

The Parsimonious Anthropologist



The Influence of string figures and a friar from the 14th Century on the preservation of digital research data.



Preservation? (1)



Preservation? (2)



Preservation? (3)



Cats Cradle



Cats Cradle



Cats Cradle



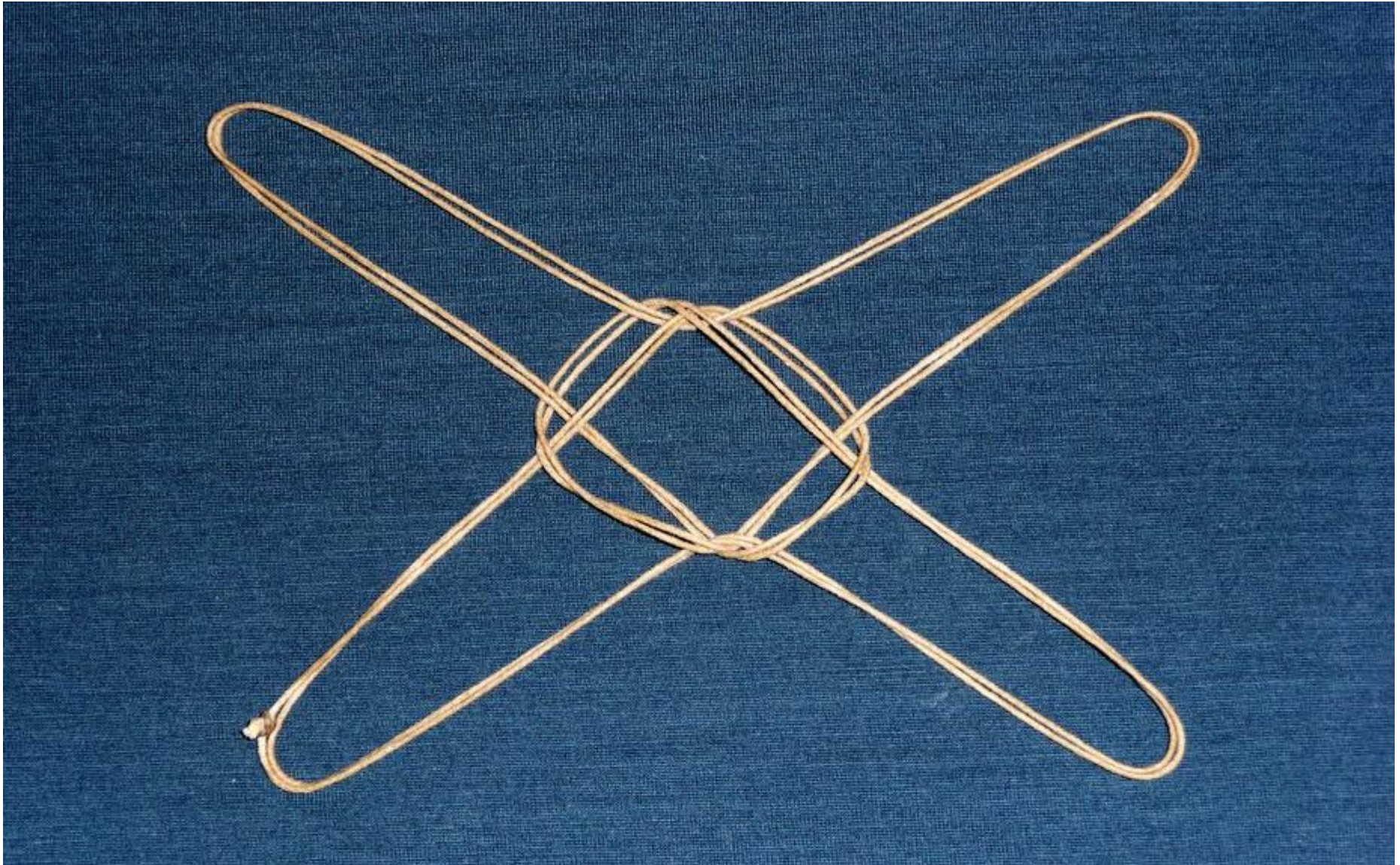
Cats Cradle



String Figure – Opening A



Preservation of The Artefact



Notation – Opening A



- P1
- rF ml-pu IPS:ex
- IF mr-th rFN pu rPS:ex
- Where is the Essence of The String Figure?

What shall we preserve ?

- Call The Archivist.
 - Only the important (valuable?) stuff !
 - The stuff which we can afford to preserve !
- How might we define “Important”?
 - Depends on the domain
 - The essence of the String Figure?
- How can we understand “Afford” ?
 - Now we need the 14th Century Friar !

William of Occam and his “Razor”

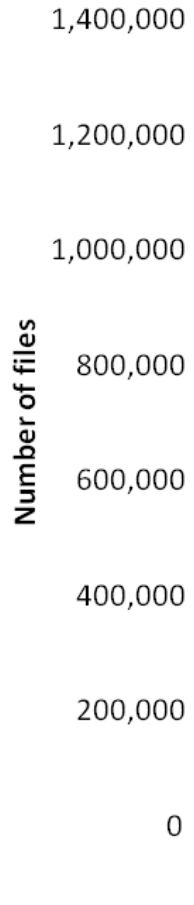
- pluralitas non est ponenda sine necessitate.
 - OR
- Frustra fit per plura quod potest fieri per pauciora.
 - Migrated user access!
- plurality should not be posited without necessity
 - OR
- It is futile to do with more things that which can be done with fewer

For Digital Preservation

- Only do the minimum needed to achieve the outcome (Preservation)!
- only do it when you have evidence you need to !
- Not as easy as it sounds
 - Lots of assumptions
 - Less evidence!

The File format Long Tail

- Administrative records (Top 25 = 98%)



File Extension	Count	%
msg	1225790	57.6
doc	437803	20.6
xls	135573	6.4
pdf	82524	3.9
jpeg	58748	2.8
url	20701	1.0
xlsx	16777	0.8
docx	16765	0.8
htm	15616	0.7
txt	14199	0.7
ppt	12834	0.6
gif	11154	0.5
rtf	11046	0.5
tif	9293	0.4
MBX	6044	0.3
png	4812	0.2
obr	4368	0.2
bmp	3286	0.2
zip	2499	0.1
html	2307	0.1
csv	2266	0.1
vsd	2088	0.1
DAT	1972	0.1
xml	1898	0.1
pptx	1700	0.1
sql		
VMBX		
dwg		
psd		
mht		
css		
log		
ico		
mp3		
tmp		
mno		
dat		
eps		
db		
exe		
sgm		
js		
dll		
com		
xlsm		
bat		
DTA		

File Format

Risk and Response

- Impact
- Probability
- Proximity
 - Not necessarily part of the institutions risk model!
- When will your Preservation Risks be realised?
- What is the lifetime of your preservation system?

Breaking News

Crystal balls are unreliable



Digital Preservation / Archiving

- Parsimonious Preservation
- Know What you have got
 - Technical Information (File Format, Where stored)
 - Cataloguing (Who, Where, When, what Context)
 - Its not Meta its Data !
- Keep the bits safe
 - Multiple Copies, Multiple Locations
 - Check Sums
 - Bit Rot and System Failure modes

String Figures again.



Questions?

Tim Gollins

Head of Digital Archiving
National Records of Scotland