



international internet preservation consortium

**GA 2010**

National  
Library  
Building  
Singapore  
3-7 May 2010

# **Program Officer Report**

**Aaron Binns  
Internet Archive**

# IIPC version 3.0

- 3-year term: 2010 to 2012
- New members
- New people in new roles
  - Program Officer: Aaron Binns (that's me)
  - Communications Officer: Abigail Potter
  - Working group co-leads

# Who is the Program Officer?

- Mr. Aaron Binns
- Software development background
- Small companies and start-ups
- Internet Archive
  - Joined 2008
  - Full-text search of web archives
- Pragmatic
- I like solving problems

# Transition from TO to PO

- Thanks to Gildas Illien (BnF)
- Impromptu tour as PO
  - Bibliothèque nationale de France
  - European Archive
  - Royal & National Library of Denmark
  - British Library
  - Hanzo Archives

# IIPC Funded Projects

- SmartCrawler -- **completed**
  - Heritrix 3
    - Released Decemeber 2009
  - Migration from 1.x
  - “Lessons learned” document
  - Ongoing work coordinated by Internet Archive and British Library
  - Monitored and reported via HWG

# Communication

- Membership is dispersed around the world
- Work often happens in silos
- Don't create new communication channels
- Edit and publish from existing channels
  - Greater frequency
  - Newsletter?
- Energy and enthusiasm from exposure

# IIPC Funded Projects

- ➔ **WARC Tools, Phase III – in-progress**
  - ➔ Work resumed after months-long delay
  - ➔ Hanzo committed to delivering on approved specifications and functionality
  - ➔ Challenge will be to manage timely delivery
  - ➔ Address performance issues from Phase 1&2
    - Deliver fixes in June, then proceed with Phase 3
    - Updated development plan forthcoming

# IIPC Funded Projects

- Best Practices Survey – **in-progress**
  - National Library Board of Singapore
  - Survey developed and distributed
  - Many responses received already
  - Full project and development plan forthcoming
  - Topic of discussion for HWG this week

# Joint Session

- Emerged from discussions with WG leads
- Streaming media applies to all WGs
  - Need coordinated approach, not silos
- It's an experiment
  - If it works, more in the future
- Topic selection by WG leads and PO
  - Work within the existing WG structure to propose topics for future joint sessions

# ISO Metrics for Web Archives

- Do not duplicate effort
- Report to IIPC in Vienna this autumn
- Possible topic for future joint session
- Topic applies to all WGs
  - Harvest: what data to collect during harvest?
  - Access: calculation, collation, and presentation
  - Preservation: interpretation and meaning
- Metrics often affect budgets

# Vision (tentative)

- Requires Steering Committee review and approval
- IIPC version 1.0: small, R&D, new
- IIPC version 2.0: expansion, dissemination
- IIPC version 3.0: ?
  - R&D
  - Packaged (vendor?) software solutions
  - Knowledge sharing
  - Data sharing (?)

# Funding Software Development

- Difficult and painful experiences so far
- Contract software development is not focus of IIPC member institutions
- Is IIPC funding truly needed?
- Smaller, narrowly-defined projects
  - Outsource to professional contract developers
- Distinguish between R&D and application development
  - Knowledge can be better than software

# Vendor solutions – when?

- Web archiving maturing as a field
- Budgets established
  - Part of digital preservation
  - Separate budget lines
- Vendors will start to take notice
  - Features of existing digital preservation systems
  - Entirely new vendor products
- New vendors

# Sharing data, sharing metadata

→ Sharing data is difficult

→ Legal limitations

- Privacy
- Copyright

→ Sharing metadata less difficult (I hope)

→ OAI-PMH interchange

→ Inter-archive linking

→ Many challenges remain

# Sharing knowledge & experience

- Strength of IIPC
- Face-to-face meetings highly valued
- Share knowledge, experience, know-how
  - More opportunities to meet?
  - More multi-institutional projects?
  - Workshops?
- Draw inspiration from OpenSource world
  - Apache, Linux, etc.

# React and Adapt

- The web changes fast!
- “web 2.0”
  - Interactive web sites
  - Flash
  - Streaming media
- Mobile devices
- Social networking and media
  - Walled gardens

# May we live in interesting times

## → Risks

- Spread ourselves too thin

- Overlap with other digital preservation groups

- Suit members' needs, big and small

## → Solve one important problem each year

- Streaming media for 2010