

## Harvesting Working Group

**Wednesday, May 11, 2011, 10:00 am – 1:00 pm**

Room B/C, KB

- 10:00 am Welcome and introductions
- How many websites are there
- Summary of TLD size data (Kristinn Sigurðsson)
  - Alias domains (Bjarne Andersen)
- 11:00 am COFFEE BREAK
- 11:30 am Harvesting of dynamic websites
- Google #fragments (Rene Voorburg)
  - Facebook (Lewis Crawford)
  - Discussion
- Heritrix 3.1 (Gordon Mohr)
- 1:00 pm LUNCH
- 2:00 pm Joint WG session – Quality Assurance

**Thursday, May 12, 2011, 10:00 am – 1:00 pm**

Boardroom, KB

- 10:00 am Quality Assurance follow-up (discussion)
- Browser as a crawler (Gordon Mohr)
- 11:00 am COFFEE BREAK
- 11:30 am Harvesting Working Group future
- Summary of e-mail poll of HWG membership
  - Discussion
- 1:00 pm adjourn

## Preservation Working Group

Tentative agenda

**Wednesday, May 11, 2011, 10:00 am – 1:00 pm**

Boardroom, KB

- 10:00 am Review and discussion of accomplishments/WP's status
- 11:00 am COFFEE BREAK
- 11:30 am Review PWG database
- Discuss relationship with UDFR/Pronom
  - Discuss repurposing PWG data for public IIPC site
- 1:00 pm Lunch
- 2:00 pm Joint WG session – Quality Assurance

**Thursday, May 12, 2011, 10:00 am – 1:00 pm**

Boardroom, KB

- 10:00 am WP 8 – Risk Assessment
- Review of database
  - Review of testing assessments
  - Future work (analysis/output tools, policy development)
- 11:00 am COFFEE BREAK
- 11:30 am WP 7 – Information Packages. Conclusions from survey
- Review of other projects outside IIPC related to web archives preservation
- 1:00 pm adjourn

## **Access Working Group Meeting**

10:00am – 1pm; Wednesday and Thursday, May 11th and 12th  
Joint Sessions, Wednesday, May 11th 2-3:30pm, 3:30-5pm  
Room A, KB

### **WEDNESDAY**

10am-11am	Memento Proposal Discussion
11am-11:30am	COFFEE BREAK
11:30-1pm	Olympics 2012
1pm	LUNCH
2pm	Quality Assurance Joint Session with PWG & HWG Case Study: NARA 2010 End of Term collection Automated QA Workflow-Based QA
3pm	COFFEE BREAK
3:30pm-5pm	Continuation of Joint Session w/breakouts

### **THURSDAY**

10am-11am	Replay of Web resources (Wayback et. al.), Interface Architectures, Interface & Navigational Design for Longitudinal Archives Includes demo of XTF by CDL
11-11:30am	COFFEE BREAK
11:30am – 1pm	Researcher & Research Initiatives & Updates, SOLR & other search Updates, Open Discussion
1pm	LUNCH - adjourn