Internet Archive

Tracey Jaquith
FOMS 2010

- Lead storage & video engineer
- Been with IA for 10 years
- Wrote highspeed multithreaded crawlers and web servers
- Machine Vision CS education

Internet Archive

- Nonprofit accredited digital library
- Hosts:
    - 400K+ digi videos
    - 2.5M+ digi books
    - 800K+ digi audio / live concerts
    - 150B+ wayback urls
- Free infinite storage and serving forever
- 10 Gbit/sec bandwidth

Focus & Activity

- Archiving television with sister org.
- Moving to mwEmbed player
- Wish to expand to mashups/editing

Problem Areas

- Permalinking multi-video/file urls
- mwEmbed integration (close! 8-)
- back-end caching could improve
- better device support (OLPC, iPad, etc.)
- waiting for right time for WebM

Dependencies

- nginx_http_h264_streaming
- oggz-tools & oggz-chop
- theora_encoder
- ffmpeg
- mplayer
- avidemux
- mythcommflag (mythTV)
- mwEmbed
- metavid, Democracy Now!, PEG and more archive with us

Next steps & goals

- video editing support and packages
- captions search (eg: metavid)
- archive with more partners!
- integrate uploading/storage with morepartners/packages (eg: http, ftp, S3)