

Digital Storage at the British Library

Dr Andy Jackson

@anjacks0n

The British Library

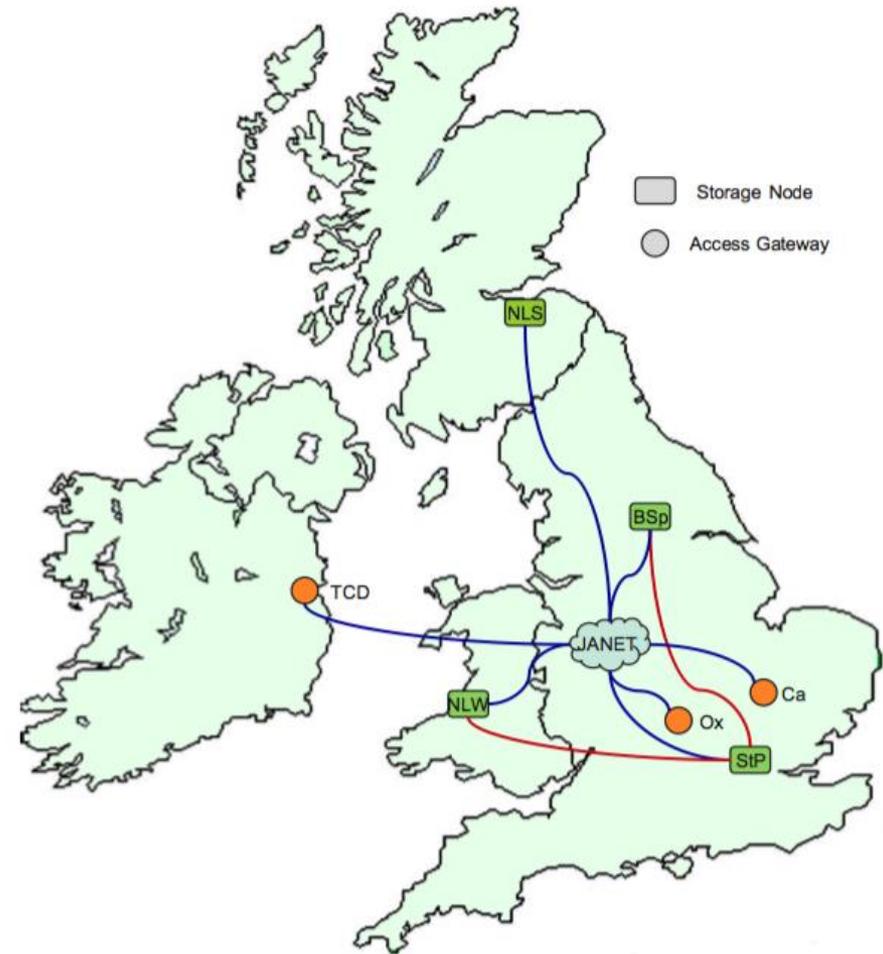
- A National Legal Deposit Library:
 - One of six for the United Kingdom
- A wide range of digital content:
 - Digitisation projects since at least 1995
 - Voluntary digital deposit programme
 - Digital Legal Deposit since 2013
 - Rapid growth in audio & video
 - Research data on the horizon
- Just over 1PB of data in our archival store:
 - Peak ingest ~600GB/day



Content Stream	2017 Total (TB)
eBooks & Journals	240
Web Archive	120
Sound & Video	240
Newspapers & Monographs	420
TOTAL	1,020

The Current Digital Library System

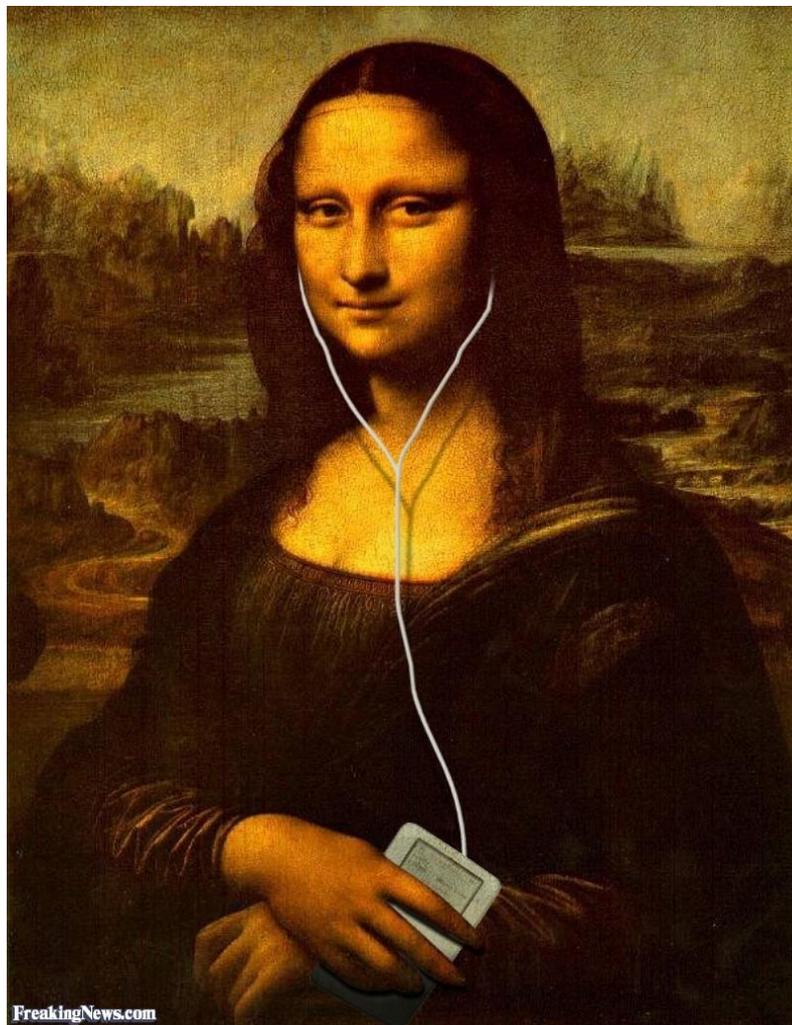
- 4 Nodes (Complete Copies)
 - British Library, St. Pancras
 - British Library, Boston Spa
 - National Library of Wales
 - National Library of Scotland
- Additional Access Points
 - Bodleian Library, Oxford
 - Cambridge University Library
 - Trinity College Library, Dublin



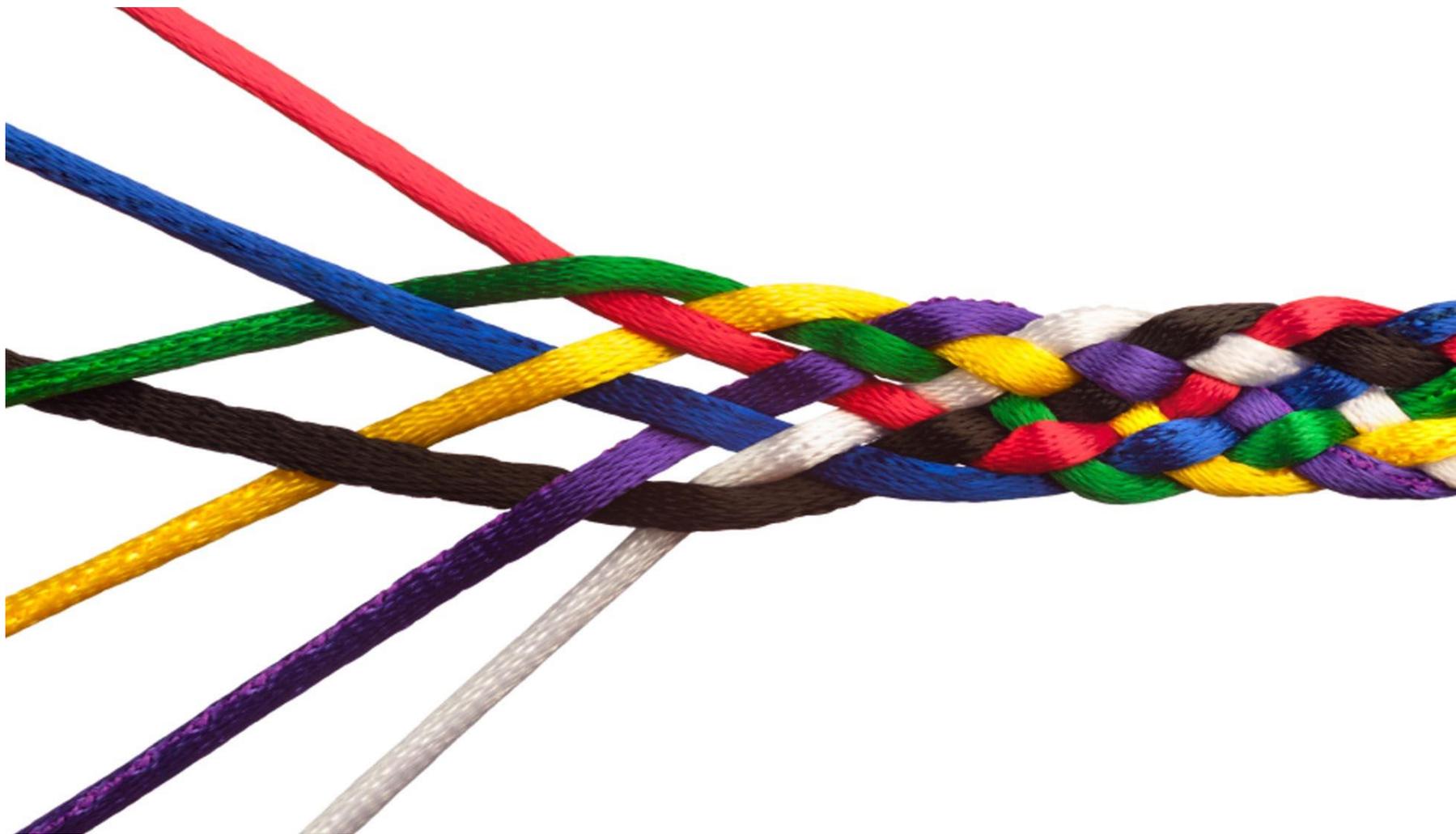
Improved Storage Architecture Needed

- Rapid growth expected:
 - Target ingest peak capacity of 4TB/day
- Threat model needs review:
 - Mitigate a wider range of risks, at the same cost?
- Replace in-house with open standards if possible:
 - Separate storage & replication from cryptographic verification
 - Ease storage growth & refresh management
- Enable data-mining all content at at least one location:
 - It's not just about single-item access anymore

Trust



Entangled Checksum Chains



Thanks!

Andrew.Jackson@bl.uk

@anjacks0n