WEB ARCHIVING OVERVIEW

National and University Library - Slovenia

Janko Klasinc | janko.klasinc@nuk.uni-lj.si | +386 01 2001 211
EARLY PROJECTS

2002 – 2004
Slovenian electronic web publications collecting and archiving methodology

2003 – 2004
Development and analysis of Slovenian digitized and electronic publications collection of national importance
LEGAL BASIS

2006

Legal deposit law
(Zakon o obveznem izvodu publikacij (Ur. list RS, št. 69/06 in 86/09)

2007

Regulation on types and selection of electronic publications for legal deposit
(Pravilnik o vrstah in izboru elektronskih publikacij za obvezni izvod, (Ur. list RS, št. 90/07)
2008 - Selective harvesting (1,400+ websites):
• government websites
• research & higher learning institutions
• online periodicals
• arts and culture institutions
• etc.

Themed crawls: parliamentary elections, local elections, important events (politics, sports, etc.).
2014 – National domain .si crawl (biannually) Heritrix 1.14.4. and 3.4
DATA COLLECTED

National domain, selective & thematic crawls:

- 560.066 domains
- 513.793.472 URLs
- 45,5 TB

Saff: 0,25 FTE?
CURRENT ACTIVITIES

- moving WCT, Heritrix & Wayback to new servers (separating crawling from access);
- focused crawl of 50 government domains before the content is moved to a single domain;
- providing access to the national domain crawls;
- rethinking legal basis for free access.
THANK YOU!

janko.klasinc@nuk.uni-lj.si
+386 01 2001 211