eScience Requirements (aka FIM4R)

David Kelsey (STFC-RAL)

AAA Study Workshop, Brussels

12 Jul 2012

Overview

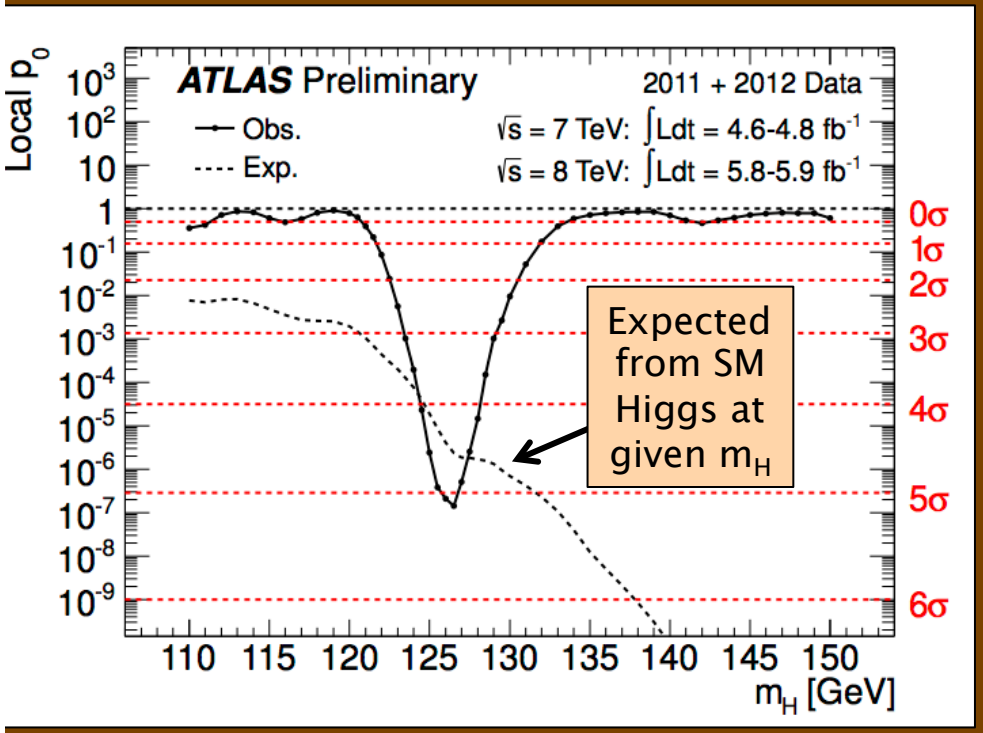
- Some background
- FIM4R workshops and our paper
- Communities and Requirements
- *Will show agreed common requirements rather than concentrate on individual communities*
- *Nothing on accounting (but EGI and others do have requirements)*

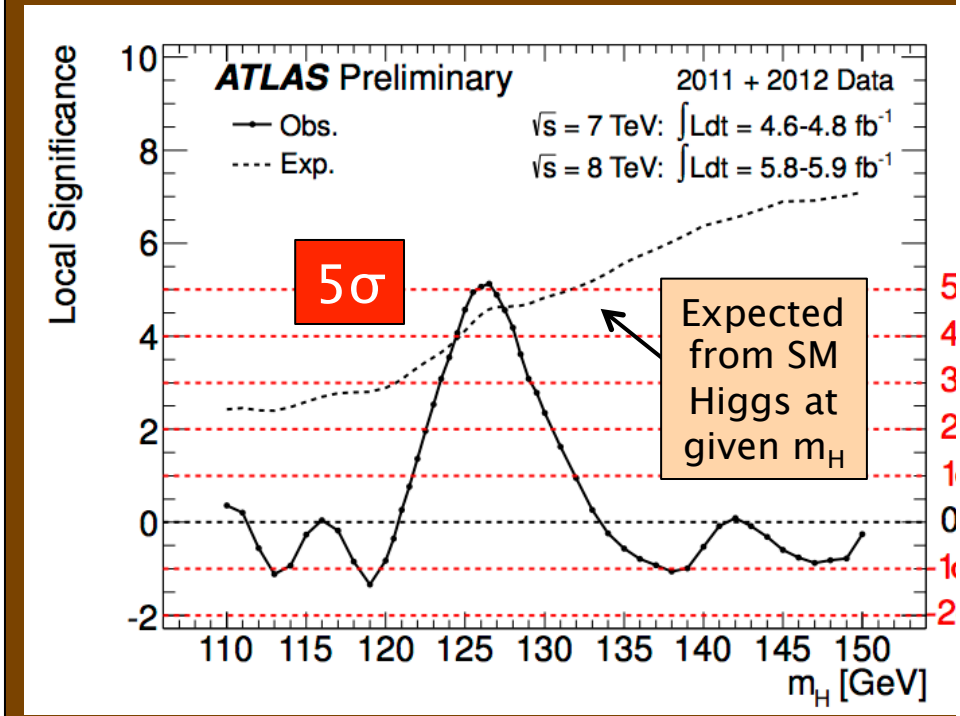
About me

- Head of Particle Physics Computing Group at STFC-RAL
- Many roles in Security, AAI, IdM etc
  - Including EGI, GridPP, WLCG, IGTF, SCI, FIM4R

- Why does “Science” need all this distributed data and computing?
- A recent example (CERN 4 July 12)
  - 2 slides (ATLAS & CMS)

Combined results: the excess

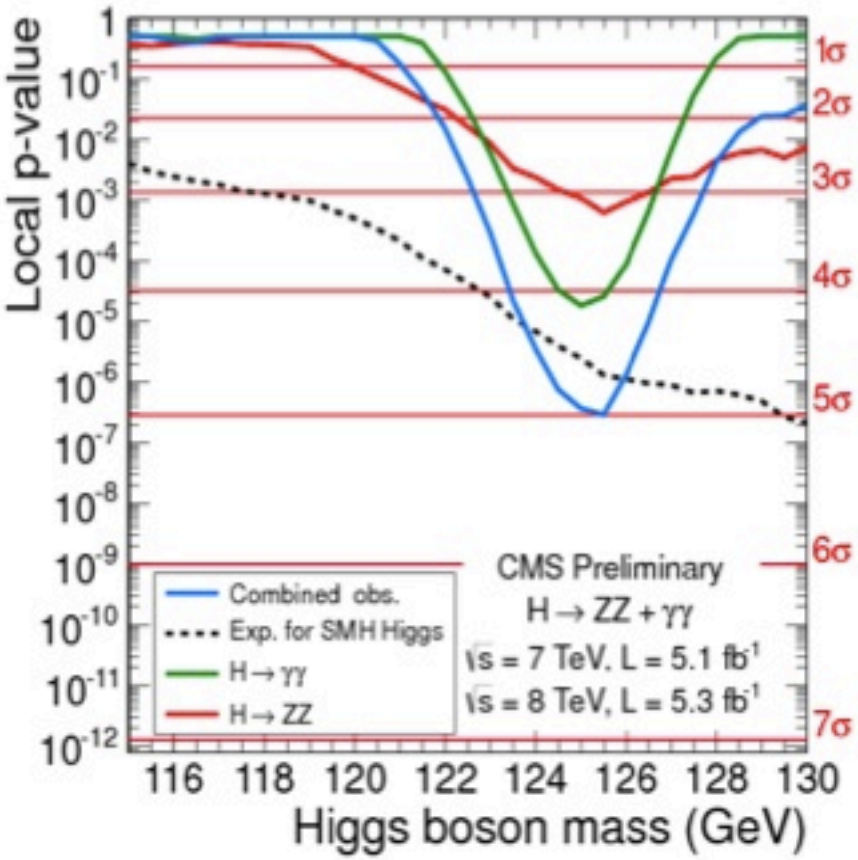




Maximum excess observed at	m_H = 126.5 GeV
Local significance (including energy-scale systematics)	5.0 σ
Probability of background up-fluctuation	3 x 10^{-7}
Expected from SM Higgs m_H=126.5	4.6 σ

Global significance: 4.1–4.3 σ (for LEE over 110–600 or 110–150 GeV)

Characterization of excess near 125 GeV



- high sensitivity, high mass resolution channels: γγ+4l
  - γγ: 4.1 σ excess
  - 4 leptons: 3.2 σ excess
  - near the same mass 125 GeV
- comb. significance: **5.0 σ**
- expected significance for SM Higgs: **4.7 σ**

- Speedy analysis of so much data only possible because of the success of WLCG, EGI, OSG and other related infrastructures
  - Including the Grid AAI !!
  - Not forgetting the accelerator, the experiments, the physicists, the engineers etc etc!

Background

- Issue of IdM raised by IT leaders from EIROforum labs during their IT working group meeting in January 2011 at ESA
  - CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL
- These laboratories, as well as national and regional research organizations, are facing similar challenges
  - Scientific data deluge means massive quantities of data
  - needs to be accessed by expanding user bases in dynamic collaborations across organisational and national boundaries
- "Facebook" generation demands all the tools (work & social) integrate smoothly
- Also encouraged by EEF and eIRG
- Global problem, not just EU
- *"Science" changed to "Research" to include Humanities*

Federated IdM for Research (FIM4R)

- A collaborative effort started in June 2011
- Not just EIROForum. Include many ESFRI projects and providers and infrastructures
  - Be inclusive
- Involves photon & neutron facilities, social science & humanities, high energy physics, climate science, life sciences and fusion energy
- Workshops included participation by HTC and HPC infrastructures, TERENA, IGTF, Geant/eduGAIN, middleware developers …

Science & Technology Facilities Council

Workshops (2)

- 4 workshops to date
  - link to Jun 2012 agenda below (other links contained within)

https://indico.cern.ch/conferenceDisplay.py?confId=191892

- Prepared a paper that documents common requirements, a common vision and recommendations
  - To research communities, identity federations, funding bodies
- Paper: CERN-OPEN-2012-006: https://cdsweb.cern.ch/record/1442597

Science & Technology Facilities Council

The communities

user community	other projects	# users	chosen technology	status	IGTF
photon/neutron facilities	EUROFEL, PanData, CRISP	>30,000 visiting researchers per year	Shibboleth/SAML	Umbrella prototype	no
Social Sciences and Humanities	DARIAH, CLARIN, CESSDA, DASISH, Bamboo	hundreds now, potential for 10000+ across SSH	Shibboleth/SAML	Prototype CLARIN Service Provider federation deployed in 3 countries	yes
high energy physics	WLCG	10,000+ globally	X509	production	yes
earth sciences	Federation, GENESI-DEC, CMIP5, Metafor, IS-ENES, CORDEX, Exarch, Climate Data Exchange	5000+ for CIMP5	OpenID, X.509 and SAML	production - earth system grid	not yet but foresee for EGI integration
life sciences	ELIXIR, BioMedBridges, BBMRI, NCoEDG & potentially 10 BMI ESFRI projects	3 million researchers access data via EBI website each year	not chosen yet	Security is included in BioMedBridges project workplan and a pilot project is being planned with EGI	no

Common Requirements

- User friendliness
  - Many users use infrequently
- Browser and non-browser federated access
- Bridging between communities
- Multiple technologies and translators
  - Translation will often need to be dynamic
- Open standards and sustainable licenses
  - For interoperability and sustainability
- Different Levels of Assurance
  - When credentials are translated, LoA provenance to be preserved
- Authorisation under community and/or facility control
  - Externally managed IdPs cannot fulfil this role
- Well defined semantically harmonised attributes
  - For interoperable authorisation
  - Likely to be very difficult to achieve!

Requirements (2)

- Flexible and scalable IdP attribute release policy
  - Different communities and different SPs need different attributes
  - Negotiate with IdF not all IdPs – for scaling
- Attributes must be able to cross national borders
  - Data protection/privacy considerations
- Attribute aggregation for authorisation
- Privacy and data protection to be addressed with community-wide individual identities
  - We need to identify individuals
    - E.g. ethical committees can require names, addresses, supervisors to grant access

Science & Technology Facilities Council

Operational Requirements

- Risk analysis
- Traceability
  - Audit trails include IdPs
- Security incident response
  - To include all IdPs and SPs
- Transparency of policies
  - To gain trust of SPs and users
- Reliability and resilience
- Smooth transition (from today's production)
- Easy integration with local SP
  - SP likely to want to support multiple AuthN technologies

Legal, Policy & Trust

- Contracts or SLAs between communities and IDFs must be scalable
  - Include maximum number of participants
  - Bi-lateral agreements will not scale
- Standards of Trust (or Codes of Conduct) similar to IGTF approach is an attractive scalable solution

Common vision statement

A common policy and trust framework for Identity Management based on existing structures and federations either presently in use by or available to the communities. This framework must provide researchers with unique electronic identities authenticated in multiple administrative domains and across national boundaries that can be used together with community defined attributes to authorize access to digital resources

Questions?