#HTRC  #DH2017

@MCDONALD  |  @HATHITRUST
HA THITRUST

A DIGITAL LIBRARY (15.7M)
Languages in HT

1. English  54%
2. German  10%
3. French  8%
4. Spanish  4%
5. Russian  3%
6. Chinese  3%
7. Other  18%
HATHITRUST RESEARCH CENTER
AN ANALYSIS CENTER
HTRC Extracted Features DS

13.7M Volumes
8.3M In-Copyright
5.4M Public Domain
5B Pages
GOOGLE BOOKS + RC CONCEPT
NON-CONSUMPTIVE RESEARCH POLICY
HTTPS://GOO.GL/LVPKSS
DISTANT PUBLISHING

1. Non-consumptive in nature
2. Removed by x degree from the creator
3. Capable of being open and re-usable
4. Outside of existing publishing workflow
HTRC USE CASES

- Extracted Features Workset
- HT+Bookworm
- HTRC Workset Ontology
- Custom HT Collection
CURRENT DP WORKSETS

1. Boris Capitanu, Ted Underwood, Peter Organisciak, Timothy Cole, Maria Janina Sarol, J. Stephen Downie (2016). The HathiTrust Research Center Extracted Feature Dataset (1.0) [Dataset]. HathiTrust Research Center, http://dx.doi.org/10.13012/J8X63JT3


HTRC AS DISTANT PUBLISHER
PEER REVIEW

DATA CURATION + DATA REPRODUCIBILITY
Acknowledgments

HTRC Executive Leadership
Beth Plale | J. Stephen Downie
Harriett Green | John Unsworth
John Walsh | Eleanor Dickson | Marie Ma

HTRC Funders
Andrew W. Mellon Foundation
Institute for Museum and Library Services

cc: Wessex Archaeology - https://www.flickr.com/photos/32884265@N00
CONTACT INFO

@MCDONALD I ROBERT@INDIANA.EDU

QR code