Warc?

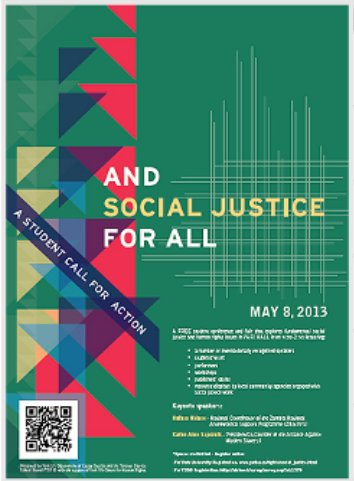
Warkwark wark? Wark Warrrk

YORK UNIVERSITÉ UNIVERSITY

this is my time.

yFile
YORK UNIVERSITY'S DAILY NEWS

HOME | TOP STORIES | CURRENT NEWS | CAMPUS NOTES | YORK IN THE MEDIA | RESEARCH | CONTACT



York University's Sakura cherry trees are in full bloom

Social justice conference and fair, a call to action, Wednesday

SAS Canada donates leading industry software to Schulich

PREVNet conference turns lens on cyberbullying

Harvard Professor Peter Galison to discuss intersection of physics, philosophy and the history of science

Summer Institute in Visual Arts considers 'Unexpected Encounters: Sexual Difference and Art Now'

FEATURED EVENTS

Mothers Speak Up! On Parenting Children with Disabilities - May 6, 9:30-1:30pm

And Social Justice for All: A Student Call for Action - May 8, 9:30-2:30pm

Muscle Health Awareness Day - May 17, 9-4pm

Speakers Series: Let's Talk - Difficult Conversations (SSDCON) - Jun 21, 8:30-10:30am

more events

YORK IN THE MEDIA

Most people don't have viable financial plan: poll

Canada's universal health-care system makes us live democratically

Canada Post issues stamps celebrating Chinatown gates, two GTA locations included

Laptops impede learning as students surf, York U study says

current news

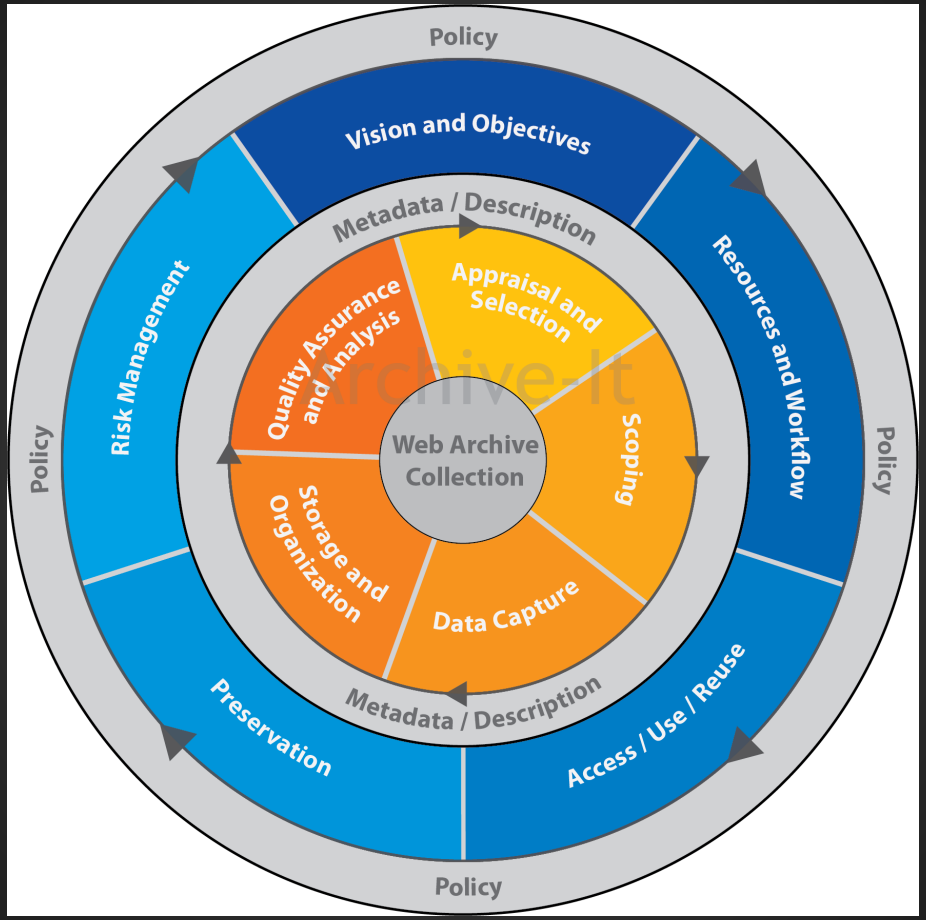
University Bookstore makes changes in time for May rush

Third TEDxYorkU enjoys record attendance

Schulich MBA team wins global real estate case competition

Marathi Dalit writer, intellectual and activist to speak at York

York celebrates innovation in teaching and learning – May 9



http://archive-it.org/static/images/WALCM_model_final.png

Software & Tools

Wget

```
$ wget --warc-file=foo http://yfile.news.yorku.ca/

Opening WARC file ‘foo.warc.gz’.

--2013-05-08 10:17:49--  http://yfile.news.yorku.ca/
Resolving yfile.news.yorku.ca (yfile.news.yorku.ca)...
130.63.173.84
Connecting to yfile.news.yorku.ca (yfile.news.yorku.ca)|130.
63.173.84|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: ‘index.html’

0K ......... ......... ......... . 180K=0.2s

2013-05-08 10:17:49 (180 KB/s) - ‘index.html’ saved [32084]
```

Heritrix

crawler.archive.org

Archive-It

archive-it.org

Twarc

github.com/edsu/twarc

Social Feed Manager

github.com/gwu-libraries/social-feed-manager

Community

International Internet Preservation Consortium (IIPC)

netpreserve.org

SAA
Web Archiving Roundtable

webarchivingrt.wordpress.com

ArchiveTeam

archiveteam.org

OCUL

OCUL Digital Curation Community

"Wark?
Warkwark wark?"
Wark
Warrrk

Questions!

Contact

Nick Ruest
Email: ruestn@yorku.ca
Twitter: @ruebot