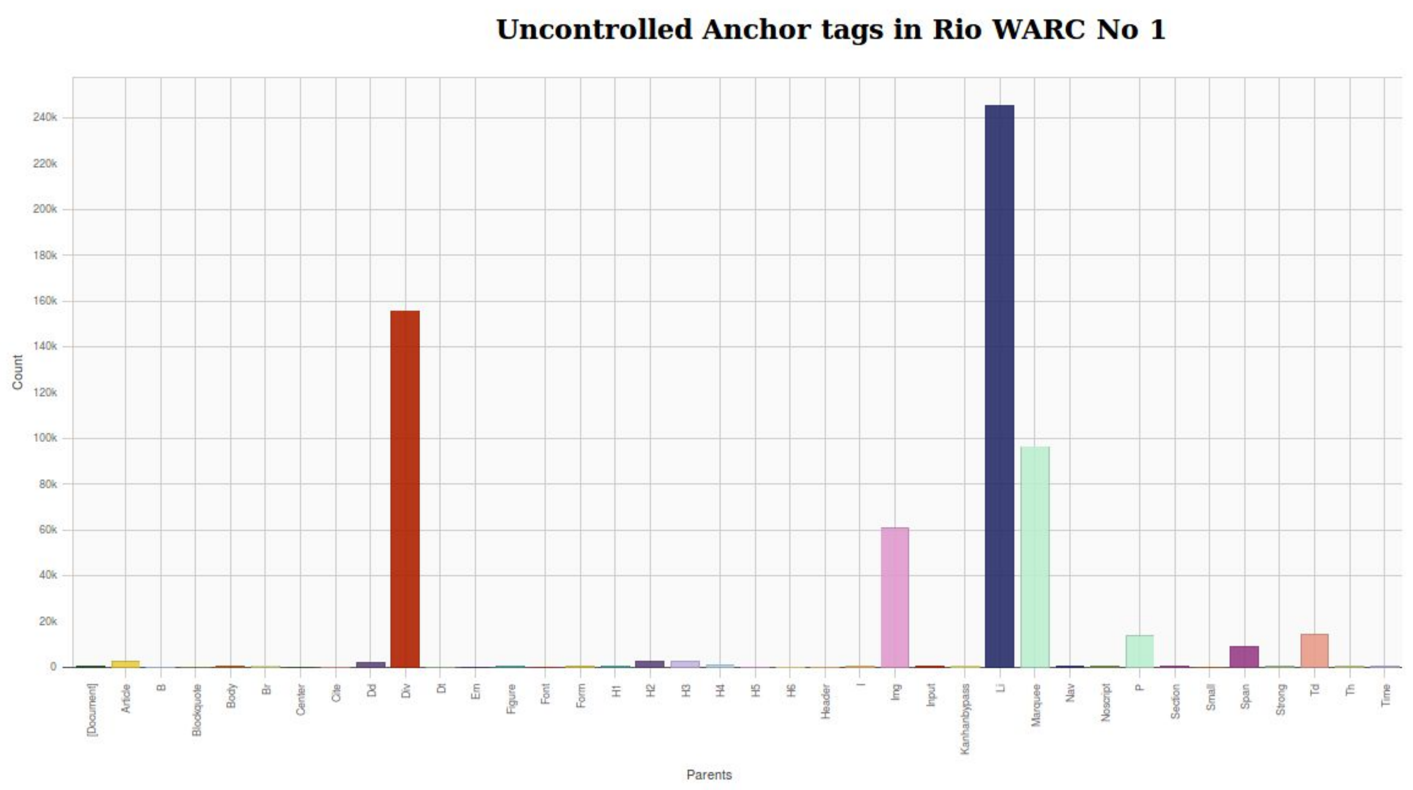
# Link Ranking group

Gregory Wiedeman (University at Albany, SUNY)
Mindaugas Vidmantas (The British Library)
Peter Webster (Independent Scholar & Consultant)
Kees Teszelszky (National Library of the Netherlands)
Richard Deswarte (University of East Anglia)

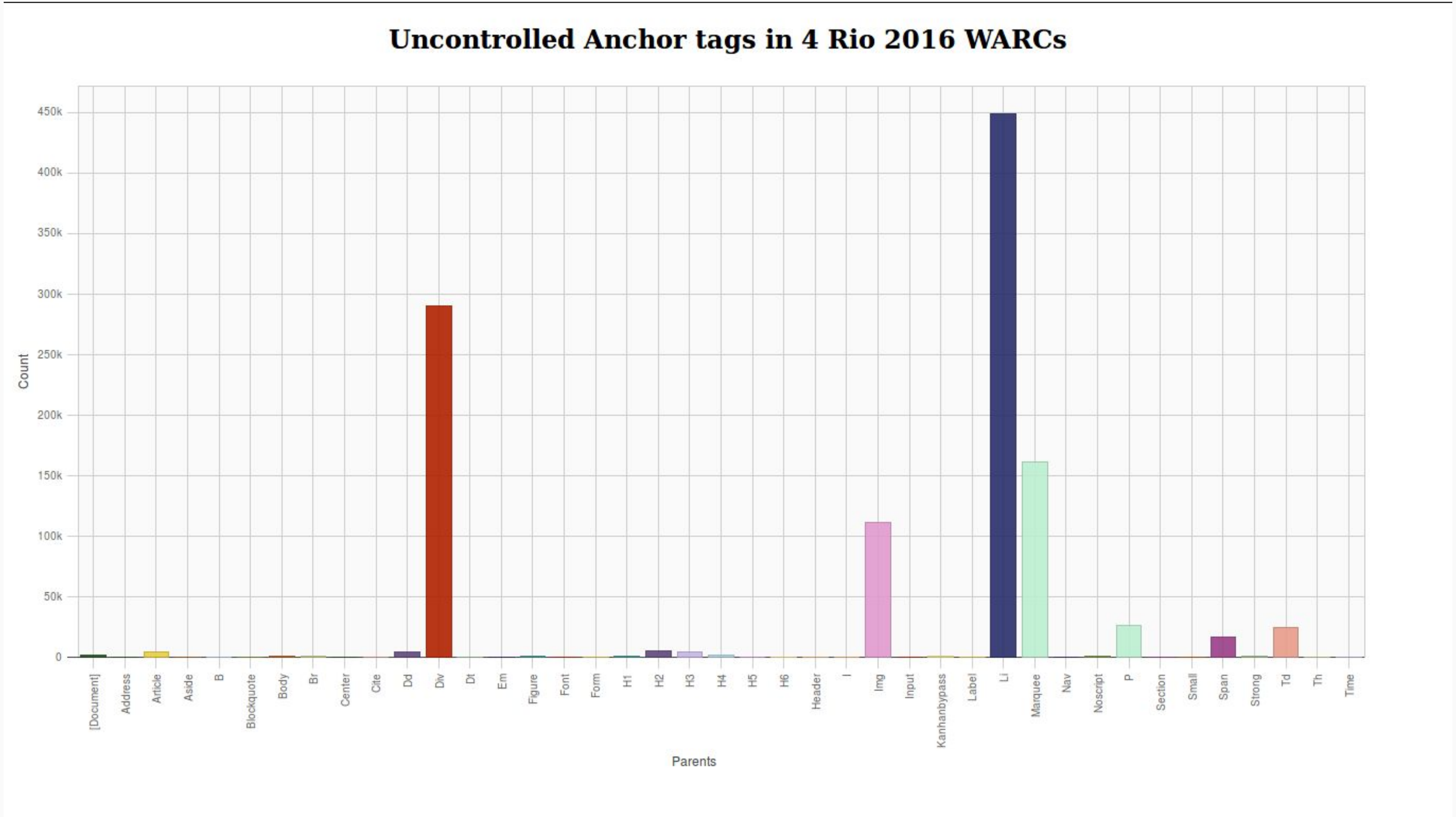
# Are All Links Created Equal?

- WarcBase scripts to export manageable raw HTML
- Load into BeautifulSoup Python library
- Look for <a> parents
- Should we weigh links with certain parents more during during link analysis?
- Are navigational links (<li>) different that content links (<p> or <div>)?

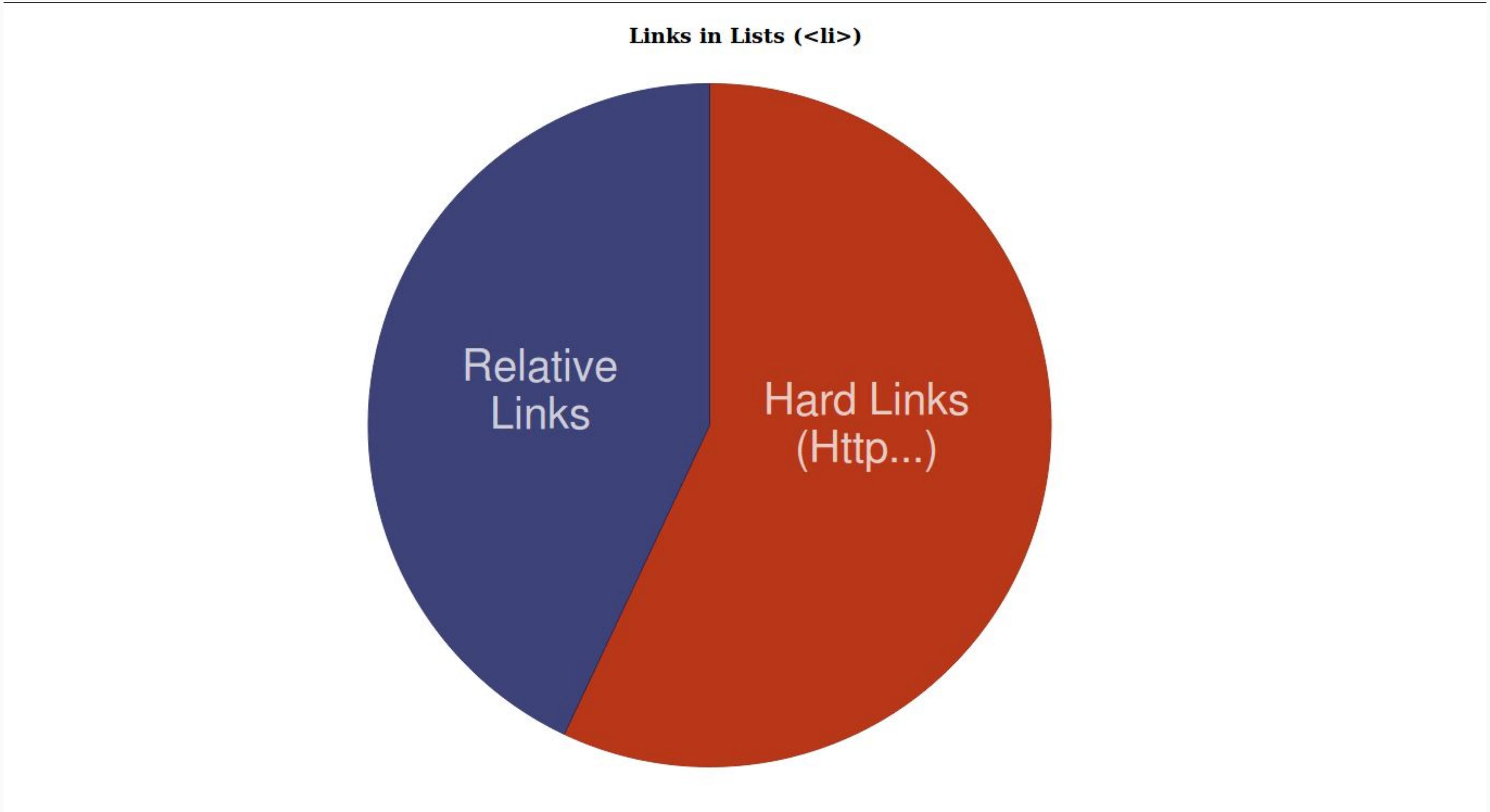
# 1 Rio WARC Link Parent Element Type Distribution



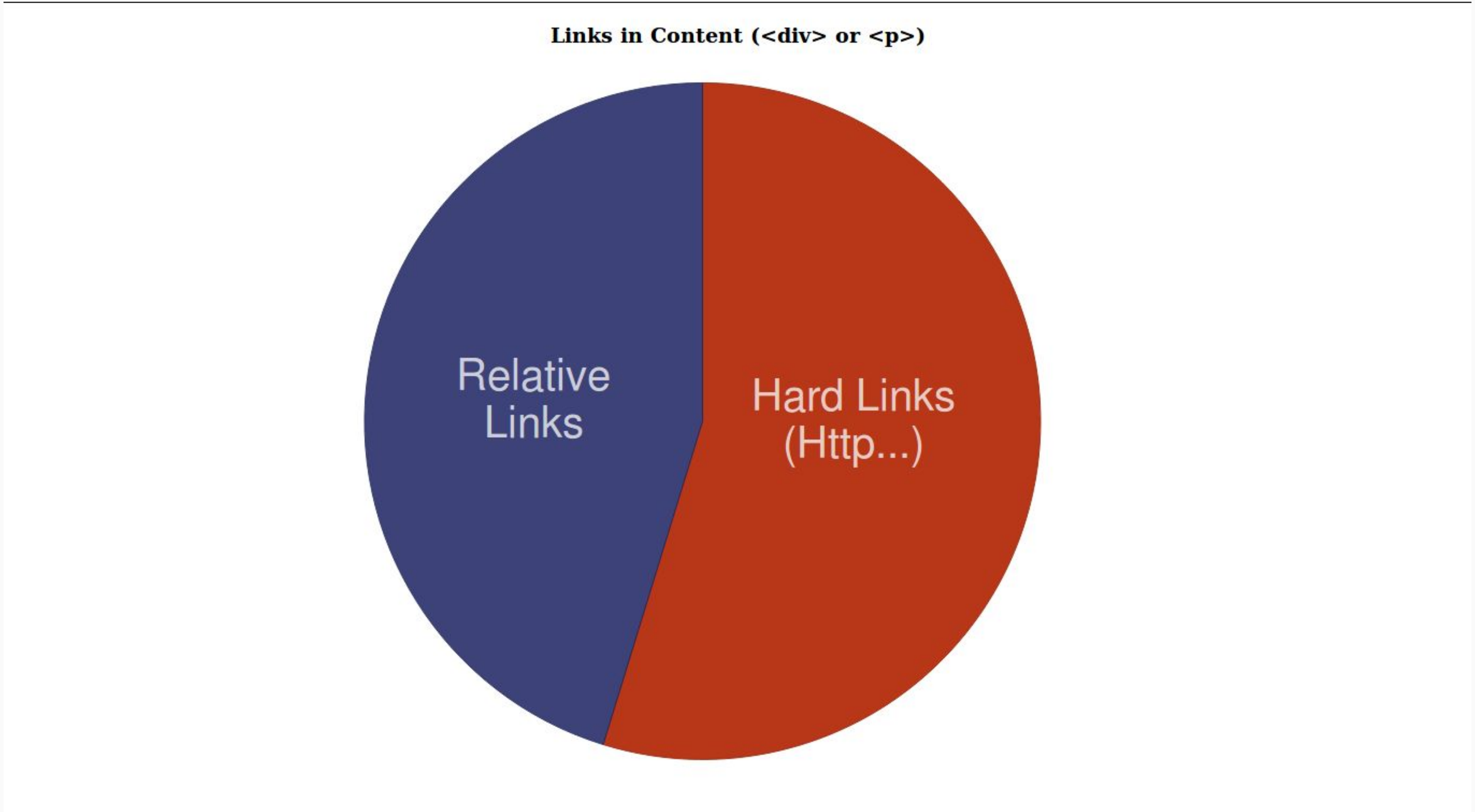
# 4 WARC Link Parent Element Type Distribution



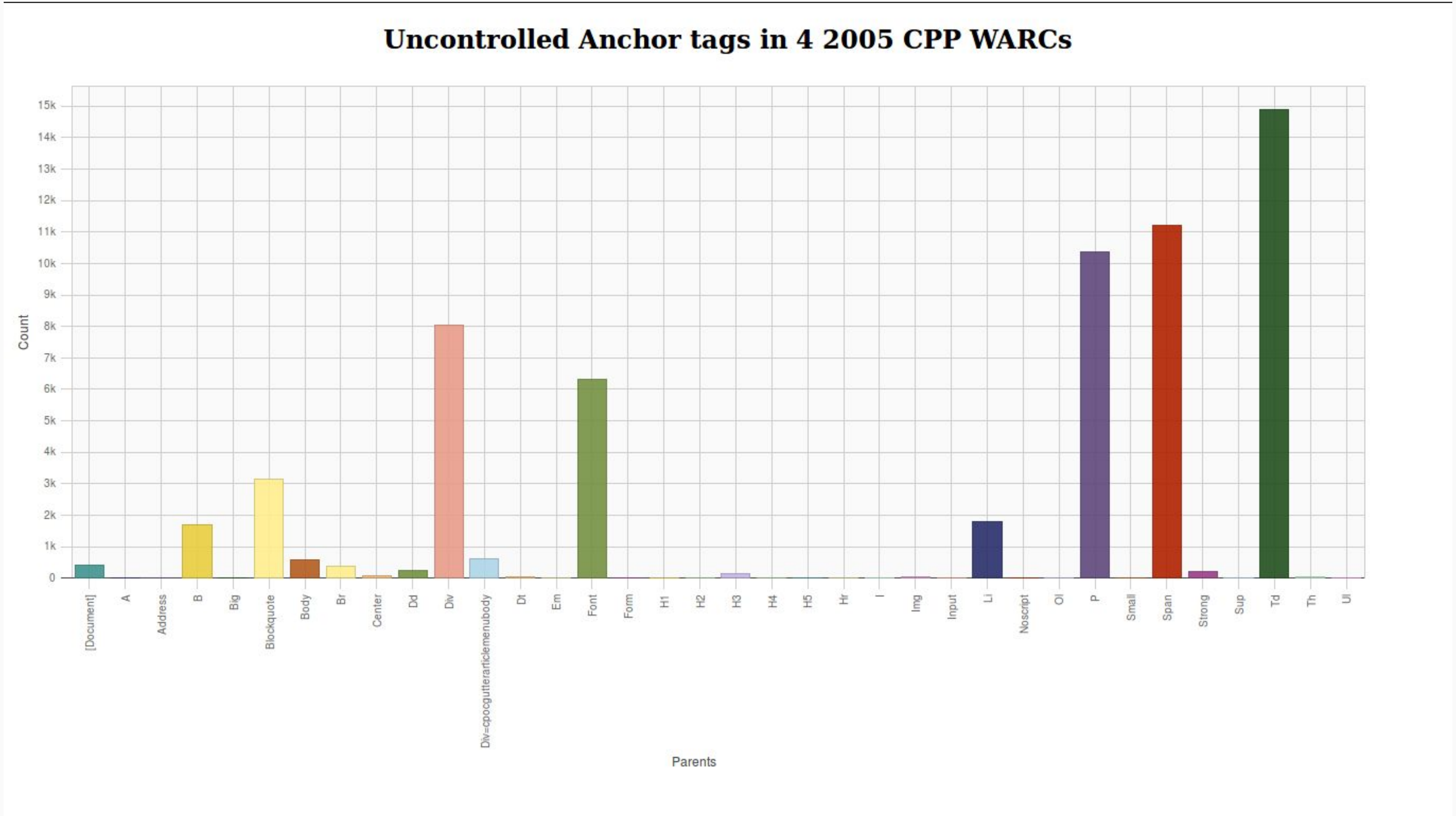
# Relative vs hardcoded in Rio (2016)



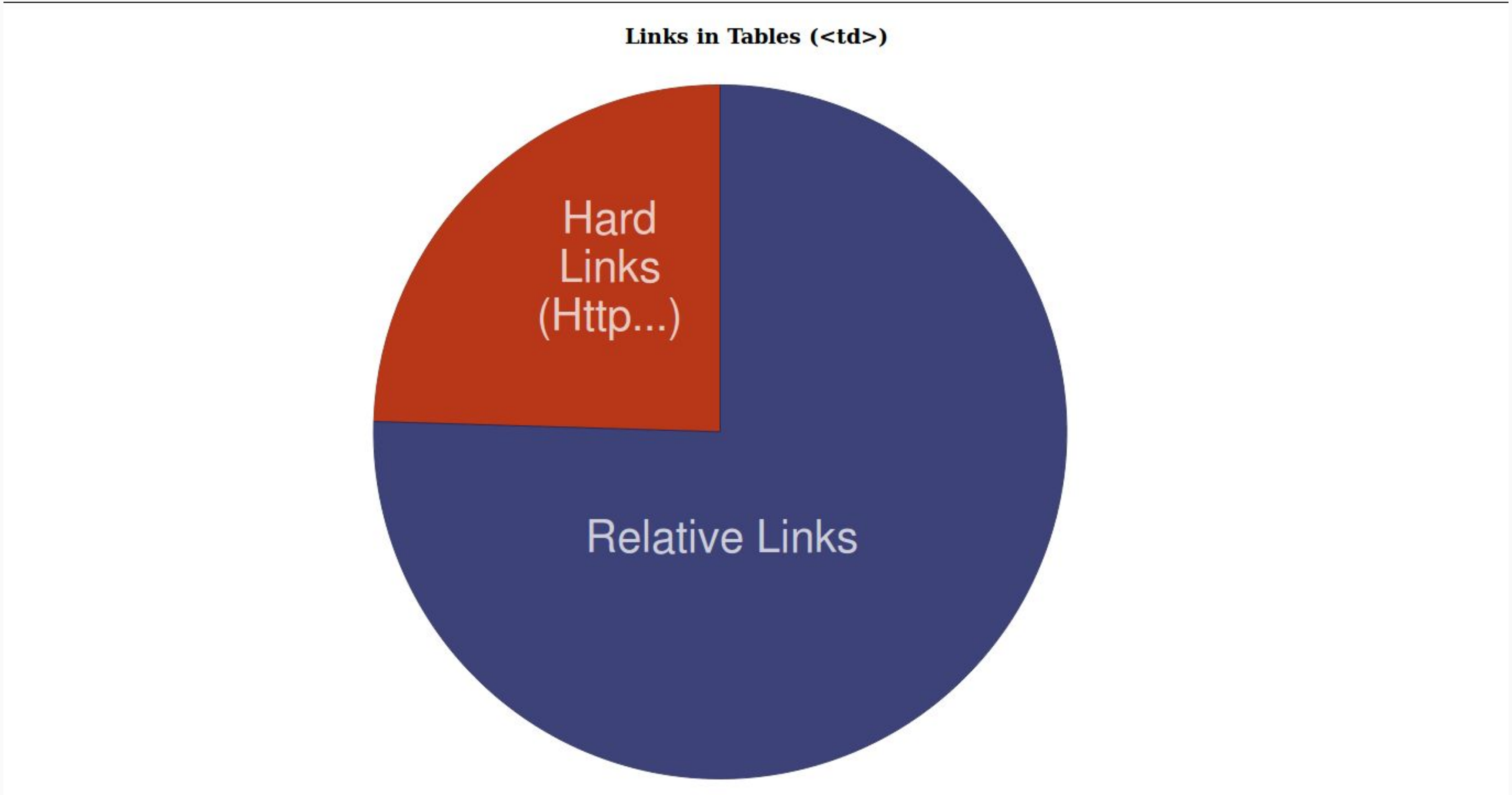
# Relative vs hardcoded in Rio (2016)



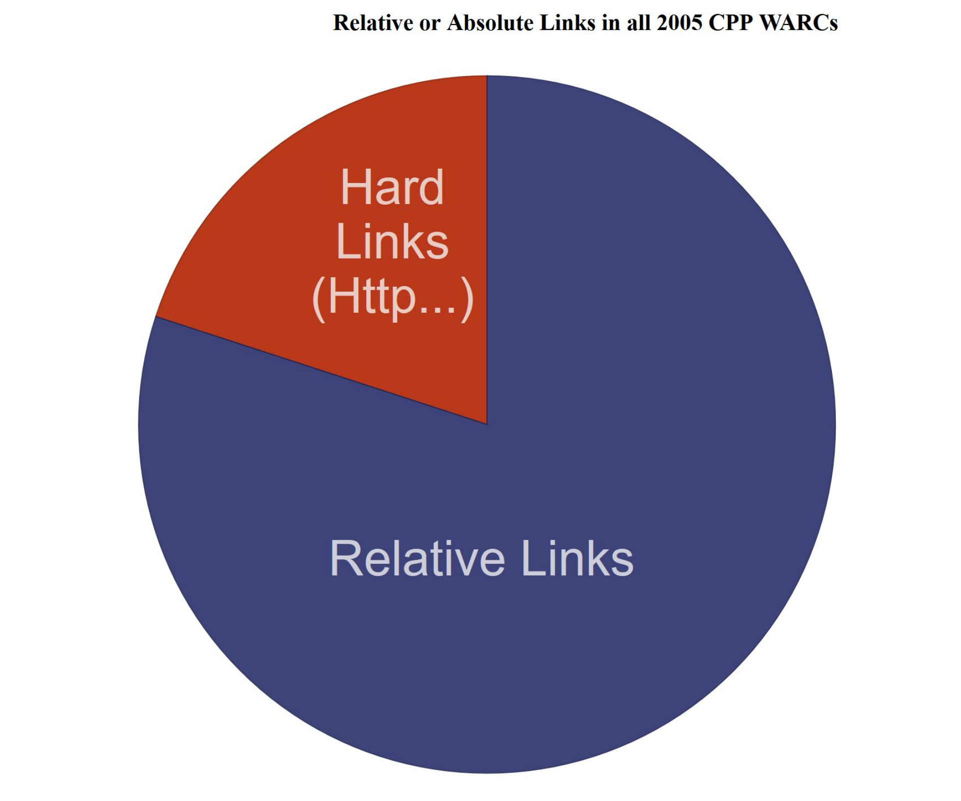
# CPP Link Parent Element Type Distribution

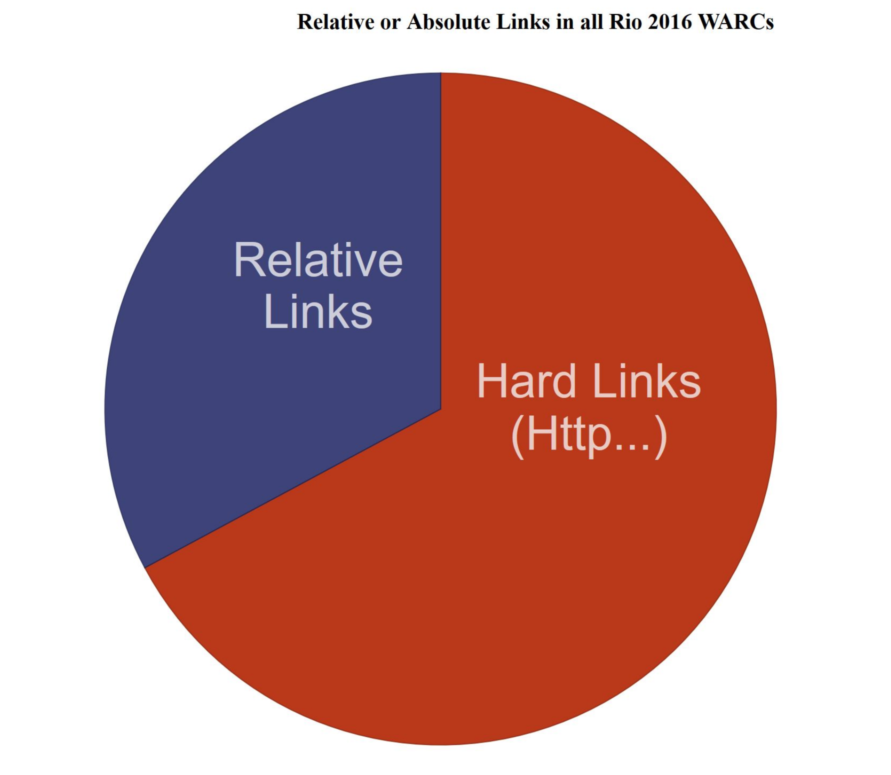


# Relative vs hardcoded in CPP (2005)



# Absolute and Relative Paths  in 2005 CPP vs. 2016 Rio WARCs


Relative or Absolute Links in all 2005 CPP WARCs
Hard Links (Http...)
Relative Links


Relative or Absolute Links in all Rio 2016 WARCs
Relative Links
Hard Links (Http...)