability model

Current GLExec model

- GLExec manages authentication, authorization & isolation and logging
- UID changes can confuse batch systems:
  - Jobs of the same user might not be isolated, esp. for (shared) storage
  - Job tracking could get confused: payload running as a separate UID
- Exposes the final user to the site, allowing *direct* trust:
  - But rely on the VO to provide a proxy matching the payload!
- Cannot provide isolation without full authentication
- Complex to deploy & operate – adoption by VOs remains low

Separating Isolation & Traceability:

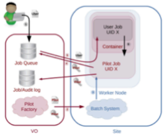
- Containers (namespaces) can provide isolation between processes (jobs):
- Unprivileged namespaces allow users to create their own containers:
  - No root privilege required, no SUID required
  - Users (and jobs) are isolated: one cannot another's container
- No need for a trusted gatekeeper within the site:
  - We already trust VO to provide matching of users and payload
  - VOs can produce enough audit logs for traceability
- Isolation has no inherent dependency on traceability
- Traceability does depend on isolation, but can be layered:
  - Batch systems isolate and track VO jobs
  - VO jobs can isolate and track VO jobs themselves



3. Current incident response workflow

Potential solution: Singularity

- Developed for HPC: http://singularity.lbl.gov/
- Unprivileged/non-daemonized tool creating containers:
  - Can provide complete isolation between the pilot and the payload
  - Can be used with CVMFS to provide a CernVM environment
  - Can mount over specific folders, e.g. job folders
- Completely unprivileged with upstream kernel:
  - No installation needed at sites, could be in CVMFS
- In RedHat/CentOS/SL 7.x requires root SUID:
  - Security assessment ongoing
  - Could offer a transition method - deployment more challenging



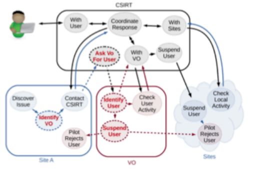
2. Proposed execution and traceability model

Simplifying site requirements

- Containers would not need special UID switch
- No need for global user to uid mapping
- Less dependencies on running environment:
  - Payload dependencies could be pulled from CVMFS
  - Payload operating system could be different from host (e.g SL6 on SL7)

VO role in Incident Response

- WLCG VOs already log job workflows to support debugging & workload management.
- Initial assessment shows that:
  - Enough data is stored to find the user/payload linked to an incident
  - Querying this data is possible but requires time and effort
- Better tools to aggregate & search workflow logs needed
- Traceability service challenges needed to test and certify VO capabilities
- Emergency credential suspension requires VOs
  - Automatic suspension feeds could be consumed by VOs
  - Sites can always suspend the entire VO as a last resort.



4. Proposed incident response workflow

Conclusions & Future work

- WLCG Traceability & Isolation Working Group will continue:
  - Investigating & testing possible isolation solutions
  - Working with VOs to develop incident response capability
- VOs may require new infrastructure to better support traceability
- New security challenges to be designed and tested

CERN WLCG GridPP UK Computing for Particle Physics

[...] Fingertips

Sync

[S]torage

[...]tD

[P]latforms

[...] more

[Com]munity

[Cloud Storage S]haring (CS3)
[...] collaborations

[C]OMMITEE
[...] Massimo Lamanna (CERN)
[...] Jakub Moscicki (CERN)
[...]n Trompert (SURFsara)

[...]ration
[...]-vendor API to [...]ect on-premise [...]ds and sync/ [...] solutions.

The Detector Final State Pattern Using the Web Ontology Language

David Carral, Wright State University — Pascal Hitzler, Wright State University

What Would It Take to Build a Search Engine With Direct Knowledge of the Paper Content

But this problem has been solved already!

Where is that paper that searched for a bump in the 2 photon mass spectra?

1 Search the Archive (arxiv.org)?

Search for the word "diphoton" in the archive?
Search for the words "two photons" in the archive?
Search for the word "photon" and window?
What if the title doesn't contain the word photon?

2 A Search Engine with semantic knowledge

What would it take to make this a valid search engine query:
P1 (photon): et>30 GeV;
P2 (photon): et>30 GeV;
Mass(P1,P2): M > 300 GeV

Paper Knowledge
Paper Knowledge
Paper Knowledge
Paper Knowledge
Search Engine

The Detector Final State OWL Pattern

Physics Objects — Relations

.cern Timeline

- 2011: CERN begins application process for ...
- 2012: .cern string delegated to CERN
- 2014 September: Name Information Centre http://nic.cern ...
- 2015 October 20 at 2PM (GMT+1): CERN moves its main Website to http://home.cern

.cern Governance

Stick Man Illustrations by Cath Noble (CERN).

The INDIGO-DataCloud AAI

A. Ceccanti, M. Hardt, B. Wegh, P. Millar, S. Licehammer, M. Caberletti, E. Vianel...

The INDIGO **Authentication and Authorization Infrastructure** (AAI) leverages and extends ... SCIM) to enable secure composition of resources from multiple providers in support of scien... **Access Management** (IAM) service provides tools to implement brokered user authenticati... well as Virtual Organization (VO) management. Identity information is provided to relying serv... simpler integration in off-the-shelf software. The **Token Translation Service** (TTS) integrates ... Connect, by creating credentials on-demand from the identity information provided by the I... AAI components and highlights the main architectural decisions that were taken during the f...

1. Identity = OpenID Connect

The **INDIGO** [1] **AAI** Identity layer leverages the **OpenID Connect** [2] standard to provide user authentication and identity information to INDIGO services. This approach provides several advantages:

- Standardized and widely adopted solution
- Accommodation of several authentication mechanisms (via identity brokering)
- Simplified integration in relying services
- Native support for delegation and offline access
- Native support for dynamic client registration

2. Identity harmonization and brokering

The **INDIGO Identity and Access Management (IAM)** [3] service deals with user authentication and identity brokering, allowing users to authenticate with local username and password, SAML, X.509 certificates, or via a remote OpenID Connect provider (e.g., Google). The different identities are linked to a single INDIGO user profile, which provides each user with a **unique, persistent identifier**. This identifier is then used for auditing, accounting and authorization at all relying services. Other attributes, like group membership information, can be linked to the user profile, and then exposed to relying services via standard **OpenID Connect** interfaces.

3. Authorization = OAuth 2 + XACML

INDIGO services expose functionality through HTTP APIs protected by **OAuth 2** [4]: only agents presenting a valid and trusted **OAuth access token** are granted access to INDIGO services. This token is obtained by client applications from the **INDIGO IAM service** and provides access to identity information (e.g., group membership and other attributes) and other authorization information (e.g., OAuth scopes). Fine-grained, powerful and distributed attribute-based authorization is implemented by integrating the **Argus authorization service** [5] with the INDIGO AAI identity layer. This provides policy distribution and centralized XACML policy management.

4. Identity pr...

INDIGO IAM leverage... **Domain Identity Man...** implement identity pr... management.

The SCIM APIs provid... **information to relyin... account creation an... various levels of the...** related to user identit...

5. Delegation

OAuth and OpenID C... **delegation and offlin...** To support **chained d... services**, in which a ... both as a service and ... another downstream ... IAM provides a partial... the **OAuth token exc... standard** [6]. The token exchange i... to implement **controll... offline access rights... applications** (i.e., the... tasks on behalf of a u...

6. Token tran...

INDIGO AAI integrates ... mechanisms (e.g., X.5... **INDIGO Token Trans...** The **TTS** translates an... one issued by the IAM... supported downstrea... TTS currently supports the generation of:

- SSH keys
- X.509 certificates
- OpenNebula username/password credentials

7. References

1. The INDIGO-DataCloud project: https://www.indigo-datacloud.eu/
2. OpenID Connect: http://openid.net/connect/
3. INDIGO Identity and Access Management: https://github.com/indigo-iam/iam
4. OAuth 2: https://tools.ietf.org/html/rfc6749
5. Argus Authorization Service: http://argus-documentation.readthedocs.io/en/latest/
6. OAuth token exchange draft standard: http://bit.ly/oauth-token-exchange
7. The INDIGO Token Translation Service: https://github.com/indigo-dc/tts
8. SCIM: http://www.simplecloud.info/

INDIGO - DataCloud receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement RIA 653549

A Security Monitoring Framework For Virtualization Based HEP Infrastructures

Andres Gomez, Camilo Lara and Udo Kebschull for the ALICE Collaboration
Infrastructure and Computer Systems for Data Processing (IRI), Goethe-University Frankfurt

Project Overview

Create a secure architecture for Scientific Grids by:
- Security by isolation
- Monitoring data mining
- Machine learning based anomaly detection
- Intrusion prevention and detection
- Automated defensive actions
- Proof of concept in Alice Grid

Security by Isolation

Grid jobs run isolated in Linux containers:
- All components run by unprivileged users
- We can run jobs by Grid users, instead of local users
- It allows to gather Job specific security data
- And to protect the underline system (physical servers)

Monitoring data mining - Trainning and Testing Dataset

"Normal" data:
- Run random Alice grid Jobs (a very limited amount)
- Collect and store system call information + parameters (e.g. network connections)
- Store all this info in a manageable way for further construction of machine learning dataset.

"Malicious" data:
- Use a set of malware samples (http://virusshare.com)
- Isolated environment (No connection with Grid, not even with Internet)
- Cuckoo sandbox => VM extra protection
- SecurityOnion (IDS Linux distribution)
- Sysdig

Machine Learning Based Security Monitoring

- Input Features: system calls and parameters
- Support Vector Machines
- Recurrent Neural Networks
- Deep Learning
- TensorFlow
- Scikit-learn

Intrusion Prevention and Detection

Create Job execution profiles:
- Raise alarms and execute predefined actions on attack attemps
- Adapt to dynamic environment
- Work together with the Alice Grid monitoring system and current available IDS
- Integrate with the Alice Grid middleware - AliEn

Proof of Concept and Testing Environment

- System calls
- Connections
- Resource usage

- 5 Ubuntu 14.04 nodes
- Centos 6 containers
- AliEn interface with kubernetes+Mesos and also Docker Swarm as cluster batch systems
- CVMFS being installed at the hosts and shared as a volume inside the AliEn container
- Currently the idea is to execute one Job per container
- An isolated hardened machine for malware data collection

Computing in High Energy and Nuclear Physics, CHEP 2016 — Contact: andres.gomez@cern.ch

---

Developing the Traceability Model to meet the Requirements of an Evolving Distributed Computing Infrastructure

Science & Technology Facilities Council — www.stfc.ac.uk

Ian Collier, STFC, Vincent Brillault, CERN, Brian Bockelmann, University of Nebraska
ian.collier@stfc.ac.uk

Introduction

...operational reality in distributed infrastructure. ...need, at a minimum, to know: ...*EN* they did it, and *WHERE* they did it. ...the impact of incidents, preserve reputations and ...available for their intended purposes. ...ves we should ensure we make the best use of all ...aintain the traceability we depend upon.

Potential solution: Singularity

- Developed for HPC: http://singularity.lbl.gov/
- Unprivileged/non-daemonized tool creating containers:
  - Can provide complete isolation between the pilot and the payload
  - Can be used with CVMFS to provide a CernVM environment
  - Can mount over specific folders, e.g. job folders
- Completely unprivileged with upstream kernel:
  - No installation needed at sites, could be in CVMFS
- In RedHat/CentOS/SL 7.x requires root SUID:
  - Security assessment ongoing
  - Could offer a transition method - deployment more challenging

Simplifying site requirements

- Containers would not need special UID switch
- No need for global user to uid mapping
- Less dependencies on running environment:
  - Payload dependencies could be pulled from CVMFS
  - Payload operating system could be different from host (e.g SL6 on SL7)

...c execution and traceability model

2. Proposed execution and traceability model

...rrent GLExec model

...tication, authorization & isolation and logging ...batch systems: ...er might not be isolated, esp. for (shared) storage ...et confused: payload running as a separate UID ...the site, allowing *direct* trust: ...provide a proxy matching the payload! ...without full authentication ...rate – adoption by VOs remains low

...g Isolation & Traceability

...can provide isolation between processes (jobs): ...allow users to create their own containers: ...ired, no SUID required ...solated: one cannot another's container ...keeper within the site: ...o provide matching of users and payload ...ugh audit logs for traceability ...dependency on traceability ...on isolation, but can be layered: ...and track VO jobs ...nd track VO jobs themselves

VO role in Incident Response

- WLCG VOs already log job workflows to support debugging & workload management.
- Initial assessment shows that:
  - Enough data is stored to find the user/payload linked to an incident
  - Querying this data is possible but requires time and effort
- Better tools to aggregate & search workflow logs needed
- Traceability service challenges needed to test and certify VO capabilities
- Emergency credential suspension requires VOs
  - Automatic suspension feeds could be consumed by VOs
  - Sites can always suspend the entire VO as a last resort.

3. Current incident response workflow

4. Proposed incident response workflow

Conclusions & Future work

- WLCG Traceability & Isolation Working Group will continue:
  - Investigating & testing possible isolation solutions
  - Working with VOs to develop incident response capability
- VOs may require new infrastructure to better support traceability
- New security challenges to be designed and tested

GridPP — UK Computing for Particle Physics

...earch: S... Lossent, I. Posada...

Open data: the portal includes interactive tools and Virtual Machines

ACA visitors can either click on word bubbles or use keywords to access a set of project specific web sites.

(University of Pittsburg...

... Fingertips — Sync — ...latforms — ...more — ...munity — ...Sharing (CS3) — collaborations — ...ration — ...vendor API to ...ect on-premise ...ls and sync/ ...solutions.


See you in 2018