

# What if Research Data were Not Forever?

Rob Cook & Rhys Francis

# Introduction

We do know about keeping data

- Storage technology => store almost all data almost forever
- Preservation technology => access all data stored almost forever
- Data technology eg FAIR => re-use much data for longer

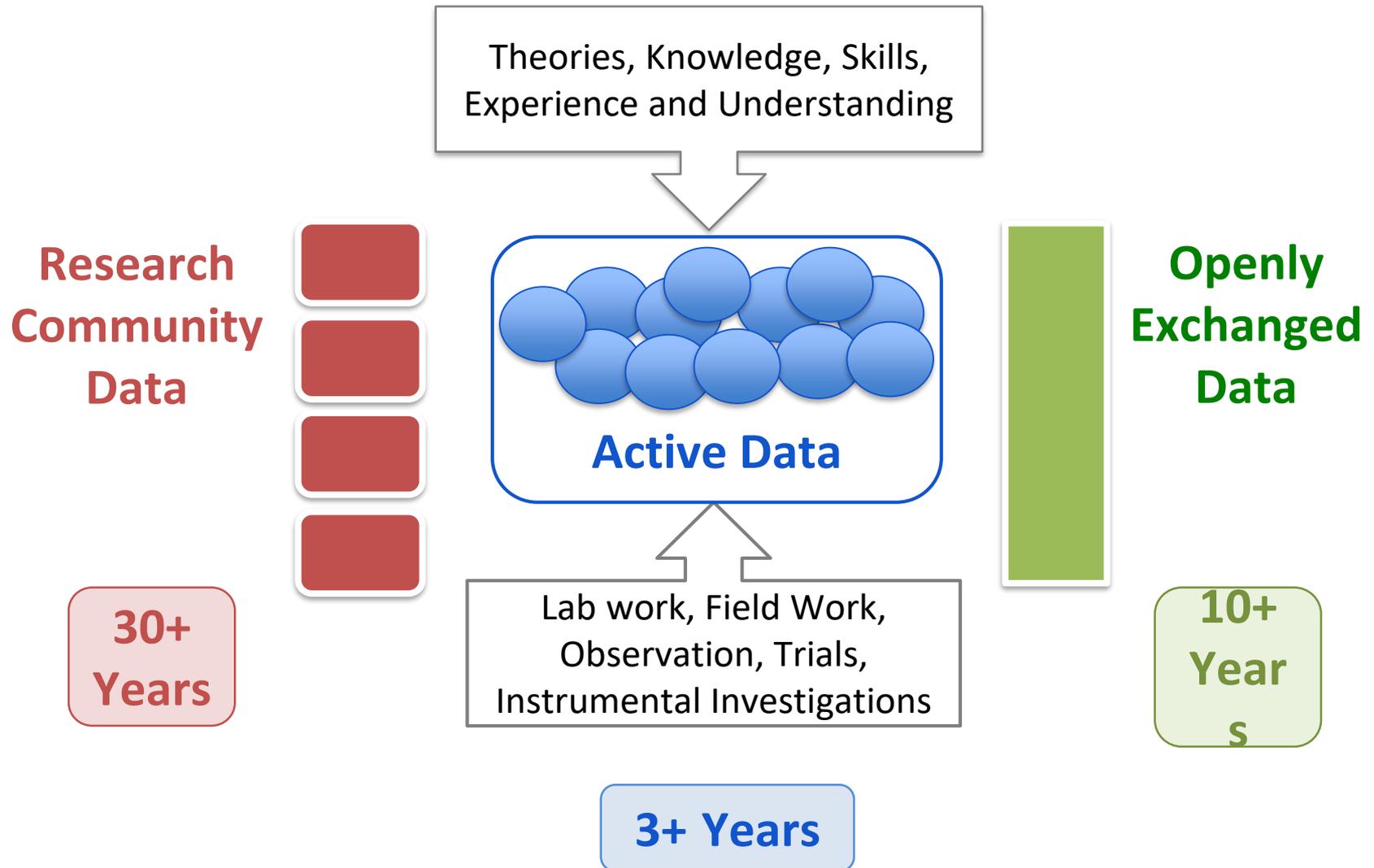
BUT

- Cost of doing so is likely to be prohibitive
- Remaking the data can be 'better' and/or 'cheaper'
- Its a zero sum game

THEREFORE

- Trade-offs and decisions need to be made

Clarifying the role of three bounded retention periods may assist develop an affordable data system design

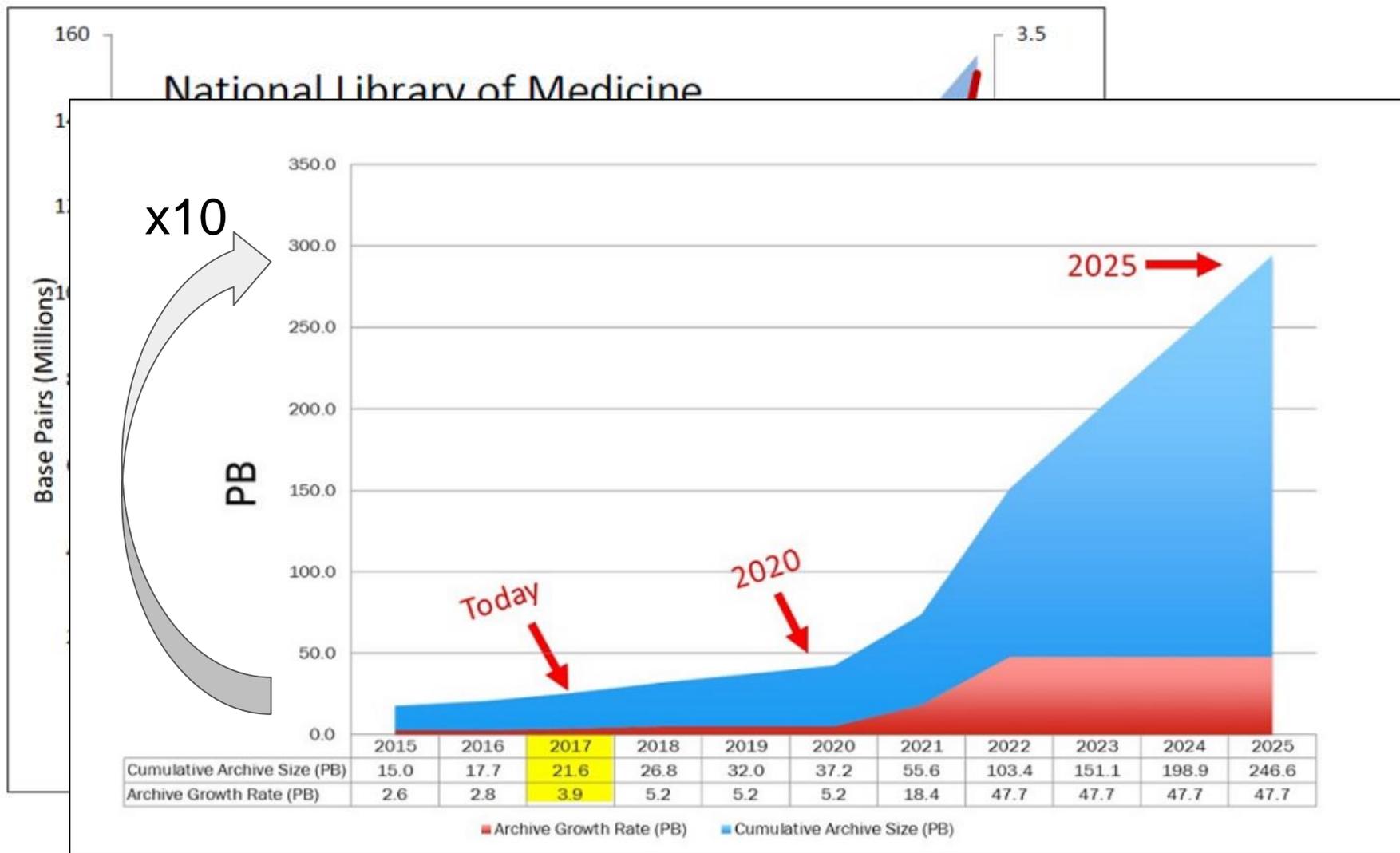


# Designing Around Three Key Timescales

	<b>Active</b>	<b>Openly exchanged</b>	<b>Referential</b>
Timescale	3+ years	10+ years	30+ years
Stakeholders	Researchers	Institutions	Research community
Cost per unit	High	Medium	Low
Curation aim	Version, Access, Use	Access, Discovery, Relate	FAIR
Curation quality	Low	Medium	High

What about forever?

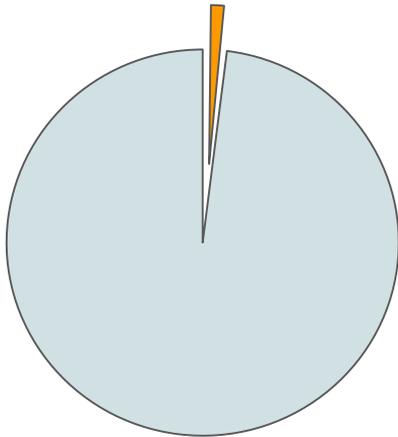
# We have all seen plenty of data growth graphs



# Spending more on data - what is the problem?

Data:

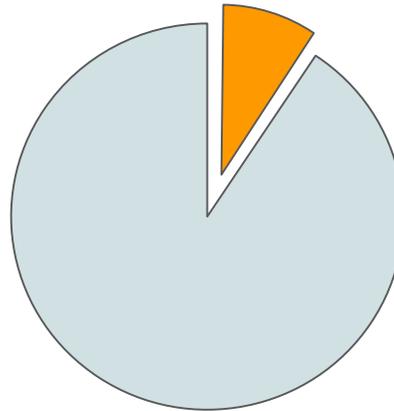
- Create
- Organise
- Use



Total Research Budget

Data:

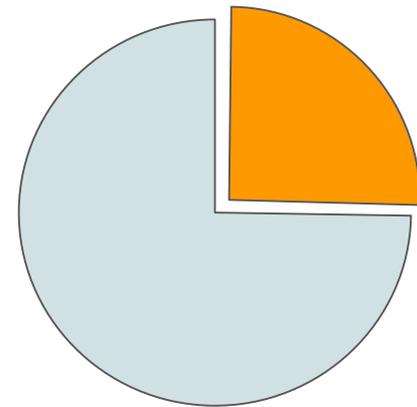
- Create
- Organise
- Use



Total Research Budget

Data:

- Create
- Organise
- Use



Total Research Budget



Time

*Something has to give - which is :  
How much data will we keep for how long?  
At what standard of curation will we keep that data?*

# Addressing the Cost of Data

Its a zero sum game

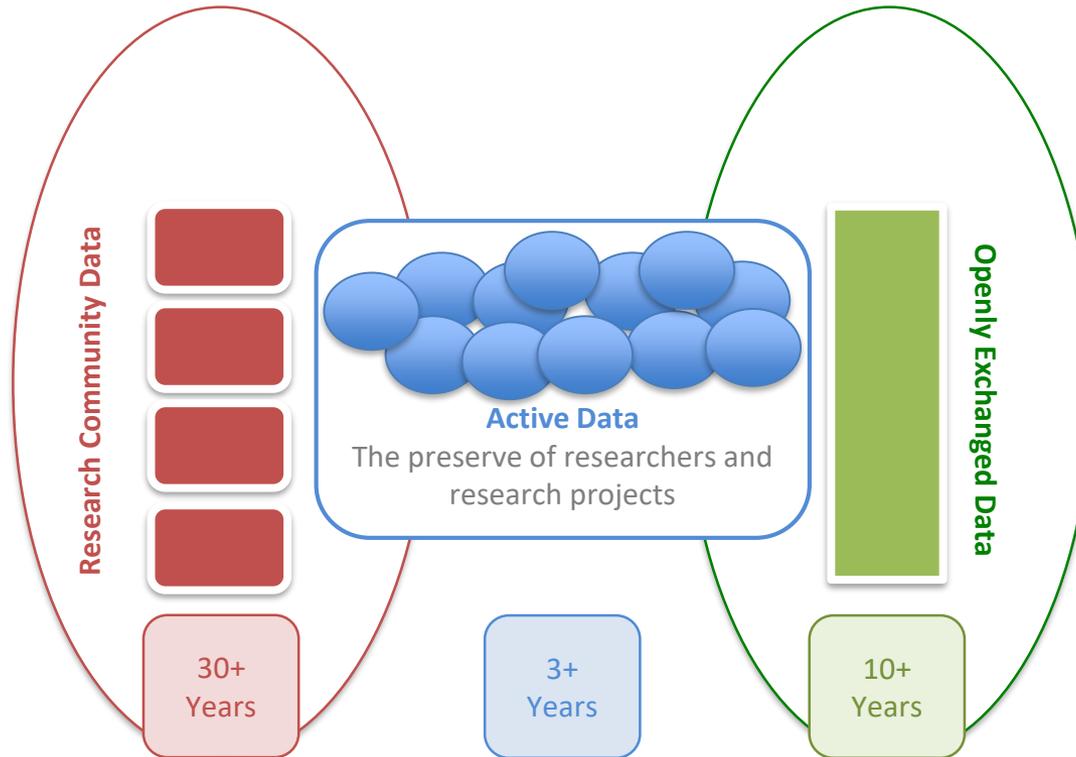
An explosion in data cost is unfundable

- Automation of all aspects of data management is essential
  - Lower the per unit cost
- Deletion of data to reduce the quantity of managed data
  - Lower the scale
- Reduce our ambitions on the quality of data curation
  - Lower the quality

HOW TO OPTIMISE?

The preserve of research communities and their social and economic beneficiaries

The preserve of research funders and research institutions



These data assets will

- relate to long term economic activity
- form an evolving component of the global stock of active knowledge
- meet domain specific requirements

These data assets will

- provide inputs for research translation and policy development
- contribute to research integrity
- support knowledge exchange

# The Benefits

- Empower decision making - data in each retention period has a life cycle managed by stewards responsible for that period
- Permit data disposal - the stakeholders involved in each retention period determine the admitted data
- Clarify admission to higher cost services and infrastructures

Not all data commences life in research projects. Some data is always destined for research community curation and is best created with that purpose in mind.

# Challenges

Sharing published and other completed research data across research teams.

- How is this selection and sharing accomplished?
- Does the researcher's institution accept the cost?

Data valuable across a research community involves a community agreed processes.

- How do communities conduct such a process?
- To what extent can the application of FAIR principles be automated?
- How do research communities fund the necessary storage?

Some research data is part of the global stock of knowledge.

- Is this funded by governments and other central funding agencies because of its value to society?
- How should data in this state be identified?

# What if Research Data were Not Forever?

Rob Cook & Rhys Francis

THANK YOU !