

micala: CERN web lecture archiving system

presentation to CERN IT-UDS group

29 November 2010

Jeremy Herr

"Web Lecture"

- Synchronized audio/video and slides
- Viewable in any web browser

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mediaarchive.cern.ch/MediaArchive/Video/Public2/M... Google

Most Visited Latest Headlines WLAP - Browse Resour... Twitter / herrj Stumble! All I like it

http://mediaa...2/flash.html

CERN Web Lecture Viewer (c) 2010

ATLAS Detector

- *Designed to explore a variety of physics from p-p collisions*
- *More collisions occur than can be recorded*
 - 1,000,000,000/s -> 200/s
 - 3 level trigger system
 - Complex subsystems and triggers require constant monitoring

ATLAS Level 1 Muon Central Trigger Processor Interface Monitoring
Olson, Timothy
12 August 2010 10:50:00
duration: 00:15

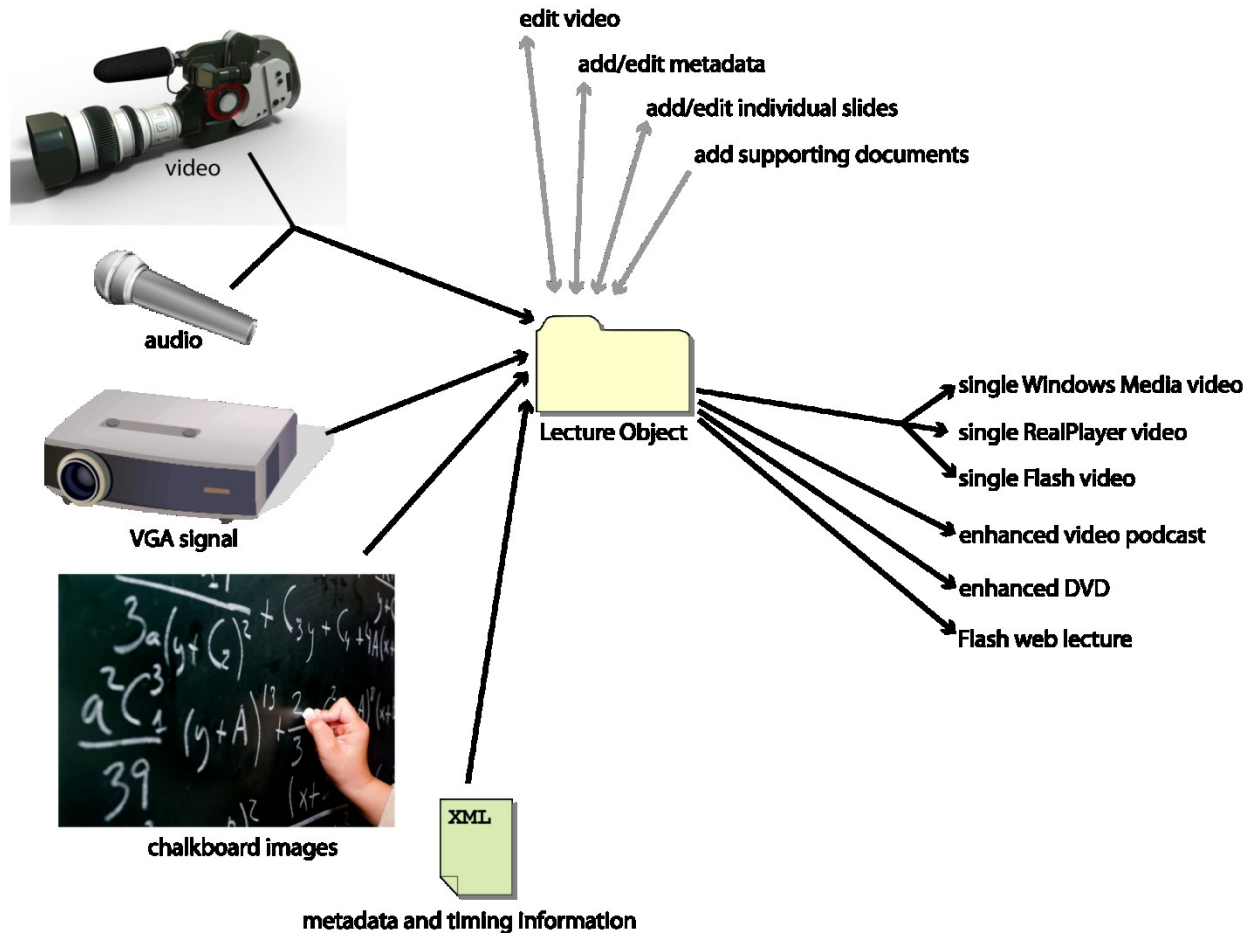
ATLAS EXPERIMENT CERN

00:01:50 / 00:15 Slide 3 / 28

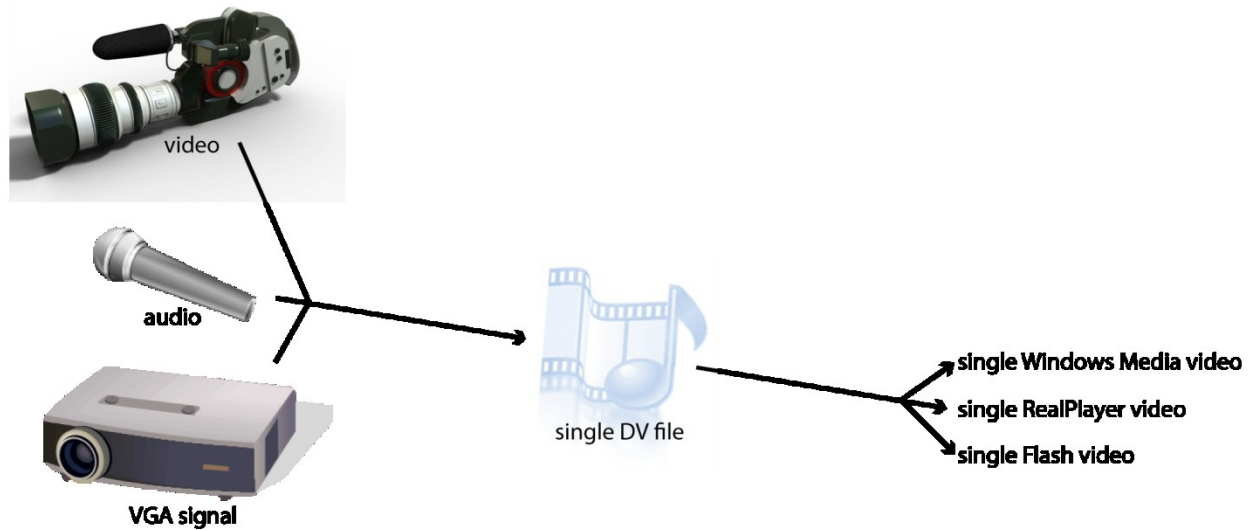
x Find: Next Previous Highlight all Match case

Done

Lecture Object



Single file



U-M ACP background

- The Web Lecture Archive Project (collaboration with U-M and CERN) recorded the Summer Student Lectures in 1999.
- The U-M ATLAS Collaboratory Project recorded ATLAS Overview weeks, software and physics tutorials since 2003.
- MScribe pilot project recorded 8 entire University courses in 2006-07.
- U-M CARMA service founded in 2008, a campus-wide recording service, using hardware/software developed in MScribe.
- Around 2500 lectures recorded and stored in U-M archive.

U-M - CERN partnership

- Technology transfer agreement between U-M and CERN Sep 2008 - Nov 2010.
- Goal: Put web lecture archiving into production at CERN.
 1. Market survey of lecture archiving technologies.
 - Compared six systems in detail
 2. Choose a system or combination of systems and extend it.
 - U-M system was chosen
 3. Integration with CERN infrastructure.

micala - an open source project

- Michigan - CERN automated lecture archiving system
 - vo.2: state of the U-M software in 2008
 - vo.3: new version developed at CERN
- Components
 - recording
 - processing
 - monitoring
 - database, web interface
 - viewing formats
 - Flash, podcast
 - analytics
 - web log analysis
 - integrated Flash player usage reporting
- Integrates with existing content management system (Indico or CWIS)

Existing CERN infrastructure

- Indico
 - Conference management software
 - User-defined metadata describing most events at CERN
- CDS - CERN Document Server
 - Metadata and document server
 - Public interface and search for digital records and media
- Media Archive
 - DFS storage for multimedia files
 - Transcoder server farm that prepares media for web viewing
- Webcast and videoconferencing hardware
 - Tandberg cameras already installed in many rooms at CERN
 - Webcast hardware/software already installed in many rooms

CERN infrastructure integration

- Indico
 - A new plugin was developed: the "Recording Manager."
 - Event metadata and access control now exported to both micala and CDS.
- CDS
 - Updated to accept lecture object records from the Media Archive.
 - Updated to accept access control information from Indico.
 - Updated to display web lectures.
- Media Archive
 - Installed micala publishing scripts on MediaArchive servers.
 - Now accepts lecture objects and produces web lectures, podcasts.
- Webcast and videoconferencing hardware
 - Videoconferencing cameras re-used for recording audio/video.
 - Webcast PCs updated to also record camera and slide feeds.



MediaArchive

**videoconference
camera**



webcast PC

CERN Document Server



MediaArchive

CERN Document Server

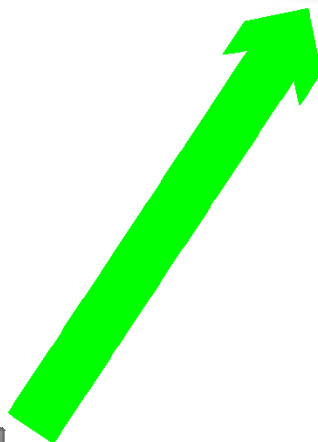
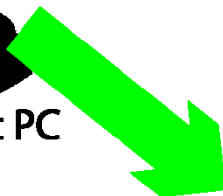
videoconference
camera

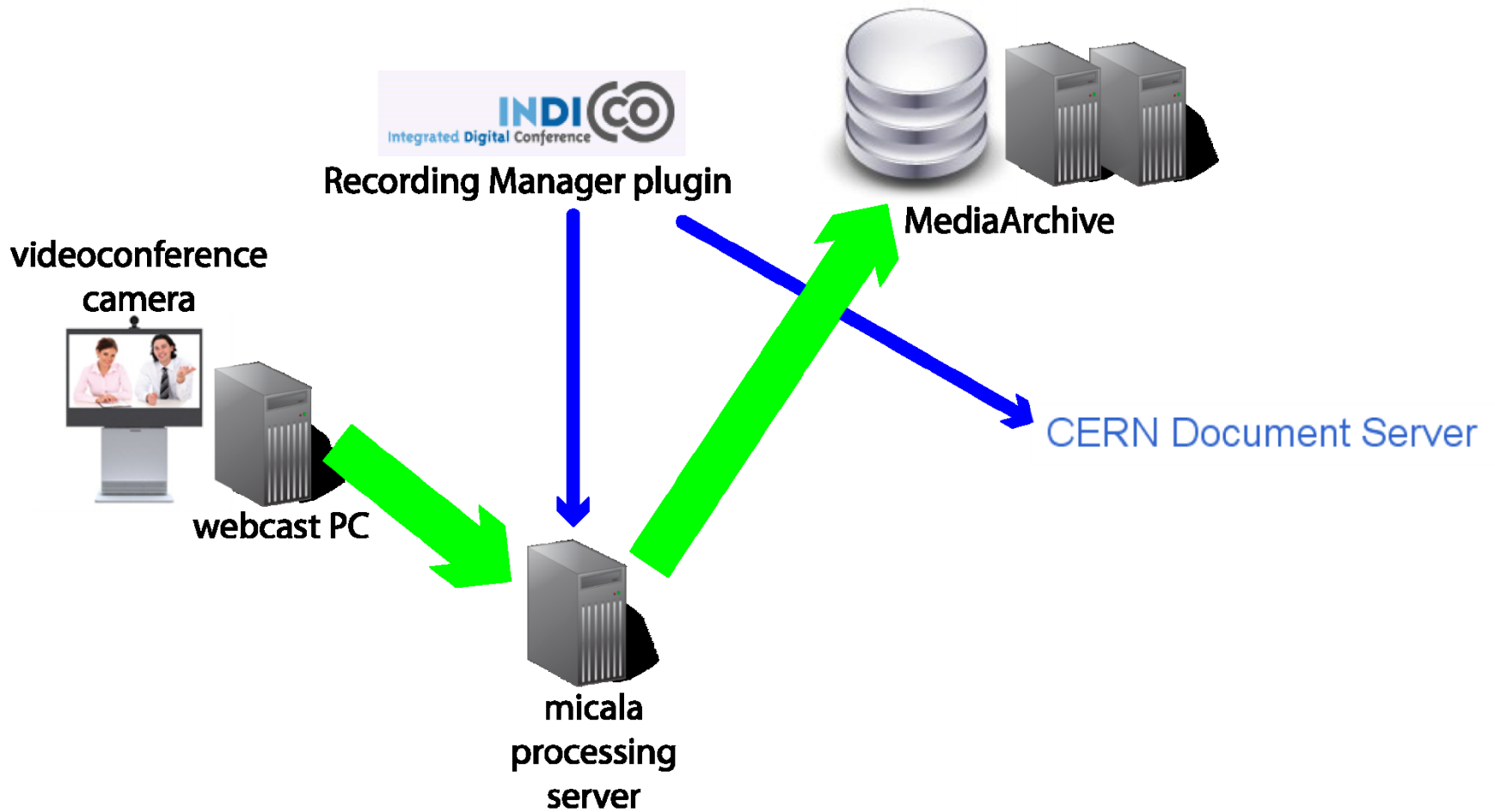


webcast PC



micala
processing
server







Recording Manager plugin



MediaArchive

videoconference camera



webcast PC



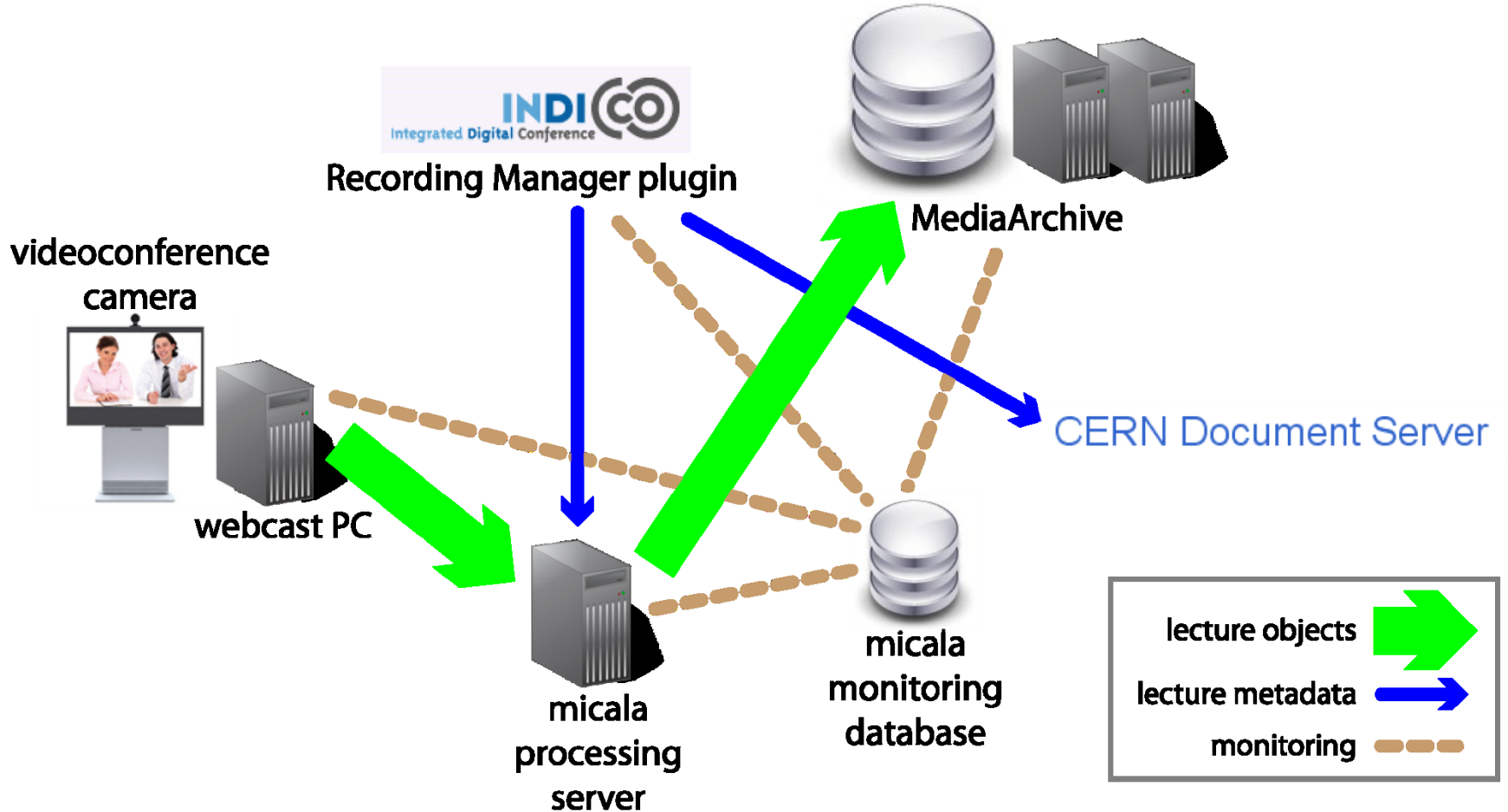
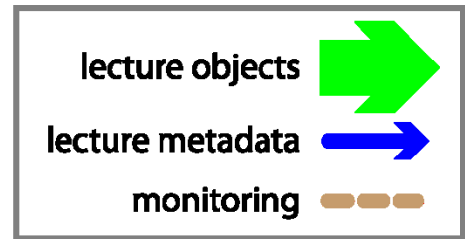
micala processing server



micala monitoring database



CERN Document Server



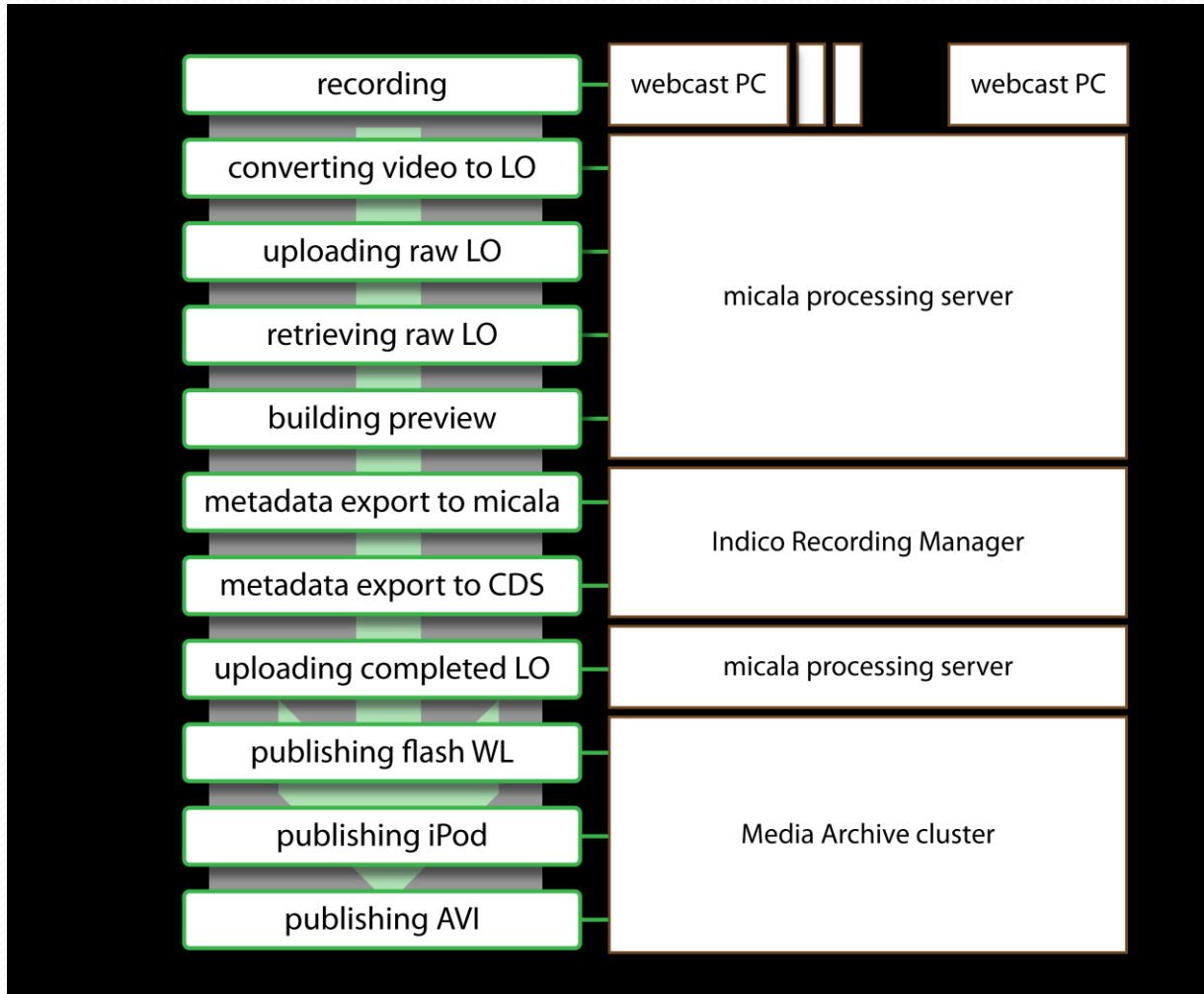
monitoring - database

- MySQL database, which ties together :
 - micala processing
 - Indico
 - CDS
 - Media Archive
- It tracks:
 - every lecture recorded
 - every machine that records or processes lectures
 - every venue where recordings take place
- When any machine does any task to any lecture, it reports the details of its status to the database; these status updates are kept indefinitely.

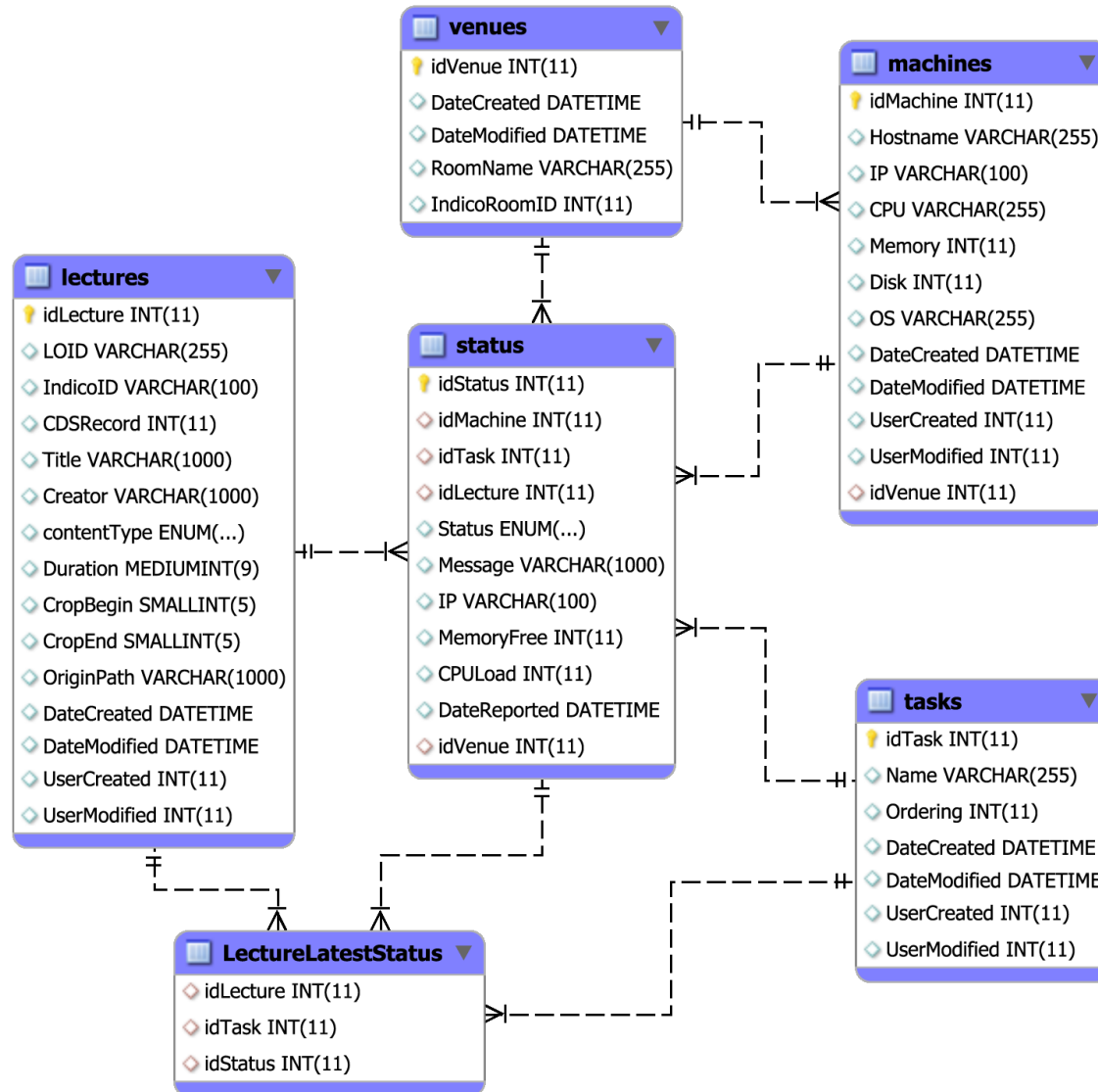
monitoring - database

- The database is backed up nightly to DFS, kept for 2 months.
- Running since 11 June 2010 (almost 6 months)
- The compressed nightly backup file is currently 192 KB
 - 300 Lectures (265 Lecture Objects, 35 plain video recordings)
 - 22,377 status updates
 - 19 machines
 - 13 tasks
 - 12 venues

monitoring - database tasks



monit datab



monitoring - web interface

Monitoring overview


FILTERS	Age:	Lecture type:			Orphan status:			Processing:			Lectures found: 10
	<input type="text" value="1week"/>	<input type="button" value="Filter"/>	<input type="button" value="Any"/>	<input type="button" value="Web Lecture"/>	<input type="button" value="Plain Video"/>	<input type="button" value="Any"/>	<input type="button" value="Orphan"/>	<input type="button" value="Non-orphan"/>	<input type="button" value="Any"/>	<input type="button" value="Finished"/>	

Lecture Object ID	Indico ID	Duration	Recording room	Date Created	Tasks Status
20100817-cernpcexternalwcd-160000	orphan	1h30m		2010-11-26 14:33:21	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
N/A	111503	TBD		2010-11-25 17:53:47	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
20101125-cernpcwebc18-162654	orphan	1h07m	500-1-001 (Main Auditorium)	2010-11-25 16:26:53	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
20101125-cernpcwebc09-155430	114573	3h26m	222-R-001 (Filtration Plant)	2010-11-25 15:54:30	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
20101125-cernpcwebc09-155250	orphan	1 minutes	222-R-001 (Filtration Plant)	2010-11-25 15:52:49	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
20101125-cernpcwebc09-152530	orphan	21 minutes	222-R-001 (Filtration Plant)	2010-11-25 15:25:30	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
20101124-cernpcwebc18-155045	99232	2h19m	500-1-001 (Main Auditorium)	2010-11-24 15:50:44	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>

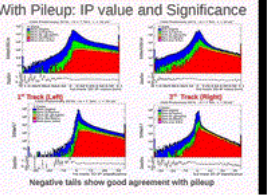
Title	CMS General Weekly Meeting GWM43
Creator	Guido Tonelli
Recording room	500-1-001 (Main Auditorium)
Recording machine	pcwebc18.cern.ch
Origin path	\\pcwebc18.cern.ch\Archive\20101124-cernpcwebc18-155045
Indico ID	99232
	View on Indico
	Recording Manager
CDS Record	1309246
	CERN Document Server

<input checked="" type="checkbox"/>	recording	DONE	1h14m
<input type="checkbox"/>	flattening f4v files		
<input checked="" type="checkbox"/>	converting video to LO	DONE	13 minutes
<input checked="" type="checkbox"/>	uploading raw LO	DONE	5 seconds
<input checked="" type="checkbox"/>	retrieving raw LO	DONE	8 minutes
<input checked="" type="checkbox"/>	building preview	DONE	1 minutes
<input checked="" type="checkbox"/>	metadata export to micala	DONE	0 seconds
<input checked="" type="checkbox"/>	metadata export to CDS	DONE	15h54m
<input checked="" type="checkbox"/>	uploading completed LO	DONE	35 seconds
<input type="checkbox"/>	publishing flash WL	REQUESTED	
<input type="checkbox"/>	publishing podcast MP4		
<input type="checkbox"/>	publishing podcast iPod		
<input type="checkbox"/>	Indico link to CDS		

Previews



With Pileup: IP value and Significance



Negative tails show good agreement with pileup

20101124-cernpcwebc18-155033	orphan	TBD	500-1-001 (Main Auditorium)	2010-11-24 15:50:33	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
20101123-cernpcwebc41-135022	72840	6h46m	40-S2-C01 (Salle Curie)	2010-11-23 13:50:22	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
20101119-cernpcuds28-093000	105451	1h07m	31-3-004 (IT Amphitheatre)	2010-11-22 10:41:00	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>

monitoring - web interface

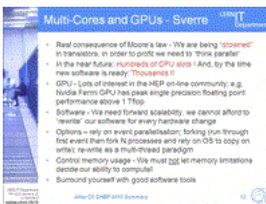
micala

HOME PAGE MONITORING ADMIN HOME PAGE MACHINES TASKS VENUES DAEMON

Report on the Taipei CHEP 2010 Conference Date: 2010-11-22 10:41:00 - Creator: Alan Silverman
Lecture Object ID #20101119-cernpcuds28-093000 - Lecture #601

Lecture details

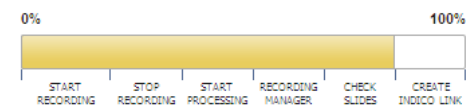
Lecture ID	601
Lecture Object ID	20101119-cernpcuds28-093000
Indico ID	105451 View on Indico Recording Manager
CDS Record	1308482 CERN Document Server
Recording room	31-3-004 (IT Amphitheatre)
Recording machine	pcuds28.cern.ch
Duration	1h07m
Crop Begin	00:10:09
Crop End	01:17:20
Origin Files	\\pcuds28\Archive\20101119-cernpcuds28-093000
Copy URL	



Tasks Status

Task name	Status	Age	Duration
recording	DONE		30 seconds
flattening f4v files			
converting video to LO	DONE		17 minutes
uploading raw LO	DONE		6 seconds
retrieving raw LO	DONE		8 minutes
building preview	DONE		38 seconds
metadata export to micala	DONE		1 seconds
metadata export to CDS	DONE		4h33m
uploading completed LO	DONE		39 seconds
publishing flash WL	DONE		8 minutes
publishing podcast MP4			
publishing podcast iPod			
Indico link to CDS			

Check-list for operators



- ### Human Tasks
- START Recording
 - STOP Recording
 - START processing [Optional CROP]
 - Recording Manager
 - Check slides and upload
 - Insert Indico link to CDS

CROP Tools

Crop START 00:10:09

Crop END 01:17:20

Processing

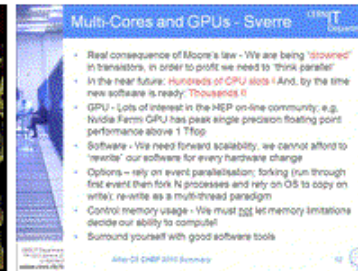
Convert raw f4v files, analyze slides, create Lecture Object

monitoring - web interface

- lecture metadata
- ID numbers
- links to related resources
- animated GIF previews

Lecture details

Lecture ID	601
Lecture Object ID	20101119-cernpcuds28-093000
Indico ID	105451 View on Indico Recording Manager
CDS Record	1308482 CERN Document Server
Recording room	31-3-004 (IT Ampitheatre)
Recording machine	pcuds28.cern.ch
Duration	1h07m
Crop Begin	00:10:09
Crop End	01:17:20
Origin Files	\\pcuds28\Archive\20101119-
Copy URL	cernpcuds28-093000



Multi-Cores and GPUs - Sverre

- Real consequence of Moore's law - We are being "disciplined" in transistors, in order to profit we need to "think parallel"
- In the near future: Hundreds of CPU slots + And, by the time new software is ready: Thousands of
- GPU - Lots of interest in the HEP on-line community, e.g. Nvidia Fermi GPU has peak single precision floating point performance above 1 Tflop
- Software - You need forward scalability, we cannot afford to "rewrite" our software for every hardware change
- Options - rely on event parallelisation, forking (run through first event then fork N processes and rely on OS to copy on write), re-write as a multi-thread paradigm
- Control memory usage - You must pgt let memory limitations decide our ability to compute!
- Surround yourself with good software tools

After DE DNP 2010 Summary

monitoring - web interface

- status of each task
- how long it took
- error messages

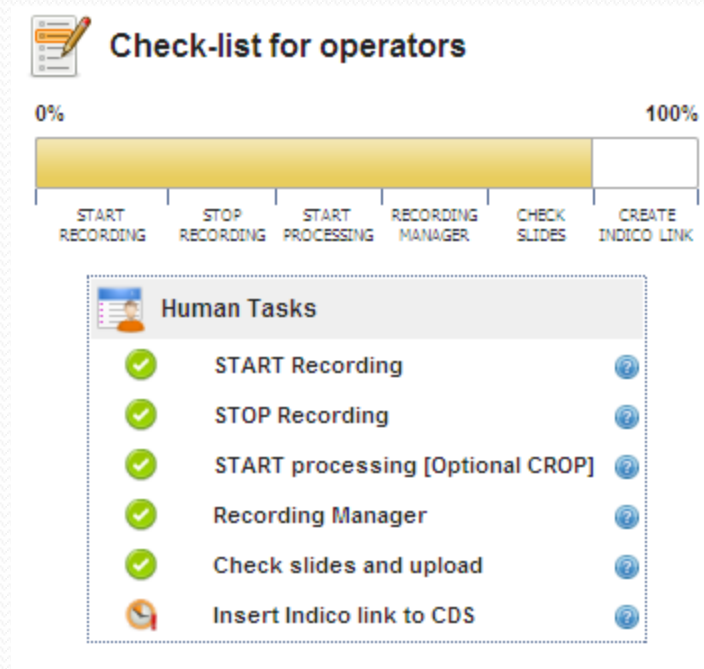


Tasks Status

Task name	Status	Age	Duration
recording	DONE		30 seconds
flattening f4v files			
converting video to LO	DONE		17 minutes
uploading raw LO	DONE		6 seconds
retrieving raw LO	DONE		8 minutes
building preview	DONE		38 seconds
metadata export to micala	DONE		1 seconds
metadata export to CDS	DONE		4h33m
uploading completed LO	DONE		39 seconds
publishing flash WL	DONE		8 minutes
publishing podcast MP4			
publishing podcast iPod			
Indico link to CDS			

monitoring - web interface

- A check-list for human operators shows what should be done next.



monitoring - web interface

- Operator can specify crop times and start processing
- Processing must be started manually because often test recordings are made and deleted.

CROP Tools

Crop START

Crop END

Processing

Convert raw f4v files, analyze slides, create Lecture Object

monitoring - web interface

- Log shows everything ever done to a particular lecture
 - Task name
 - General status (START, RUNNING, COMPLETE, ERROR, REQUESTED)
 - Task details (e.g. "FLV conversion: 45%")
 - Machine name performing task
 - Date/time



Log

Task Name	Status	Task Details	Machine Name	Date/Time
publishing flash WL	RUNNING	untarring master tarball	mediaarchive.cern.ch	2010-11-22 17:37:05
publishing flash WL	RUNNING	master tarball untarred	mediaarchive.cern.ch	2010-11-22 17:38:49
publishing flash WL	RUNNING	copying files	mediaarchive.cern.ch	2010-11-22 17:38:49
publishing flash WL	RUNNING	creating FLV file	mediaarchive.cern.ch	2010-11-22 17:38:49
publishing flash WL	RUNNING	FLV conversion: 15%	mediaarchive.cern.ch	2010-11-22 17:39:41
publishing flash WL	RUNNING	FLV conversion: 30%	mediaarchive.cern.ch	2010-11-22 17:40:31
publishing flash WL	RUNNING	FLV conversion: 45%	mediaarchive.cern.ch	2010-11-22 17:41:21
publishing flash WL	RUNNING	FLV conversion: 61%	mediaarchive.cern.ch	2010-11-22 17:42:11
publishing flash WL	RUNNING	FLV conversion: 76%	mediaarchive.cern.ch	2010-11-22 17:43:01
publishing flash WL	RUNNING	FLV conversion: 91%	mediaarchive.cern.ch	2010-11-22 17:43:51
publishing flash WL	RUNNING	FLV file finished	mediaarchive.cern.ch	2010-11-22 17:44:18
publishing flash WL	RUNNING	copying FLV file	mediaarchive.cern.ch	2010-11-22 17:44:19
publishing flash WL	RUNNING	resizing slides	mediaarchive.cern.ch	2010-11-22 17:44:25
publishing flash WL	RUNNING	creating thumbnails	mediaarchive.cern.ch	2010-11-22 17:44:53
publishing flash WL	RUNNING	installing Flash web lecture	mediaarchive.cern.ch	2010-11-22 17:45:08
publishing flash WL	COMPLETE		mediaarchive.cern.ch	2010-11-22 17:45:09
metadata export to CDS	COMPLETE	CDS record: 1308482	indico.cern.ch	2010-11-22 20:05:33


[Go to last log](#)


monitoring - web interface













- Venue management interface
- ## micala


HOME PAGE MONITORING ADMIN HOME PAGE MACHINES TASKS VENUES DAEMON

Venues

 **Venues List**

 [New venue](#)

Room name	Indico ID	
222-R-001 (Filtration Plant)	184	
31-3-004 (IT Ampitheatre)	33	
354 (CMS Centre)	0	
40-S2-A01 (Salle Andersson)	51	
40-S2-C01 (Salle Curie)	53	
500-1-001 (Main Auditorium)	57	
503-1-001 (Council Chamber)	0	
503-2-007	0	
513-R-069 (Computer Centre)	0	
external	0	
mobile	0	
SCX1 (ATLAS Control Room)	0	

 **Venue details**

Room name

Indico RoomID

To get the Indico Room ID, [goto to Indico - Search rooms](#), search the desired room, click on 'Room details' and get the room ID by the URL parameter 'roomId'

monitoring - web interface

- machine management interface

micala

The screenshot shows the 'micala' web interface. At the top, there is a navigation bar with buttons for 'HOME PAGE', 'MONITORING', 'ADMIN HOME PAGE', 'MACHINES', 'TASKS', 'VENUES', and 'DAEMON'. Below this is a yellow header for 'Machines'. The main content is split into two panels: 'Machines List' and 'Machine details'.

Machines List

[New machine](#)

Hostname	IP	
cernwlcd0001.cern.ch		
cernwlcd0002.cern.ch		
externalwlcd.cern.ch		
indico.cern.ch	137.138.38.22	
lectureprocessing01.cern.ch	137.138.4.159	
mediaarchive.cern.ch		
pcituds15.cern.ch		
pcuds28.cern.ch		
pcuds40.cern.ch		
pcuds41.cern.ch		
pcuds42.cern.ch		
pcuds46.cern.ch		
pcwebc09.cern.ch		
pcwebc12.cern.ch		
pcwebc13.cern.ch		
pcwebc14.cern.ch		
pcwebc18.cern.ch	137.138.47.16	
pcwebc41.cern.ch		
umwlcd0010.cern.ch	137.138.38.41	

Machine details

Hostname:

IP:

CPU:

Memory:

Disk:

OS:

Venue:

monitoring - web interface

- Tasks can be added, renamed, arranged in different order.

micala

HOME PAGE MONITORING ADMIN HOME PAGE MACHINES TASKS VENUES DAEMON

Tasks

Task Ordering

[Click here to enable/disable sorting list](#)
(drag and drop tasks up/down to sort)

recording	
flattening f4v files	
converting video to LO	
uploading raw LO	
retrieving raw LO	
building preview	
metadata export to micala	
metadata export to CDS	
uploading completed LO	
publishing flash WL	
publishing podcast MP4	
publishing podcast iPod	
Indico link to CDS	

New Task

Task name insert after --- As first element ---

Caution

Pay attention when you insert a new task or edit a task name.
Do it only if you are aware of the consequences!

If you want insert a new task, read this:

- Video processing daemon must be stopped
- After inserted the new task, you have to wait few minutes that some data of the database will be re-generated
- When the operation is complete, you can start again the video processing daemon

monitoring - web interface

- Operator can start/stop daemon through web interface.
- This ensures that the three threads complete tasks safely.

micala

The screenshot displays the 'micala' web interface. At the top, there is a navigation bar with tabs: HOME PAGE, MONITORING, ADMIN HOME PAGE, MACHINES, TASKS, VENUES, and DAEMON. Below this is a yellow header for the 'Daemon manager' section. The main content area is divided into three columns:

- Daemon Info:** Contains an information icon and a list of actions: 'Convert Webcasts to RawLO', 'Convert RawLO to LO, create previews and move LO to Lecture Objects folder', 'Wait for human checks', and 'Move the LO to Media Archive'. Below the list, it states 'Another parallel thread match the LO with Indico'.
- Daemon Console:** Shows the current status as 'Stopped' in red text. Below this, it indicates 'Last Alive Feedback: 2010-11-26 13:50:12'. There are two buttons: a green 'Start daemon' button and a red 'Stop daemon' button.
- Daemon logs:** A scrollable log area with a 'Refresh status' button. The log entries include: 'Sending stop signal to install metadata parallel Thread.', 'micala daemon stopped.', 'Daemon is starting up.', 'micala daemon started to work.', 'Requested to stop daemon.', 'User stopped daemon.', 'Sending stop signal to flatten parallel Thread.', 'Sending stop signal to install metadata parallel Thread.', 'micala daemon stopped.', 'Daemon is starting up.', 'micala daemon started to work.', 'Requested to stop daemon.', 'User stopped daemon.', 'Sending stop signal to flatten parallel Thread.', 'Sending stop signal to install metadata parallel Thread.', 'micala daemon stopped.', and 'micala daemon stopped.'. Each entry is followed by a timestamp, such as '2010-11-25 19:04:26'.

recording

- Linux (Ubuntu 9.10 tested)
 - Used at U-M, tested on laptop- and PC-based systems.
 - Mobile kit includes prosumer camera, mic mixer, lapel mic and receiver.
 - Video capture: Firewire (laptop) or Osprey 100 card (PC).
 - Slide capture: Epiphan VGA2USB device, 1fps JPEG images.
 - Produces a "raw" lecture object: master.avi and JPEGs.
- Windows (XP tested)
 - Used at CERN in 5 auditoria.
 - Mobile unit planned.
 - Video capture: Osprey 240e card (PC only).
 - Slide capture: Epiphan VGA2USB device, 4fps f4v video.
 - Produces two f4v video files.

processing - micala server

- micala daemon operates three concurrent threads:
 - Major processing tasks - high CPU and I/O usage, run sequentially.
 - conversion of captured f4v files to a "raw" lecture object, video cropping
 - slide analysis and chapter creation, building standard lecture object
 - upload of completed lecture object to Media Archive **Masters** DFS directory
 - Metadata installation - very small task to be run every 5 seconds.
 - retrieval/installation of Indico metadata into lecture object.
 - f4v flattening - to be run as soon as each recording completes.
 - "raw" f4v files created by Adobe FMLE must be "flattened" before usable

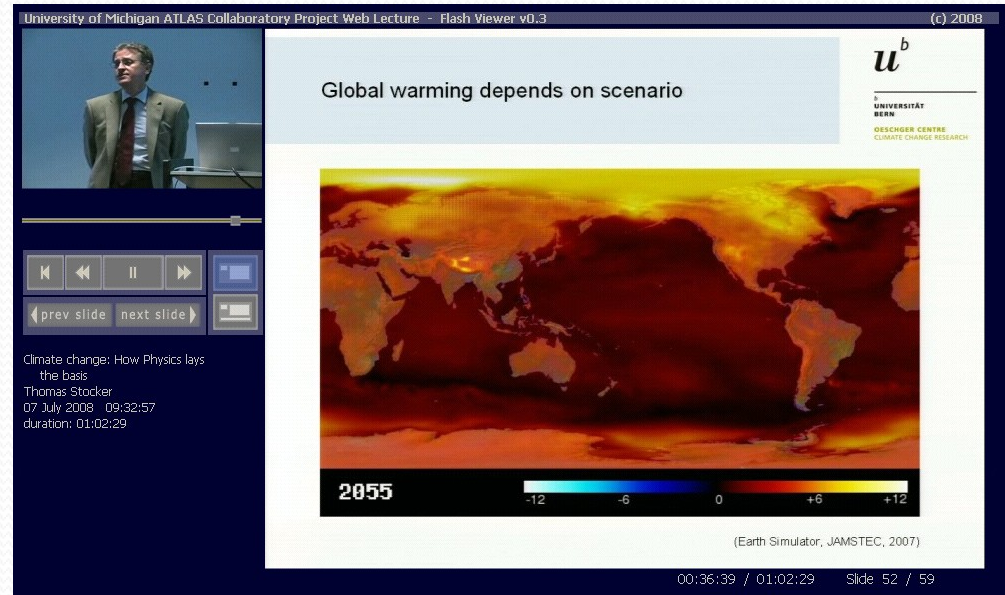
processing - publish

- Convert lecture object ("**master**") to viewable format ("**slave**")
- These could be run by the daemon, but at CERN run by Media Archive.
- Media Archive server farm monitors Masters directory.
- Three formats created for every new/changed lecture object:
 - Flash web lecture - side-by-side synchronized video and slides
 - Podcast MP4 - single video file switches between video and slides
 - Podcast iPod - same as MP4, suitable for playback on iPod/iPhone

viewing formats

- Web Lecture

- plays in any browser with Flash plugin.
- video of speaker
- synchronized slide images
- navigate directly to chapters



The screenshot shows a web lecture player interface. On the left, a video window displays a man in a suit speaking. Below the video are playback controls (play, stop, previous, next) and navigation buttons for 'prev slide' and 'next slide'. Text below the controls reads: 'Climate change: How Physics lays the basis', 'Thomas Stocker', '07 July 2008 09:32:57', and 'duration: 01:02:29'. The main slide area on the right has the title 'Global warming depends on scenario' and the University of Bern logo. It features a world map with a color scale from -12 to +12, labeled '2055'. A color bar below the map shows a gradient from blue (-12) to red (+12). At the bottom right of the slide, it says '(Earth Simulator, JAMSTEC, 2007)'. The player's status bar at the bottom shows '00:36:39 / 01:02:29' and 'Slide 52 / 59'.

- Podcast (full-resolution MP4)

- Single video file, switches between video of speaker and slides

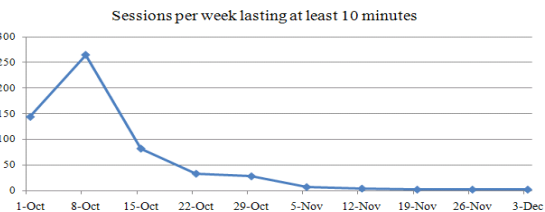
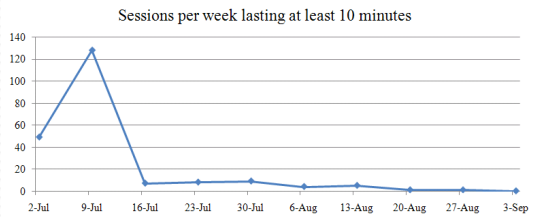
- Podcast (iPod/iPhone)

- Formatted for playback on iPod/iPhones (using HandBrake)

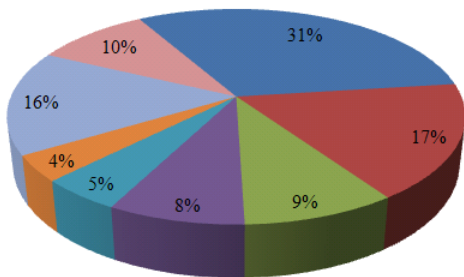
analytics

- Perl script that parses existing web logs to find:
 - how long users stay on site.
 - how long users view each lecture.
 - how popular each lecture is.
- Flash viewer logging (under construction):
 - each Flash viewer reports back to micala server.
 - can see which Flash features are most useful.
 - which slides are viewed the most.
 - understand how users view multimedia content.

analytics

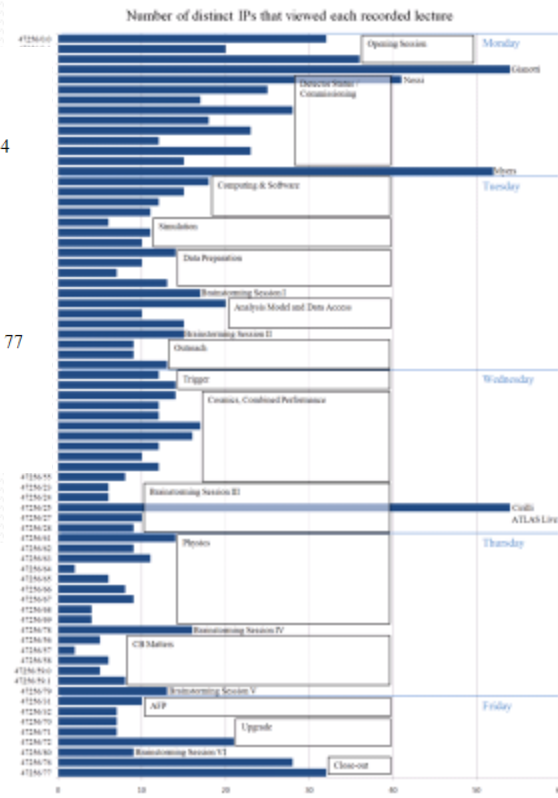
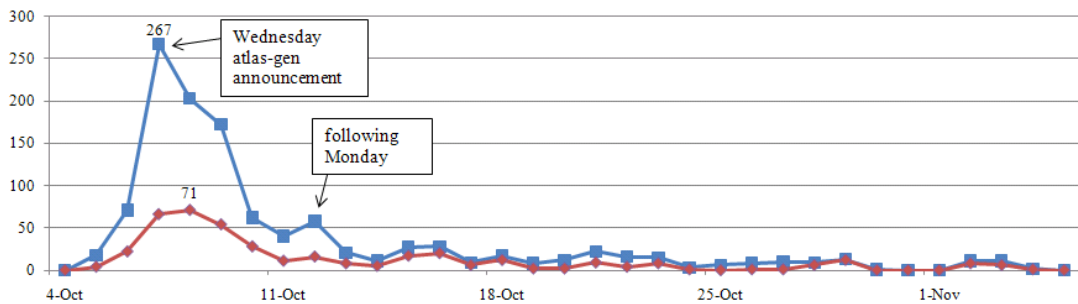


491 unique IP addresses, categorized by country
5 October - 5 December 2009



- Switzerland: 154
- USA: 85
- Germany: 45
- France: 40
- Italy: 24
- UK: 19
- other countries: 77
- unknown: 47

ATLAS Week October 2009: daily sessions of any length,
and daily sessions 10 minutes or longer



Content Management System

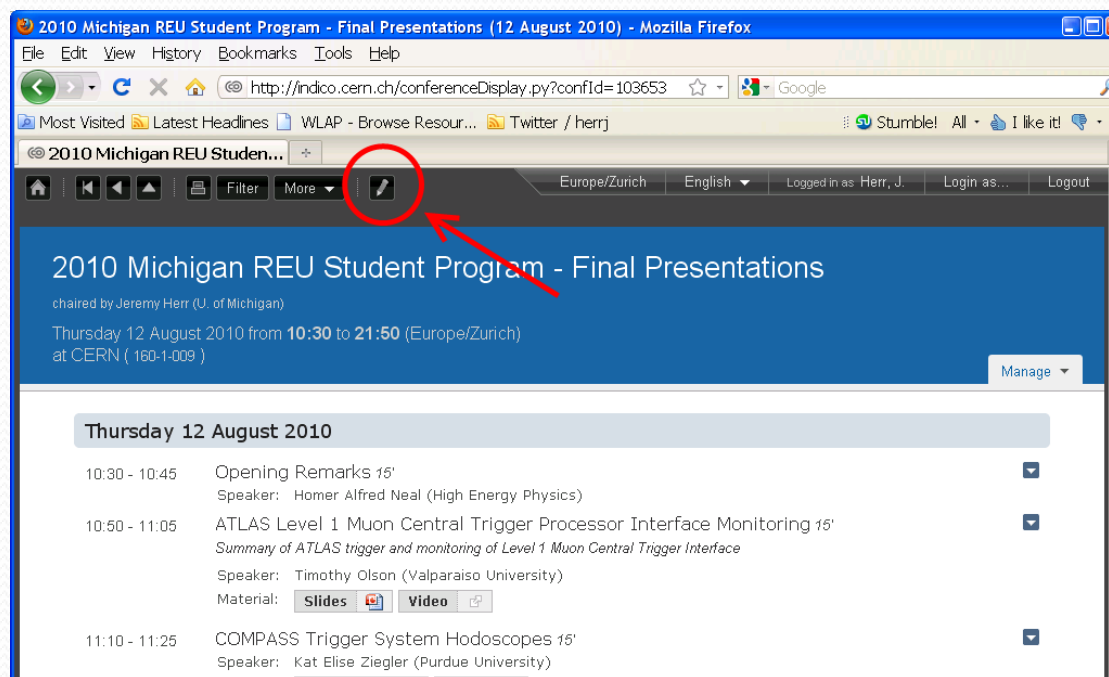
- micala does not include a CMS for browsing through lectures.
 - At U-M, we use CWIS, modified to export metadata to micala.
 - At CERN, we use Indico and CDS
- The "Recording Request" is a new Indico plugin
 - Form for users to request recordings for a given event
 - Management interface to accept/deny requests
- The "Recording Manager" is a new Indico plugin
 - Matches recorded lectures with Indico records:
 - events
 - sessions
 - contributions
 - subcontributions
 - Submits metadata, access control list and collection information to CDS.
 - Submits metadata to micala.

Indico - Recording Request

- The following rooms are currently outfitted with the new Windows recording system:
 - 500-1-001 (Main Auditorium)
 - 40-S2-A01 (Salle Andersson)
 - 40-S2-C01 (Salle Curie)
 - 222-R-001 (Filtration Plant)
 - 31-3-004 (IT Ampitheatre)
- All rooms that have a videoconferencing system can be upgraded for recording.
- Users can request to have any event in these rooms recorded.
- The recording is operated remotely by AVC team.

Indico - Recording Request

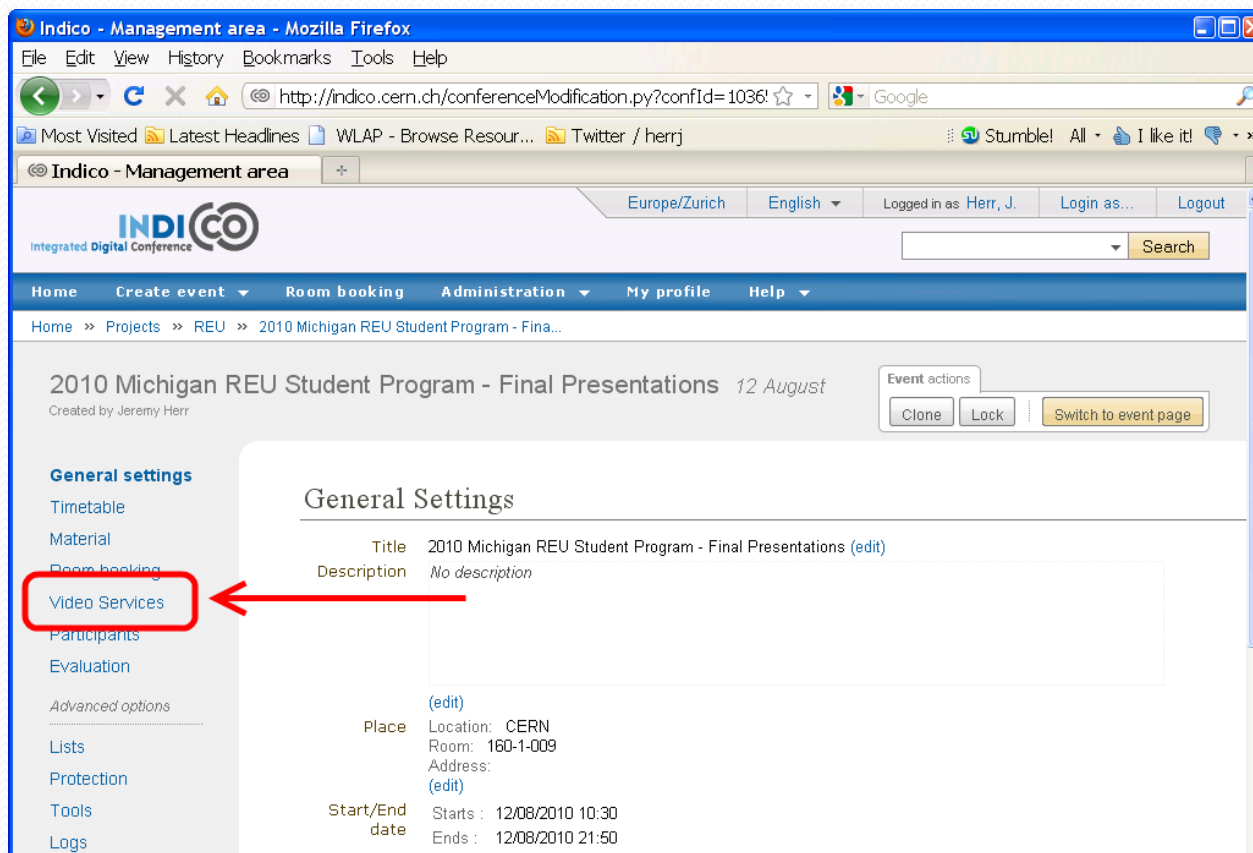
- Book a room and create an event in Indico.
 - Remember to include dates, times, locations, people!



- Click on the pencil (management area).

Indico - Recording Request

- In management area, click on "Video Services."



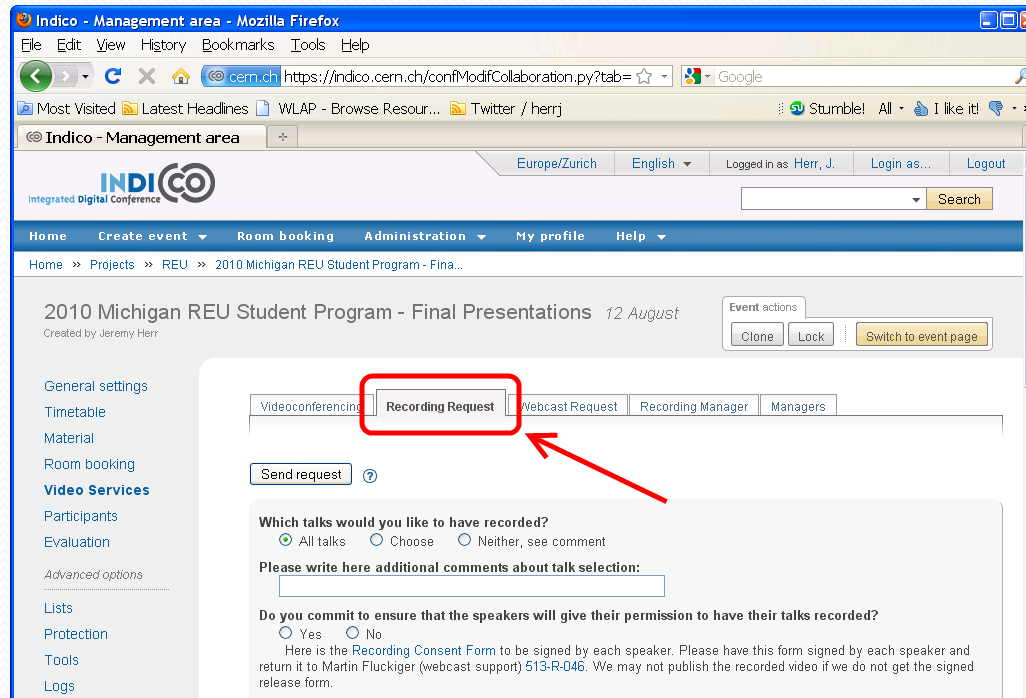
The screenshot shows the Indico Management area in Mozilla Firefox. The browser address bar displays the URL: <http://indico.cern.ch/conferenceModification.py?confId=1036>. The page title is "Indico - Management area". The user is logged in as "Herr, J.". The main content area shows the "General Settings" for the event "2010 Michigan REU Student Program - Final Presentations" (12 August). The "Video Services" link in the left sidebar is highlighted with a red box and a red arrow pointing to it.

General Settings

Title	2010 Michigan REU Student Program - Final Presentations (edit)
Description	No description
Place	(edit) Location: CERN Room: 160-1-009 Address: (edit)
Start/End date	Starts : 12/08/2010 10:30 Ends : 12/08/2010 21:50

Indico - Recording Request

- User clicks on the "Recording Request" tab.



- User fills out form and submits, and AVC is notified.

Indico - Recording Manager

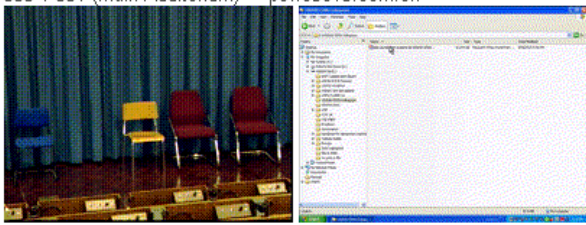
1. Select a talk:

E	2010 Michigan REU Student Presentations ...	14:00, Thu 22 Jul
C	Opening Remarks	14:00, Thu 22 Jul <i>Homer Alfred Neal</i>
C	Missing Transverse Energy and Jet Selec...	14:22, Thu 22 Jul <i>Judson Benton Locke</i>
C	Data analysis of lambda-0 polarization....	14:29, Thu 22 Jul <i>Ray Zhang</i>
C	Data Analysis of a Low Momentum Detecto...	14:36, Thu 22 Jul <i>Michael Glidden</i>
C	Stability in VPT signal in the ECAL End...	14:43, Thu 22 Jul <i>Len Evans</i>
C	Gustav Gandara	14:50, Thu 22 Jul <i>Gustavo Gandara Montano</i>
C	Z boson reconstruction with Monte Carl...	14:57, Thu 22 Jul <i>Adam Wallace Sekou Lowery</i>
C	CMS Dashboard System	15:04, Thu 22 Jul <i>Aram Apyan</i>
C	CMS Cathode Strip Chamber Track Finder	15:11, Thu 22 Jul <i>Alex Ji</i>
C	Rebecca Pankow	15:18, Thu 22 Jul <i>Rebecca Pankow</i>
C	Semilanthonic Boosted Top Decays	


2. Select content type: plain video or **web lecture**

3. Select an orphan lecture object:

30.09.2010 16:00:29 54 minutes
500-1-001 (Main Auditorium) pcwebc18.cern.ch



30.09.2010 16:31:01 2 hours 54 minutes
500-1-001 (Main Auditorium) pcwebc18.cern.ch



08.10.2010 11:35:17 0 seconds
503-2-007 pcuds41.cern.ch

4. Select language(s) in which the talk was given

English French Other

5. Create CDS record (and update micala database)

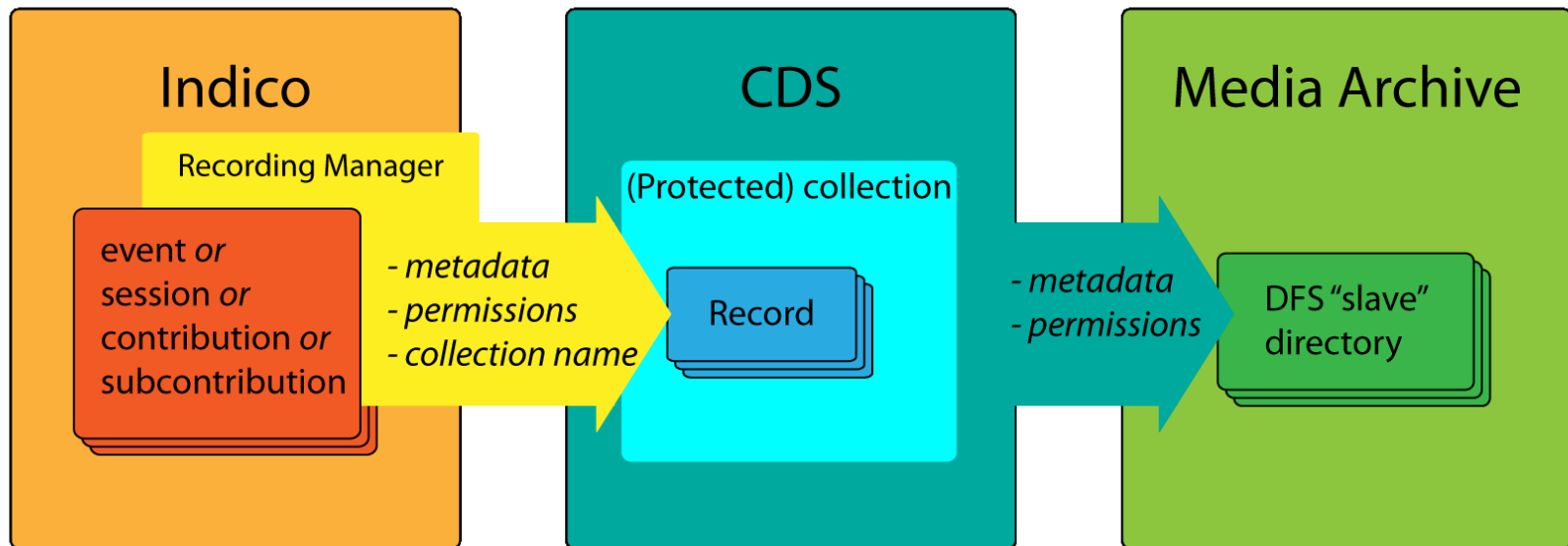
talk: 101928c8, web lecture: 20100930-cernpcwebc18-163101, languages: eng

6. Create Indico link to CDS record

Access control for web lectures

- ACL originates from Indico
 - It is a list of allowed e-groups and individual e-mails
 - Recording Manager recursively checks Indico permissions on:
 - categories
 - event
 - session
 - contribution
 - subcontribution
- Then it exports the ACL with metadata and collection name to CDS.
- The same ACL must then be applied to:
 - CDS records
 - DFS Media Archive directories

Access control for web lectures



Scalability

- Recording:
 - There is one dedicated recording machine per room.
- Processing:
 - There is currently one processing server.
 - micala is designed to allow multiple processing servers.
- Publishing:
 - Media Archive transcoding server farm has 10 machines.

Human operator tasks

- Accept recording request
 - Indico Recording Request plugin
- Log into recording PC, press START/STOP
 - This could be automated, but talks often start late
 - Indico schedules are changed/ignored at last minute
- Match recording to metadata
 - Indico Recording Manager plugin
- Check slides for duplicates
 - VGA2USB hardware bug can create extra chapters
- Indico Recording Manager – create Indico link
 - This can eventually be automated with help from CDS

After more automation...

- Accept recording request
 - Indico Recording Request plugin
- Log into recording PC, press START/STOP
 - This could be automated, but talks often start late
 - Indico schedules are changed/ignored at last minute
- Match recording to metadata
 - Indico Recording Manager plugin
- ~~• Check slides for duplicates~~
 - ~~• VGA2USB hardware bug can create extra chapters~~
- ~~• Indico Recording Manager – create Indico link~~
 - ~~• This can eventually be automated with help from CDS~~

Activity in 2010

- Recordings include:
 - ATLAS Overview Weeks
 - ATLAS Weekly meetings
 - CMS weekly meetings
 - IT Seminars
 - Software tutorials
 - Shift tutorials
 - Technical training
- 324 lectures archived in 2010, expected to increase.

Code and documentation

- Go to **<http://micala.sourceforge.net>** to download micala, and to see screen shots and examples.
- version control: **git**
 - git is a very powerful version control system ideal for distributed developers
 - It is already used by Indico and CDS
- documentation: **Trac** (wiki)
 - Description of micala components
 - Description of individual scripts, database
 - Installation instructions
 - Development practices
- ticket system: **Trac**
 - Trac integrates the wiki, ticket system, and git

Code details

- Languages:
 - Perl, Python, MySQL (currently translating Perl to Python)
 - HTML (Mako)/CSS/Javascript
 - Older code includes some Java, C, PHP
- Requires:
 - ffmpeg, HandBrake, AviSynth
 - ImageMagick, Python Image Library (migrating all to PIL)
 - jQuery, jQuery plugins, Mako template engine
- Statistics:
 - 20,000 lines of code

Copyright/License

- Copyright by
 - The Regents of the University of Michigan
 - CERN
- Educational Community License 2.0
 - Almost identical to Apache 2.0
 - Used by Opencast Matterhorn project
- Planning to also release under GPL v2
 - This makes our code re-usable by other GPL projects

The future

- Nicola Tarocco is micala's new principal developer.
- This week, I am announcing micala to Opencast collaboration, and the University of Michigan community.
- Planned improvements:
 - OCR metadata extraction from slide images.
 - Regular podcasts of prominent CERN lectures.
 - Automated DVD creation.
 - Perl to Python migration.
 - Simplified installation procedure.
 - Smarter, more automated processing.
 - Web interface for analytics reports.
 - Multiple processing servers for faster turnaround.

Thank you!

This was a team effort!

- Thomas Baron - AVC section leader and micala contributor, a constant source of guidance, support and a pleasure to work with.
- The Indico team, especially Jose Benito Gonzalez Lopez, David Martin Clavo and Jose Pedro Ferreira, for their extensive knowledge and patient help.
- Michal Budzowski - Media Archive integration
- Jerome Caffaro - CDS integration
- AV team - always ready to help with equipment and recordings.
- Nicola Tarocco - for working incredibly hard to get up to speed in one month!
- Tim Smith - UDS group leader
- Homer A. Neal - director of the U-M ATLAS Collaboratory Project, my supervisor and mentor since 2001, who was dreaming about automated lecture recording before that.
- ATLAS Collaboration - for helping fund the project and providing a demanding test environment for the micala system.