



# DIY Digital Preservation

Neil Jefferies, Head of Innovation, Bodleian Digital Libraries Systems and Services



# DP is more than just technology...

## 10 core requirements from TRAC

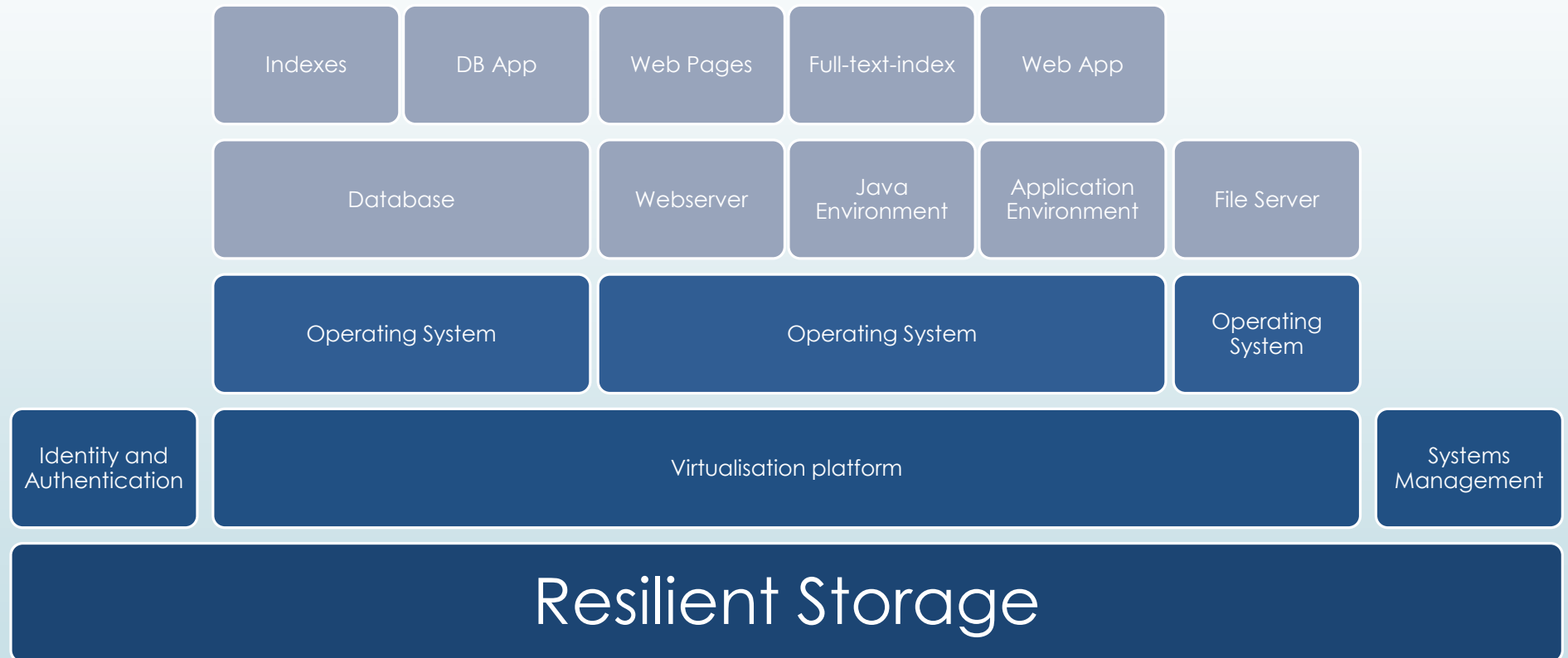
1. Mandate and commitment to digital object maintenance
2. Organizational fitness
3. Legal and regulatory fitness
4. Efficient & effective policies
5. Adequate technical infrastructure
6. Acquisition and ingest
7. Preservation of digital object integrity, authenticity & usability
8. Metadata management & audit trails
9. Dissemination
10. Preservation planning and action



# Why DIY?

- ▶ OTS packages/service offerings don't meet requirements (or not for a reasonable price!)
  - ▶ Workflow customisation
  - ▶ Integration with existing/legacy systems
  - ▶ Trust – retain control/eliminate dependencies
- ▶ DIY Development/hosting or both?
  - ▶ In house skill requirements
    - ▶ Training
    - ▶ Retention
- ▶ Risk management – cost/benefit

# Not Really DIY!





# A couple of quotes

If you want to go fast, go alone. If you want to go far, go together.

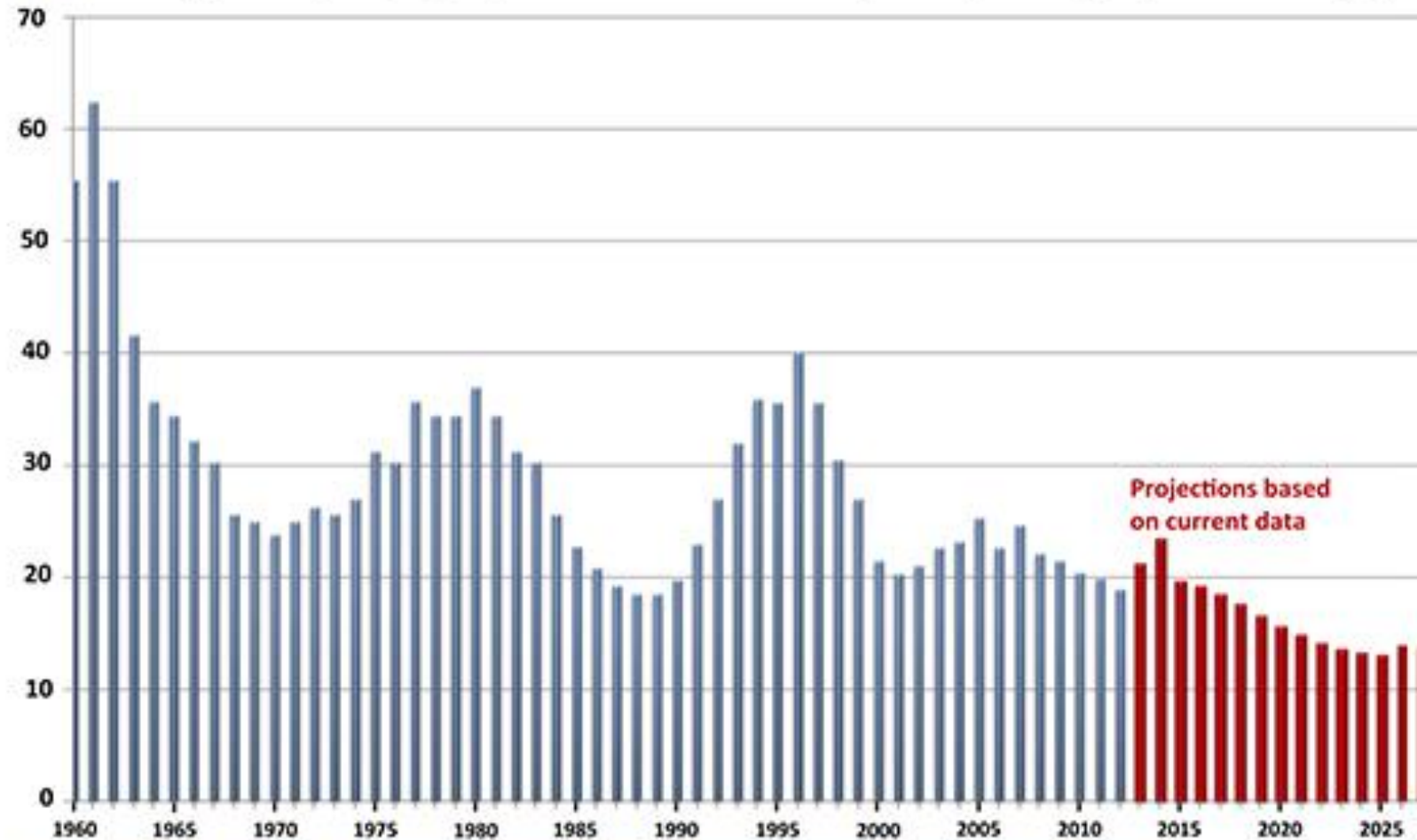
*African proverb (allegedly)*

Digital preservation is the series strategies and actions taken to promote the availability and usability of digital information over time.

*Tom Cramer, Stanford University*

Make use of communities like DPC

**Average company lifespan on S&P Index in years (rolling 7-year average)**



*Year (each data point represents a rolling 7-year average of average lifespan)*

DATA: INNOSIGHT/Richard N. Foster/Standard & Poor's

# Why preserve? Use cases

## ► **Compliance archiving**

- Often structured – financial and HR records from databases
- Defined retention period – often with a strong incentive to destroy afterwards
- Emphasis on security and authenticity
- Penalties for failure

## ► **Corporate memory**

- Less structured - documents and images
- Indefinite/extended retention period
- Emphasis on discovery, usability
- Benefits for success





# Procuring communities instead of vendors

## ► **Technical Communities**

- Reasons for DIY - Customisation, Integration, Trust
  - Open Source (in some form)
- Community should provide **vision**, support, continuity, training, documentation
  - Basis for evaluation
- Single product (Fedora) vs stacks (Hydra)
- Fit with organisational standards/skills

## ► **Communities of Practice**

- Shared concerns (up to a point!)
- User groups for systems that need integration

## ► **Standards Groups**

- Shape how products are developed

## ► **Still need to evaluate products against requirements**



A decorative graphic on the left side of the slide. It features a dark blue vertical bar on the far left. A black arrow points to the right from the top of this bar. Several thin, light blue lines curve upwards and to the right from the bottom of the bar, overlapping the main content area.

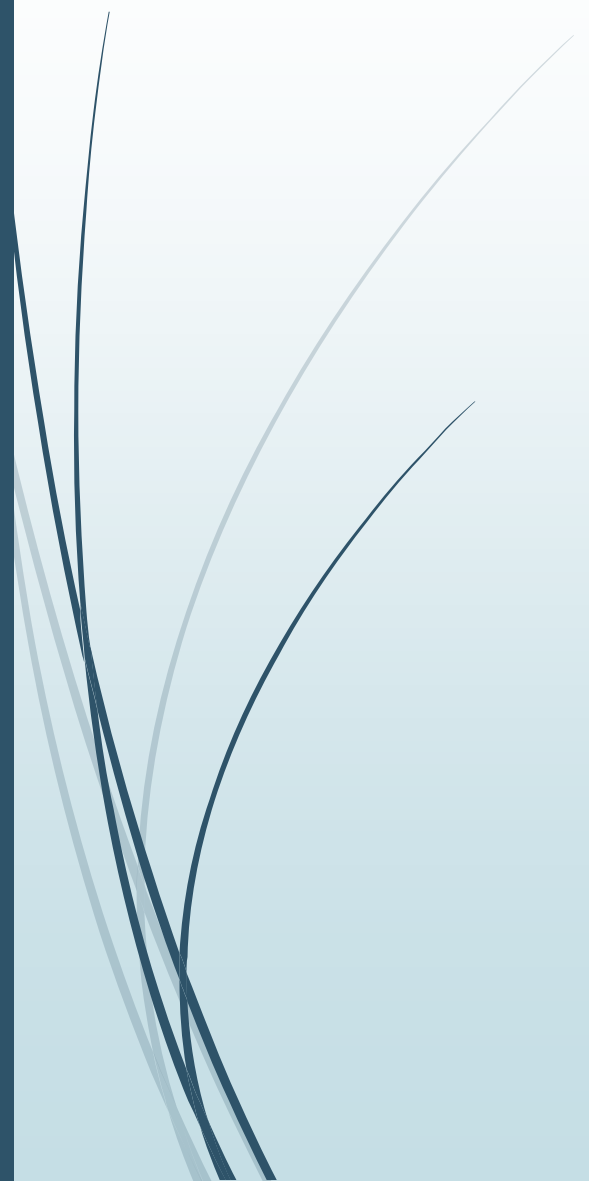
# Levels of engagement

- ▶ Open Source is not free
  - ▶ Much value lies in the community which relies on contribution
  - ▶ Costs are much harder quantify
    - ▶ And may vary over time
  - ▶ Contribution can be made in many ways
    - ▶ Cash
    - ▶ Developer time
    - ▶ Working group participation
    - ▶ Hosting events
    - ▶ Doing documentation
    - ▶ Opportunities for creative accounting



# Revisiting the idea of Corporate Memory

- ▶ Objects are stored for their knowledge content
  - ▶ A lot of meaning depends on information that is not intrinsic to the object
  - ▶ Metadata about context
    - ▶ People (corporate biographies) and places (especially corporate locations)
      - ▶ Digital objects with their own metadata
  - ▶ Metadata about process (W3C PROV-O)
    - ▶ Events, provenance, versioning
  - ▶ Corporate ontologies to reflect culture, brands, language (e.g. Mars Inc)
- ▶ Graphs are a better fit than hierarchies
  - ▶ Typed relationships rather than relational databases
  - ▶ Linked Data/Semantic Web technologies



# **PASIG 2017**

**Preservation and Archiving Special Interest Group**

**11 - 13 September 2017, Oxford, UK**



# DIY Digital Preservation

Neil Jefferies, Head of Innovation, Bodleian Digital Libraries Systems and Services