

Program Officer Report

**Aaron Binns
Internet Archive**

Who is the Program Officer?

→ Aaron Binns

→ Internet Archive

→ January 2008

→ Full-text search

→ aaron@archive.org

→ IIPC

→ Program Officer: 2010-2012

→ This is my second GA :)

Themes

- Rich Media
- Quality Assurance
- Research of Web Archives
- Social Web....archiving?
- Mobile devices

Social Web Archiving

- Social web of IIPC members
- Crowdsourcing
 - Volunteer curators
 - Volunteer harvesters
 - Volunteer archivists
- Reacting to world events
 - Subject matter experts
- Monitoring “endangered species” websites

Projects: current

→ Harvest Practices Report

→ Michaela Mayr, Austrian National Library

→ WARC Tools

→ Phase II redux

- Python implementation
- Testing & use by IIPC members

→ Phase III

- ARC->WARC conversion

Projects: future

- Olympics 2012
 - British Library
 - Focus: access
- WebAnalyzer
 - Czech Republic
 - Heritrix plugin(s)
- JHOVE2: WARC support
 - California Digital Library
 - Bibliothèque nationale de France

Programs

→ Staff exchange

→ “Our most valuable time is when we are meeting together in person”

→ Soliciting ideas & discussion

→ Other ideas?

→ Use the IIPC mailing list

Friday Workshop

- Technical platform/environment for research
 - Hadoop (distributed computation)
 - Pig (scripting)
 - Library to read WARC files
- Freebies!
 - Software & 10GB dataset
 - `$ tar xvfz pig-0.8.0.gz`
 - `$ cd pig-0.8.0`
 - `$./bin/pig`