Software and Computing for Hyper-Kamiokande

Francesca Di Lodovico

18 June 2014
Second EU Hyper-Kamiokande Open Meeting
CERN, Geneva

Outline

- Agreement w/ Super-Kamiokande and T2K
  - Description of the Agreement
  - Current Status
- Release
  - First Hyper-K release
- Production
  - Overview
  - Data Retrieval
- Web Pages (documentation)
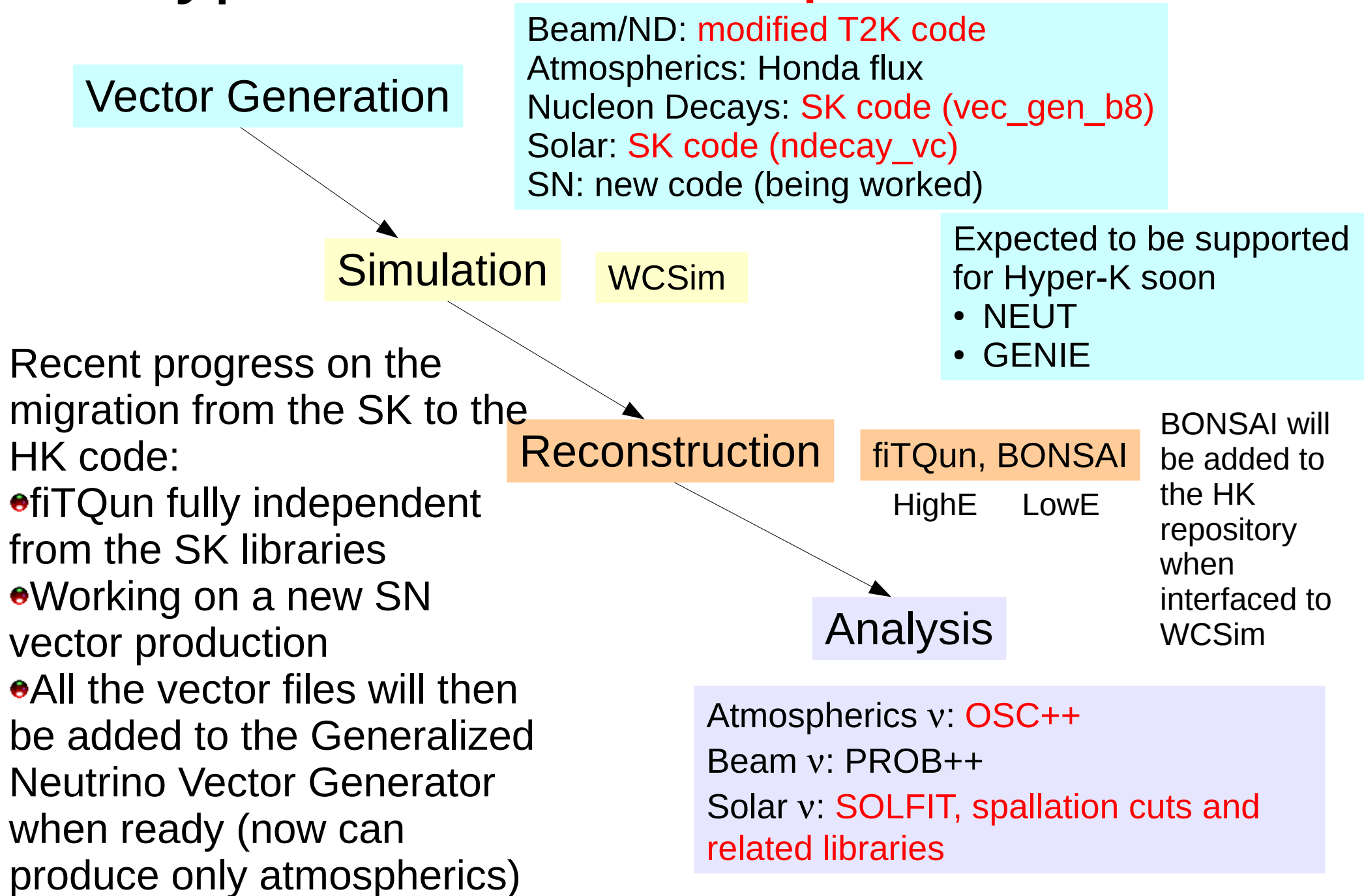 - Wiki
- Computing Model Strategy

Report

- Request from HK working group for an use of SK softaware and SK data in the HK software delopement work.
  - They asked to make a similar agreement between SK and T2K

Answer to the request

1) The SK Executive Committee feels that Hyper-Kamiokande is a natural extension of Super-Kamiokande, therefore, we should cooperate as much as possible.

2) We think that a HK working group will develop their own simulation code based on the 'real' HK geometry and other parameters since the HK detector size and shape are quite different from SK. As HK develops its own simulation code, then HK software becomes less dependent on SK software. We believe that HK should start their code development now.

3) Certainly, we agree that HK may adopt concepts of the SK software and if necessary to copy SK codes into HK codes freely. Once done, HK may modify the codes as they wish without reporting all the modifications back to SK. We would like to be informed of any modifications that could also improve the SK code.

4) SK software should not be used for any other purpose than for HK simulation and analysis software development.

5) You mentioned the T2K-SK agreement as an example, but the situation of T2K is quite different from HK. T2K uses SK detector as a far detector, therefore, T2K definitely needs SK software and SK data, which is the reason we have made such an agreement. HK will not use the SK detector, so we do not need the same kind of agreement, as you have proposed.

6) SK will not form a group to have responsibility for maintaining analysis software of the SK detector for the HK group.

7) SK does not see any reason that the HK working group needs to use SK data for the HK software development. SK believes that HK needs only the MC generator, detector simulator and data analysis codes, which we are willing to provide, but not data.

3

Hyper-K code vs Super-K vs T2K

Vector Generation

Beam/ND: modified T2K code
Atmospherics: Honda flux
Nucleon Decays: SK code (vec_gen_b8)
Solar: SK code (ndecay_vc)
SN: new code (being worked)

Simulation   WCSim

Expected to be supported for Hyper-K soon
- NEUT
- GENIE

Recent progress on the migration from the SK to the HK code:
- fiTQun fully independent from the SK libraries
- Working on a new SN vector production
- All the vector files will then be added to the Generalized Neutrino Vector Generator when ready (now can produce only atmospherics)

Reconstruction   fiTQun, BONSAI

HighE   LowE

BONSAI will be added to the HK repository when interfaced to WCSim

Analysis

Atmospherics ν: OSC++
Beam ν: PROB++
Solar ν: SOLFIT, spallation cuts and related libraries

Hyper-K Software Release

- All needed documentation is stored in https://wiki.hyperk.org/Software

- Use git (https://github.com/) as code-repository
  - Git is a modern code management system, being used by several particle physics experiments (eg. SNO+)
  - Hosted at QMUL. Aim to mirror to other sites.
  - To get an account email your public key to both:
    - Francesca Di Lodovico f.di.lodovico@qmul.ac.uk
    - Alex Owen r.a.owen@qmul.ac.uk

- Add the relevant packages to the repository.
  - All open-source packages (e.g. ROOT, GEANT4, etc.) are treated as "third party" packages (no developement in git).
  - Production scripts for HK are added to git.

Hyper-K Repository

- A total of 7 external open source packages:

CLHEP   fiTQun   GEANT4   Prob3++   ROOT   WChSandBox WCSim

- are in the git repository as:

hk-clhep hk-fitqun hk-geant4 hk-prob3++ hk-root hk-wchsandbox hk-WCsim

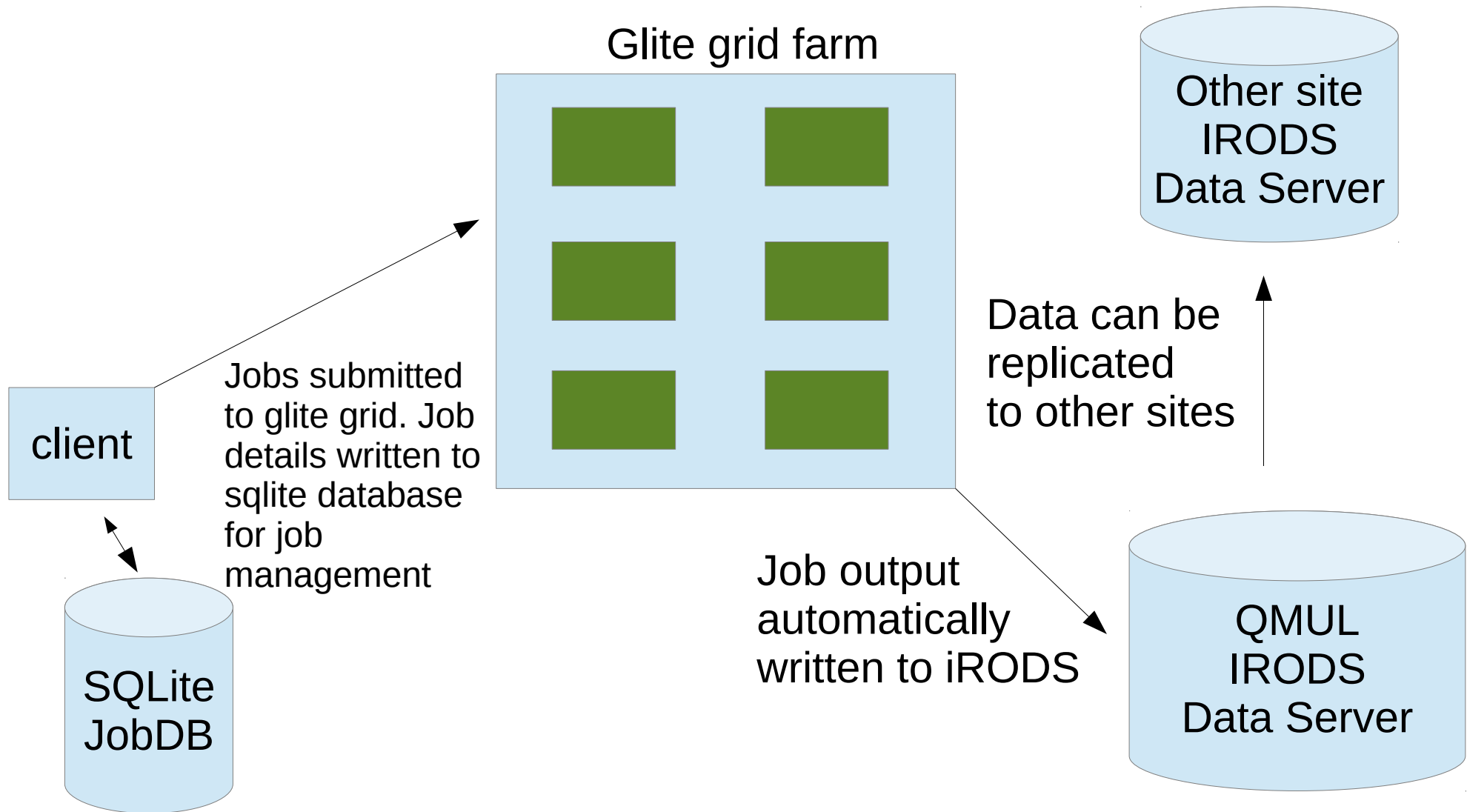- Also a total of 7 Hyper-K packages are in the repository:

hk-eventdisplay hk-geant4 hk-hyperk hk-irods hk-simplefitter hk-TITUSanalysis hk-utils

- More packages (eg for vector file production) will be added as needed.

Hyper-K Repository

- Instructions to download the files are on the Hyper-K Wiki
https://wiki.hyperk.org/

  > git clone ssh://git@poset.ph.qmul.ac.uk/hk-hyperk hk-hyperk
  > cd hk-hyperk
  > ./hk_get_release.sh

- The above command will retrieve only basic files to produce reconstructed events.

- Planning to move to a more refined way to download the release assuming several configurations:
Hyper-K production, Hyper-K analysis, Hyper-K sensitivity, Hyper-K TITUS etc. More will be added when the packages are uploaded.

Hyper-K Production

Glite grid farm

Other site IRODS Data Server

client

Jobs submitted to glite grid. Job details written to sqlite database for job management

Data can be replicated to other sites

Job output automatically written to iRODS

SQLite JobDB

QMUL IRODS Data Server

Hyper-K Production

- Release deployed to CVMFS. Binaries read from there

- Jobs managed through scripts that interface to local sqlite database

  - Database schema has hooks for cloud jobs (not implemented yet)

  - Schema keeps track of job status

  - In principle, it allows running production at different centres (although need to use run-ranges to avoid clashing jobs)

- Fairly stable simulation release of WCSim, but aiming to add information from HPD testing in the eGADS tank soon.

- Almost final version of fiTQun for July production (170s/event, it could be sligthly reduced). Validating current results before starting the production.

Hyper-K Simulation

- Data files read from CVFMS along with executable

- Output written to iRODS

  - Log files written to 'success' collection or 'failed' collection depending on return code

- Currently run 50K nu-mode beam events (stopped to check output then run for 1M events):

  - 5 runs (10K per run). No failures. Jobs fit within expontential profile below (from previous run)

Job Duration

10k events per job
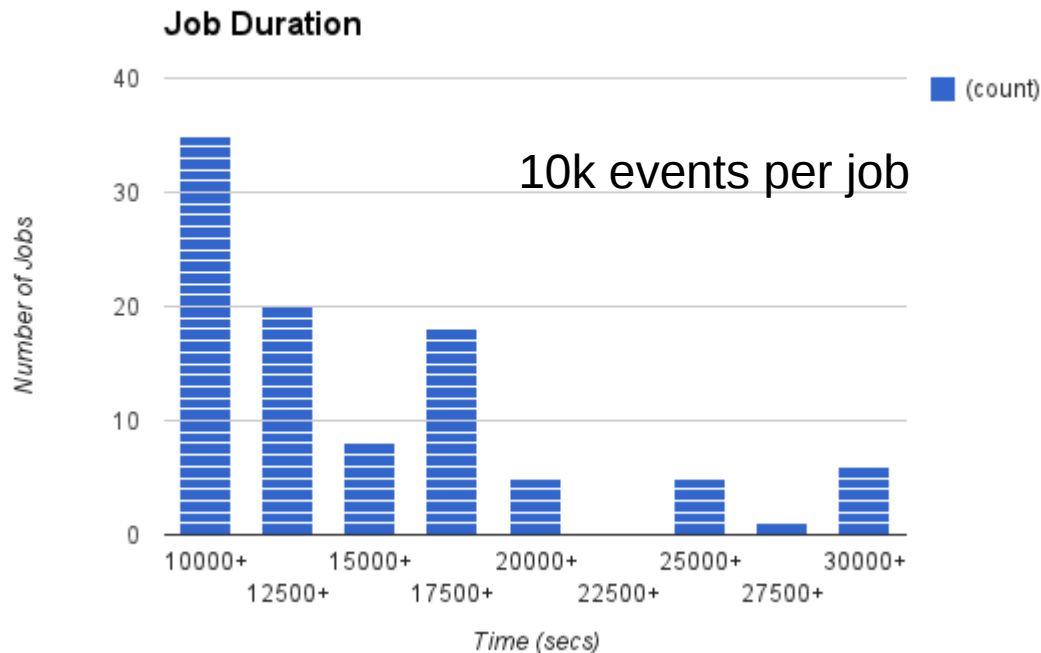
Job duration exhibit exponential decay – long tail due to other processes running on grid Nodes, filesystem activity etc. Majority of jobs take <18s/evt.

Hyper-K Reconstruction

- Read simulated data from iRODS to compute node, process and write out to iRODS

- Improvements to reco mean we can safely run 200 events/job:

    - Safely within the the max allowed time per grid job (some events take 5 min to process)

    - Running with 23hr proxy cert

- Split jobs:

    - 1 simu job = 50 reco jobs (200 events/job)

- Processed 29K (146 jobs) so far of test sample

Reco rocessing time profile
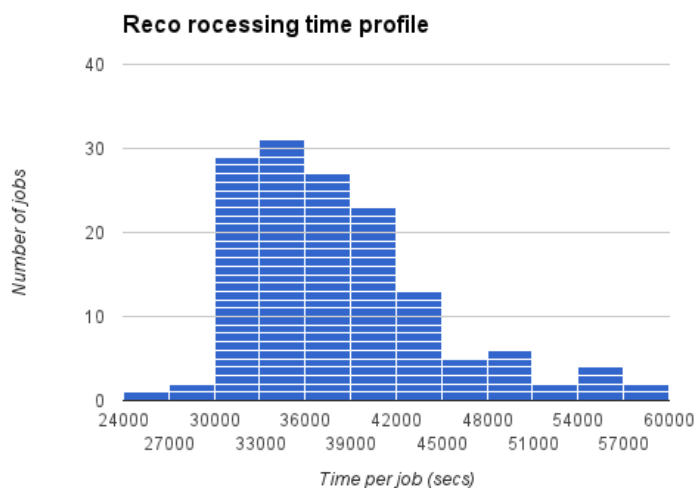
On limited stats average time ~3min/evt.

Hyper-K Data Storage

- **Data available in iRODS (no Grid certificate needed):**
  - /QMULZone2/home/hyperk/production/simu/
  - /QMULZone2/home/hyperk/production/reco/
- Data can being replicated to sites
- Production version maps to frozen versions of WCSim and fiTQun. File names have form
  - neutvect-<run#>-<sim#>.root
  - fitqun-<sim#>-<rec#>.root
  - In this case 1 run = 1 sim file
  - Log files written to 'success', 'failed' collections.

Next Steps for Hyper-K Production

- Validate current small production. Aim to start producing next week beam simulated events to be used for the July collaboration meeting.

- In the longer term:

  - setup web page that will automatically be populated with current production

  - Implement data replication

  - investigate deployment on other Grids

Web Site

As a by-product of the Grid VO (Virtual Organization), we have a domain hyperk.org.

Two web sites:

Public web site: http://www.hyperk.org/

Working wiki web site: http://wiki.hyperk.org
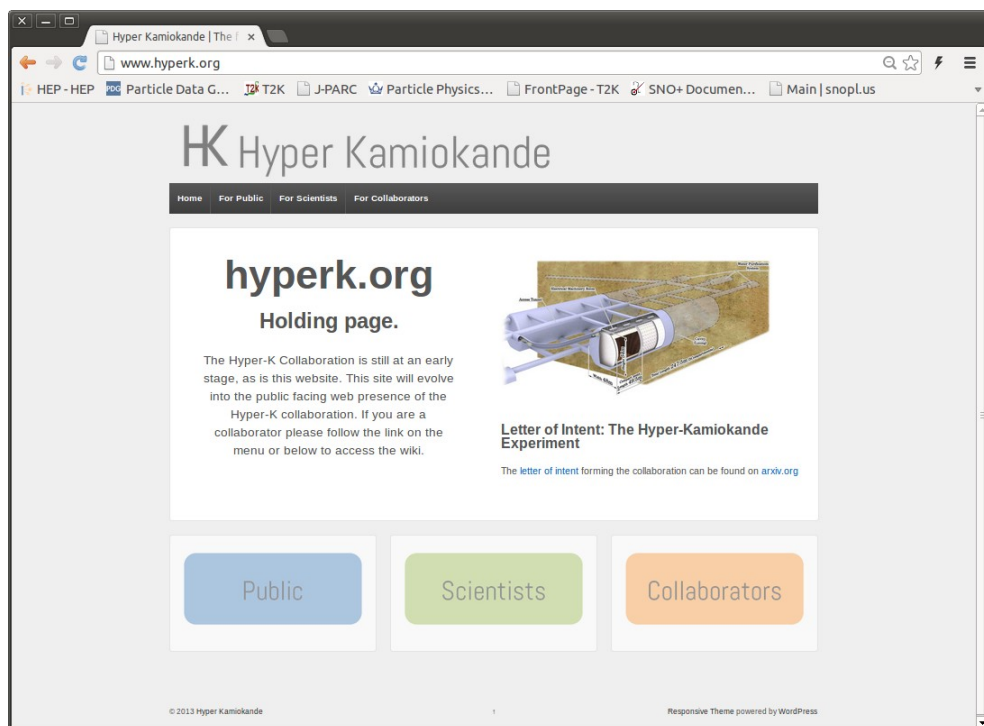
Added basic info to the public web site.

We will need to buy more domains for Hyper-K

Hyper-K Wiki

Working wiki web site: http://wiki.hyperk.org

The Wiki is meant to be the working web site for the experiment.
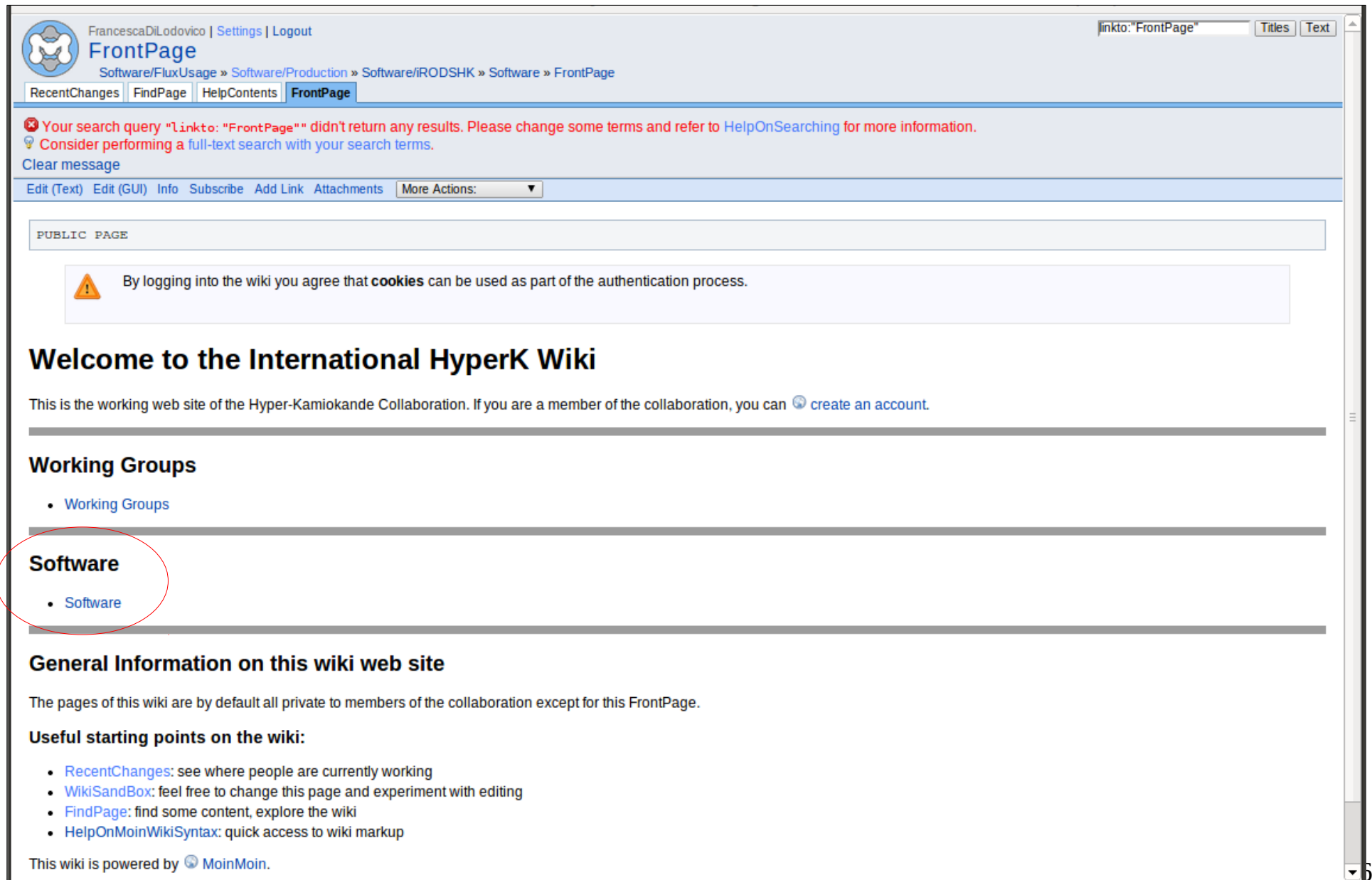
Currently contains all the needed software documentation:

- how to install a release

- how to retrieve the production files

If you are a T2K member and use the same email address as for t2k.org you will be automatically registered, otherwise I will get an email requesting to approve the account.

For questions on the registration, email Alex Owen (r.a.owen@qmul.ac.uk) and myself (f.di.lodovico@qmul.ac.uk)

International wiki https://wiki.hyperk.org

Computing Model

Working on two major timelines:

Short Term future:
- Need to have a system automated and up and running for the current simulation production: use the Grid (VO: hyperk.org) - ongoing

Middle/Longer Term Future:
- Keep up to date with the current developments towards a Computing Model for Hyper-Kamiokande
- Current proposal presented @ the Second Open HK Meeting:
  http://indico.ipmu.jp/indico/contributionDisplay.py?contribId=5&confId=10
- Main aspects:
  - Cloud
  - Virtualization
  - Digital Preservation

Summary

- Developed a fully comprehensive system for

    - Release code management

    - Processing on the Grid

    - Data Storage

    - Documentation

- The system is complete, but being being extended.

- First MC simulated events are being production. Full beam production expected for the July collaboration meeting.

Additional Slides

Distributed Processing

- Generally want to keep jobs within few hours
  - Limit exposure to infrastructure 'glitches' (network problems, disk problems etc)
- For very long jobs ideal to have checkpointing
  - If job fails 90% into processing don't lose everything
- For reco is it possible to have a number of detailed passes:
  - Find 'obvious' rings output results
  - Read in results in next pass and look for more combinations output results
  - iterate

Data Replication

- Currently setting up iRODS system at KEKCC for HK

  - But, total available space will be 1TB until HK approved expt at KEK

  - Can cache most recent data there for people at KEK to use and have 'older' data only available at QMUL

  - Can setup another iRODS system and replicate data to there as well to provide larger pool of storage.

How to get the Data

- There are currently 90K simulation events in iRODS. The reconstruction is processing and will match.

- Make sure you have the iRODS icommands in your path.

- Make sure you have an .irodsEnv file for the HYPERK_Reader account

- Log onto the QMUL iRODS system:

    - iinit

    - Ask me for the account details

- To dowload simulation data:

    - iget /QMULZone2/home/hyperk/production/simu/V00-00-01/neutvect-<run>-<sim>.root

- To download simulation data:

    - iget /QMULZone2/home/hyperk/production/reco/V00-00-01/fitqun-<sim>-<rec>.root

Short-term needs: MC Production

Job Manager/Bookkeeping

Grid

Store

Cloud (future)

Inst 1 Inst 2

... Inst N

Currently using GridPPP in the UK. Will be expanded to other Grids.

Files can be replicated to different sites

Starting to test Cloud systems (KEKCC), QMUL.
The will not be used for now.

usr1 usr2

usr3 usrN

Users – can retrieve files without need of Grid certificate,

23