



Australian Research Data Commons

Data Retention Project

Building a Foundation for Impact

J Max WILKINSON
ARDC Data Infrastructure Architect

eResearch NZ 2021



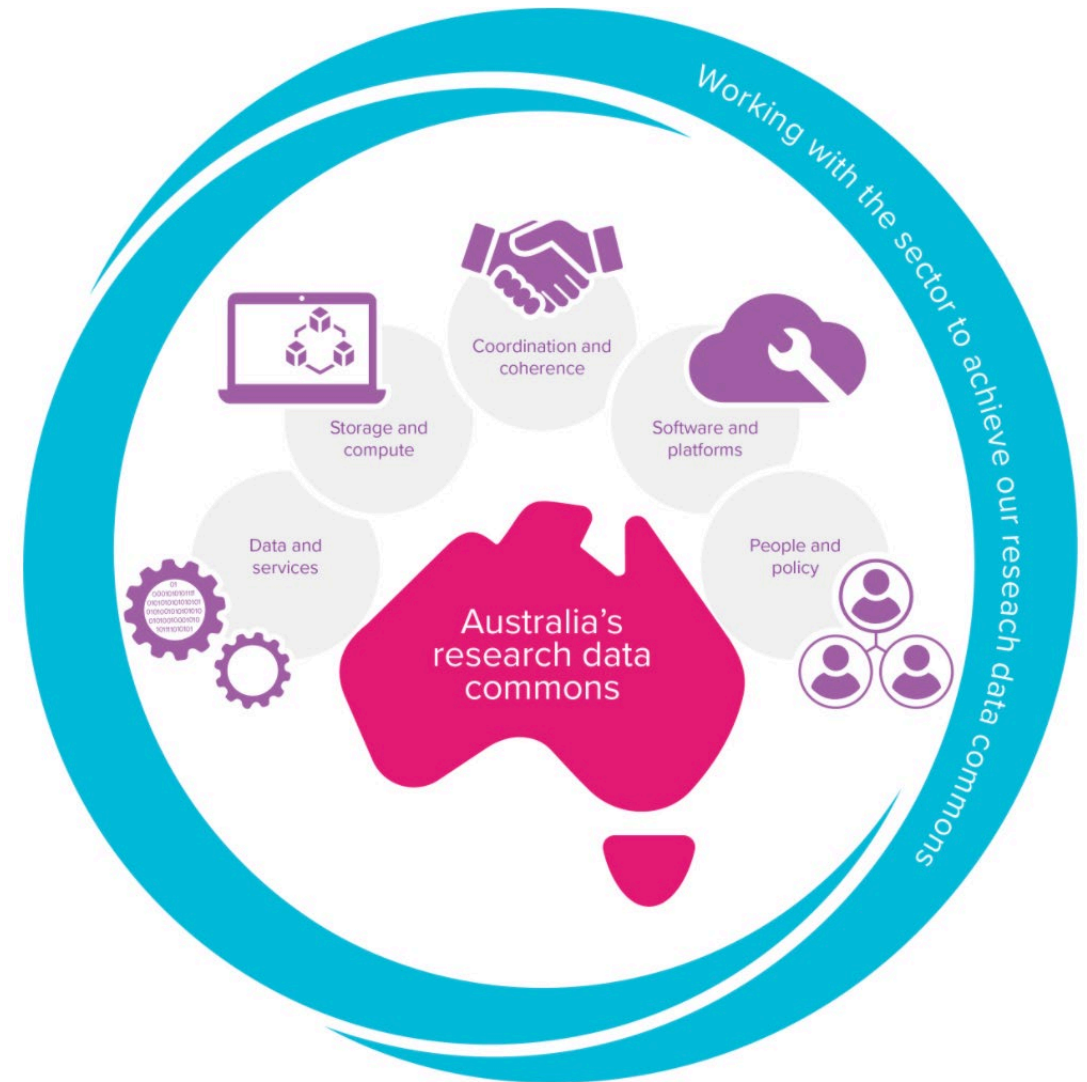
ARDC Themes

Purpose:

to provide Australian researchers with competitive advantage through data.

Mission:

to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.



Australian Research Data Commons

NCRIS

National Research
Infrastructure for Australia

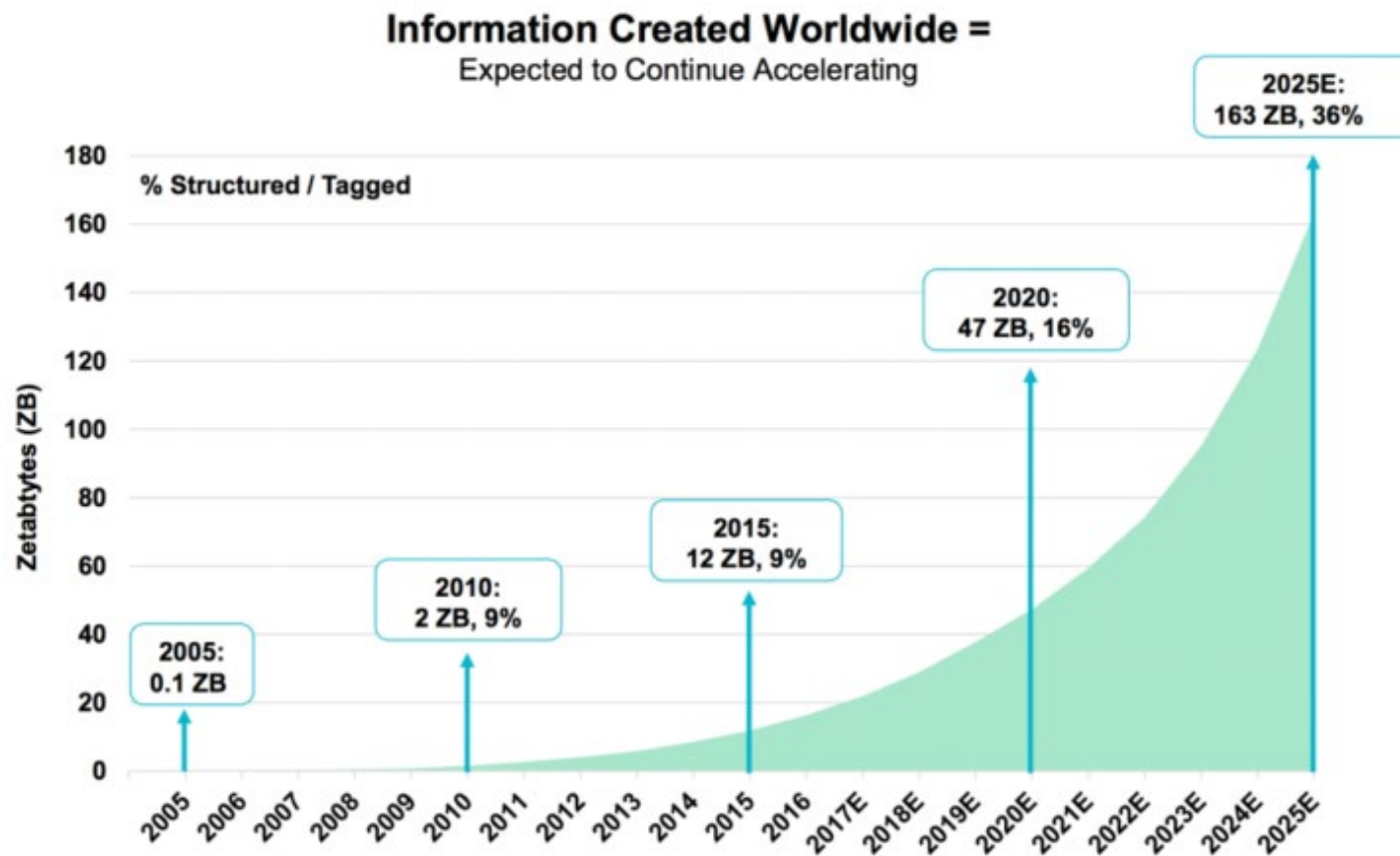
An Australian Government Initiative

ARDC Themes

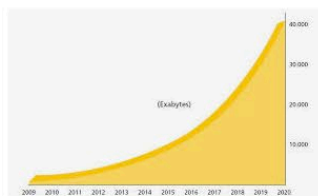


The Challenge (part1)

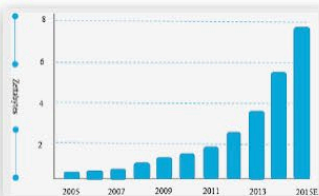
Data Volume Growth Continues @ Rapid Clip...
% Structured / Tagged (~10%) Rising Fast...



The Challenge (part1)



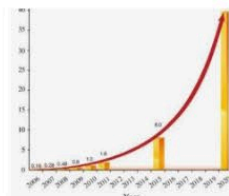
The exponential data growth estimated ...
researchgate.net



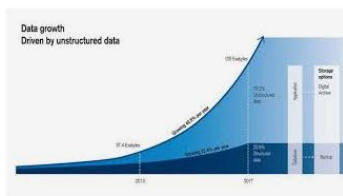
rapid growth rate of data in Zettabytes ...
researchgate.net



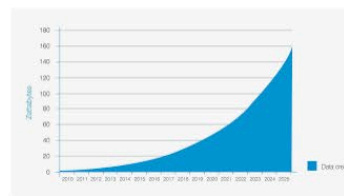
Data growth and expansion (IDC, 2009 ...
researchgate.net



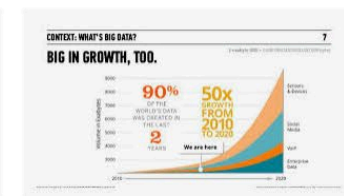
Global growth trend of data volume ...
researchgate.net



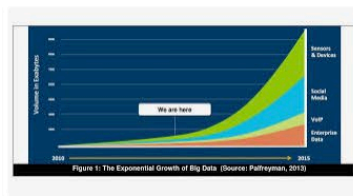
Everything a Data Scientist Should Know ...
kdnuggets.com



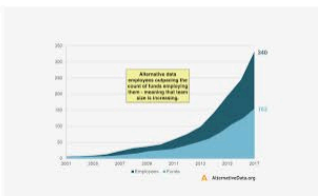
Forecast of exponential growth of ...
reddit.com



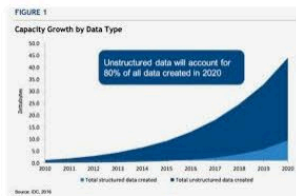
Rise of the Data Warehouse | Avora
avora.com



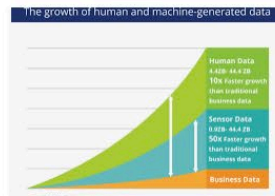
Data Analytics: Concepts, Technologies ...
semanticscholar.org



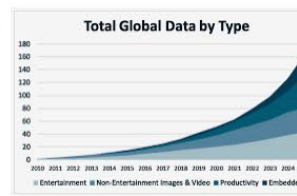
Buy-side Alternative Data Employee ...
alternativedata.org



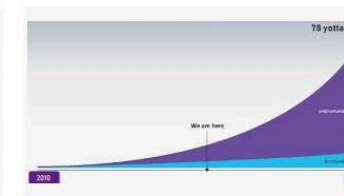
Industry Verticals Tackle Unstructured Data
kevinjackson.blogspot.com



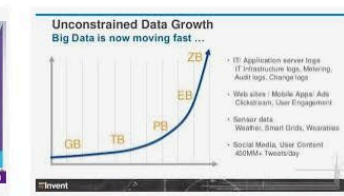
IoT, Big Data and AI – the New ...
business2community.com



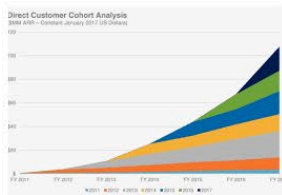
The Data Deluge - Drowning in Data ...
uncommonlogic.com



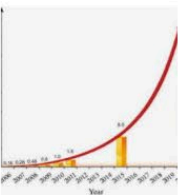
Introduction to BIG DATA: What is ...
bigdatapath.wordpress.com



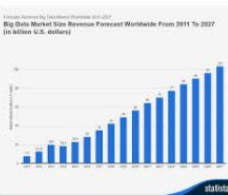
big data growth - Google 搜尋 | Big ...
pinterest.com



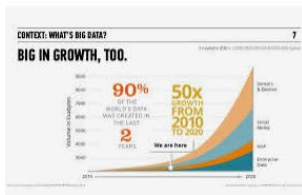
MongoDB: Riding The Data Wave (NASDAQ ...
seekingalpha.com



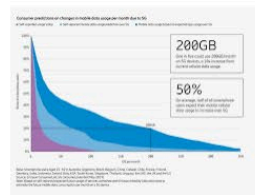
Global growth trend of data...
researchgate.net



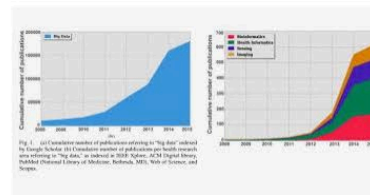
10 Charts That Will Change Your ...
forbes.com



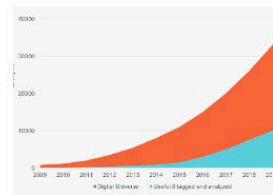
Ensure Business Growth via Big Data ...
promptcloud.com



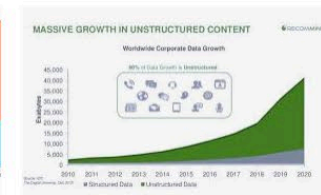
How Much Will 5G Data Usage Increase ...
spectrumblog.blogspot.com



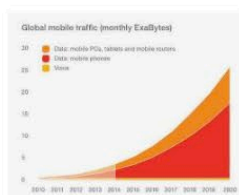
Healthcare Big Data Analytics
healthitanalytics.com



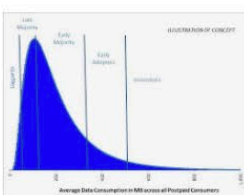
Data growth between 2009 and 2020 ...
researchgate.net



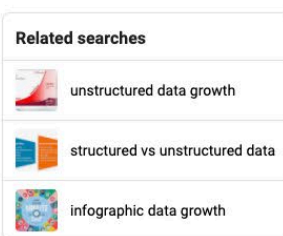
The Massive Growth in Unstructured Data ...
researchgate.net



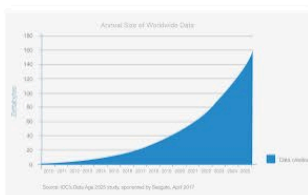
Global Mobile Data Traffic 2010–202...
whatsthebigdata.com



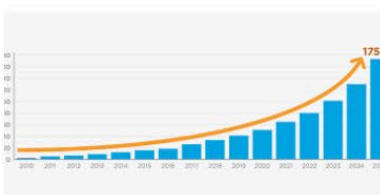
Mobile Data Growth ... The Perfect Stor...
technoeconomyblog.com



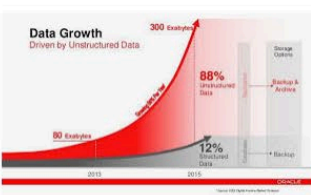
Orchestrating Enterprise Data with Data ...
virtustream.com



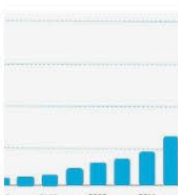
Orchestrating Enterprise Data with Data ...
virtustream.com



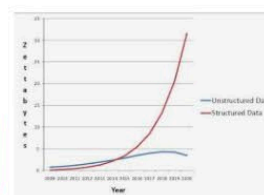
Big data overview | AP CSP (article ...
khanacademy.org



data growth driven by unstructured ...
pinterest.com

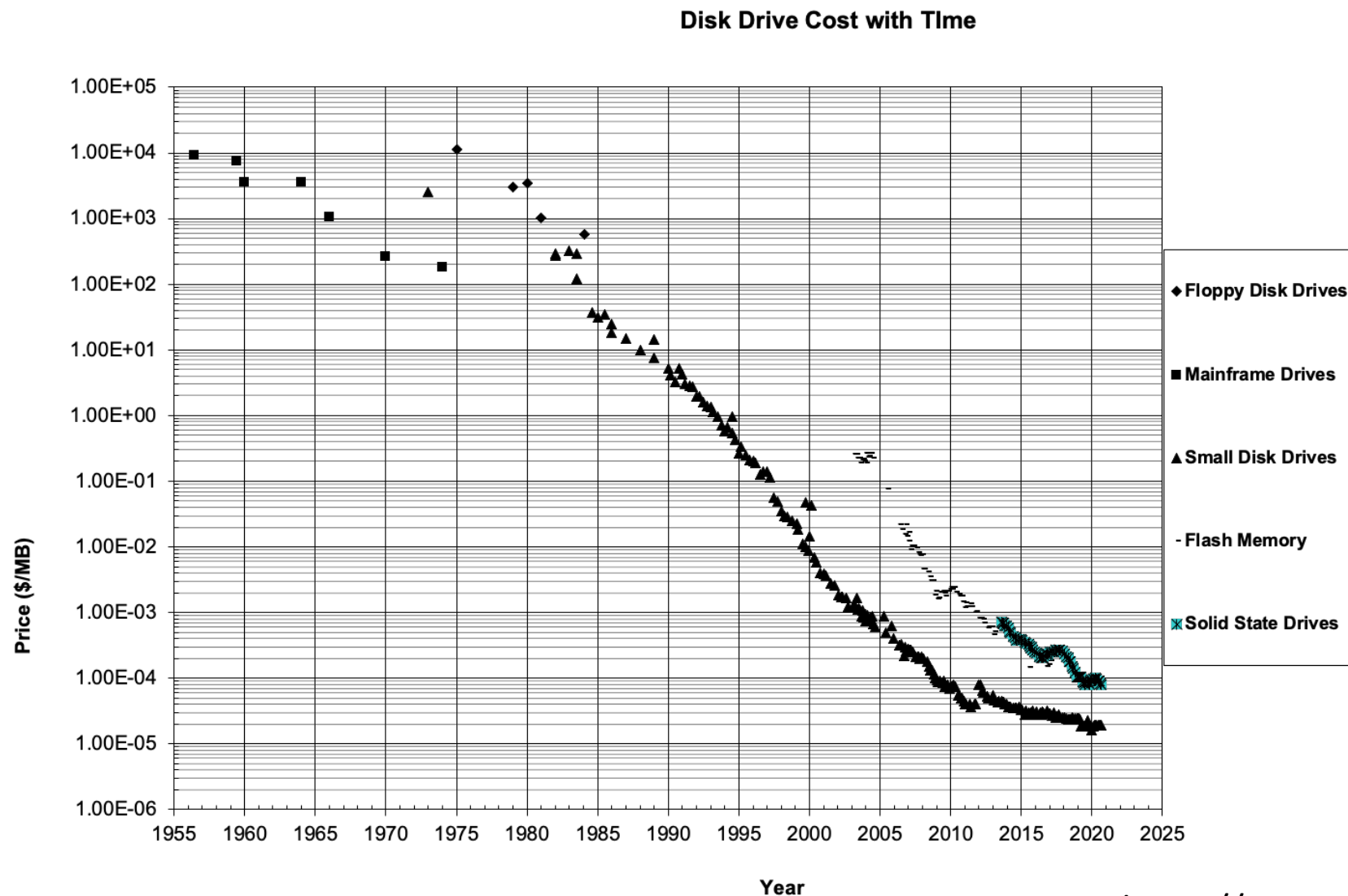


rapid growth rate of data in Z...
researchgate.net

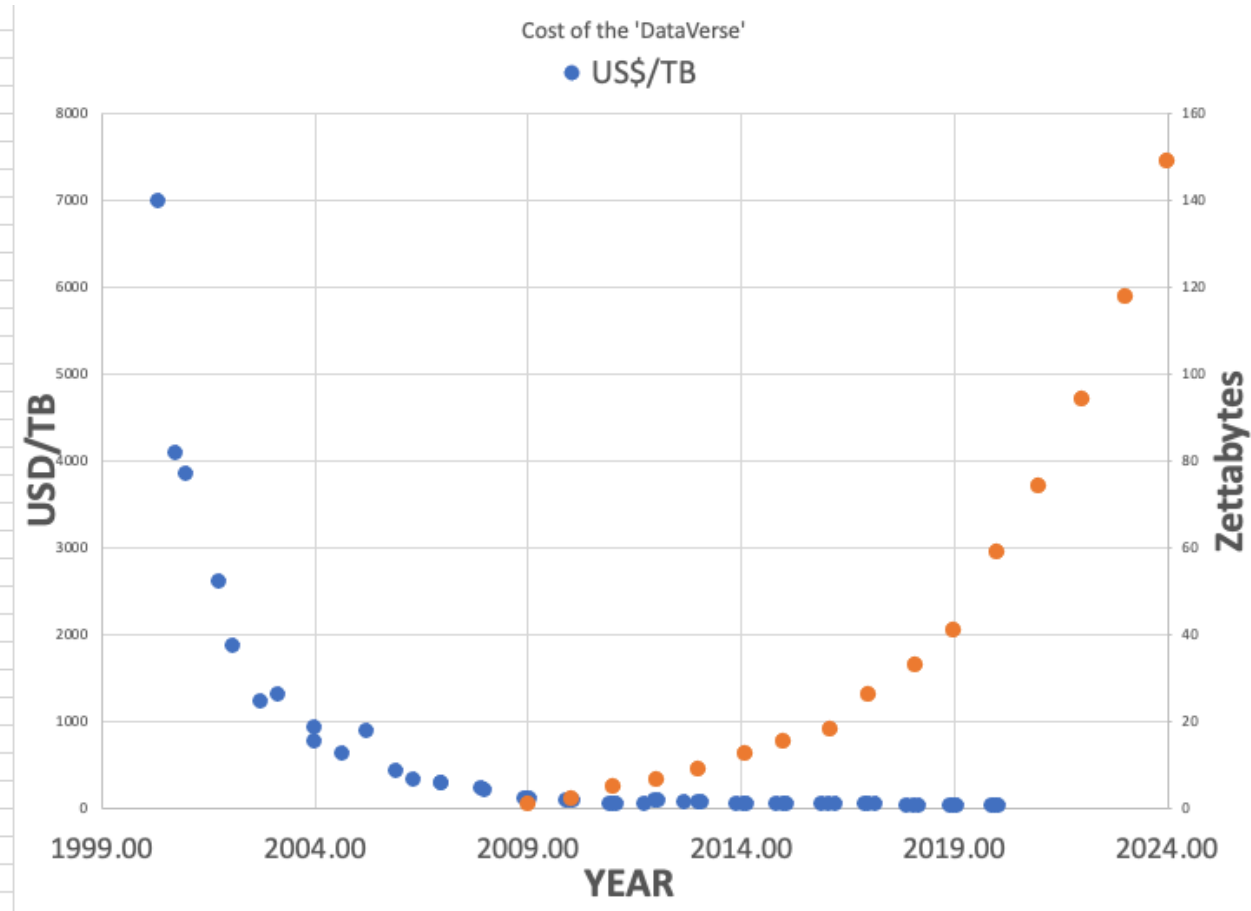
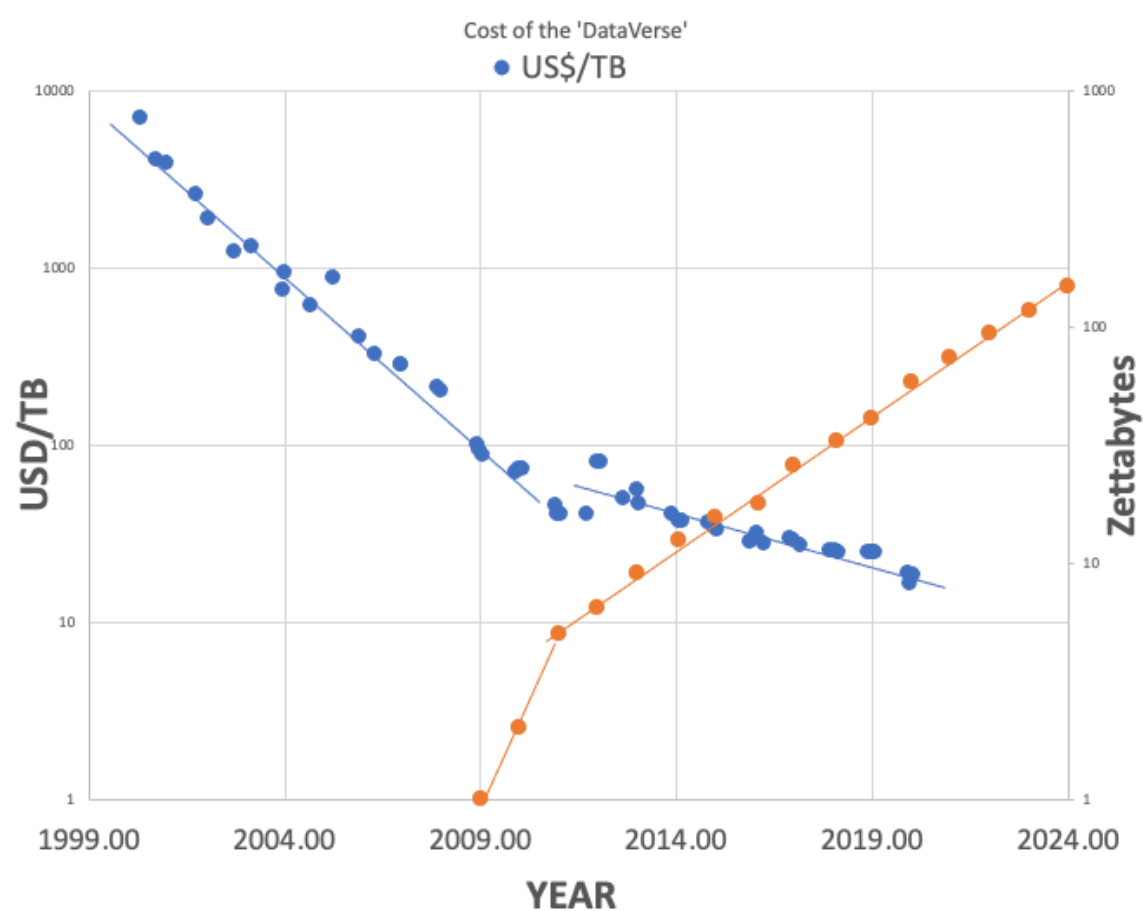


structured vs unstructured data growth ...
tomkendig.wordpress.com

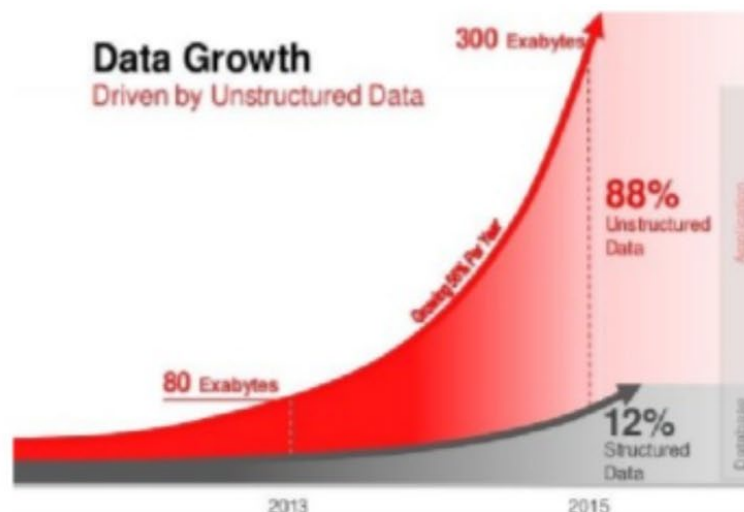
The Challenge (part 2)



The Challenge (part 2)



The Challenge (all is not lost- characterisation)



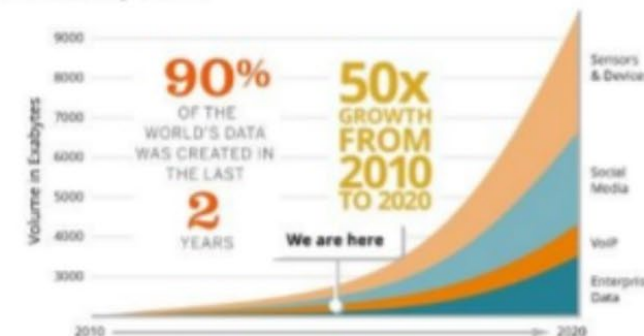
<https://www.linkedin.com/pulse/world-today-data-rich-information-poon-gurup-mohapatra-pmp/>

CONTEXT: WHAT'S BIG DATA?

7

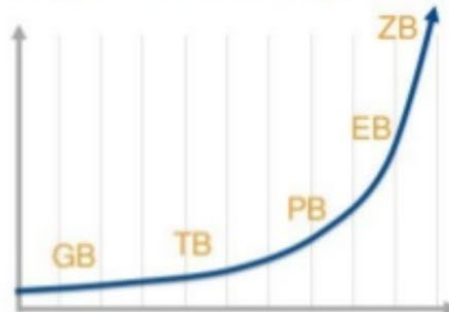
BIG IN GROWTH, TOO.

1 exabyte (EB) = 1,000,000,000,000,000 bytes



<https://www.promptcloud.com>

Unconstrained Data Growth
Big Data is now moving fast ...



- IT/ Application server logs
IT Infrastructure logs, Metering, Audit logs, Change logs
- Web sites / Mobile Apps/ Ads
Clickstream, User Engagement
- Sensor data
Weather, Smart Grids, Wearables
- Social Media, User Content
450MM+ Tweets/day

Invent

<https://www.signiant.com>



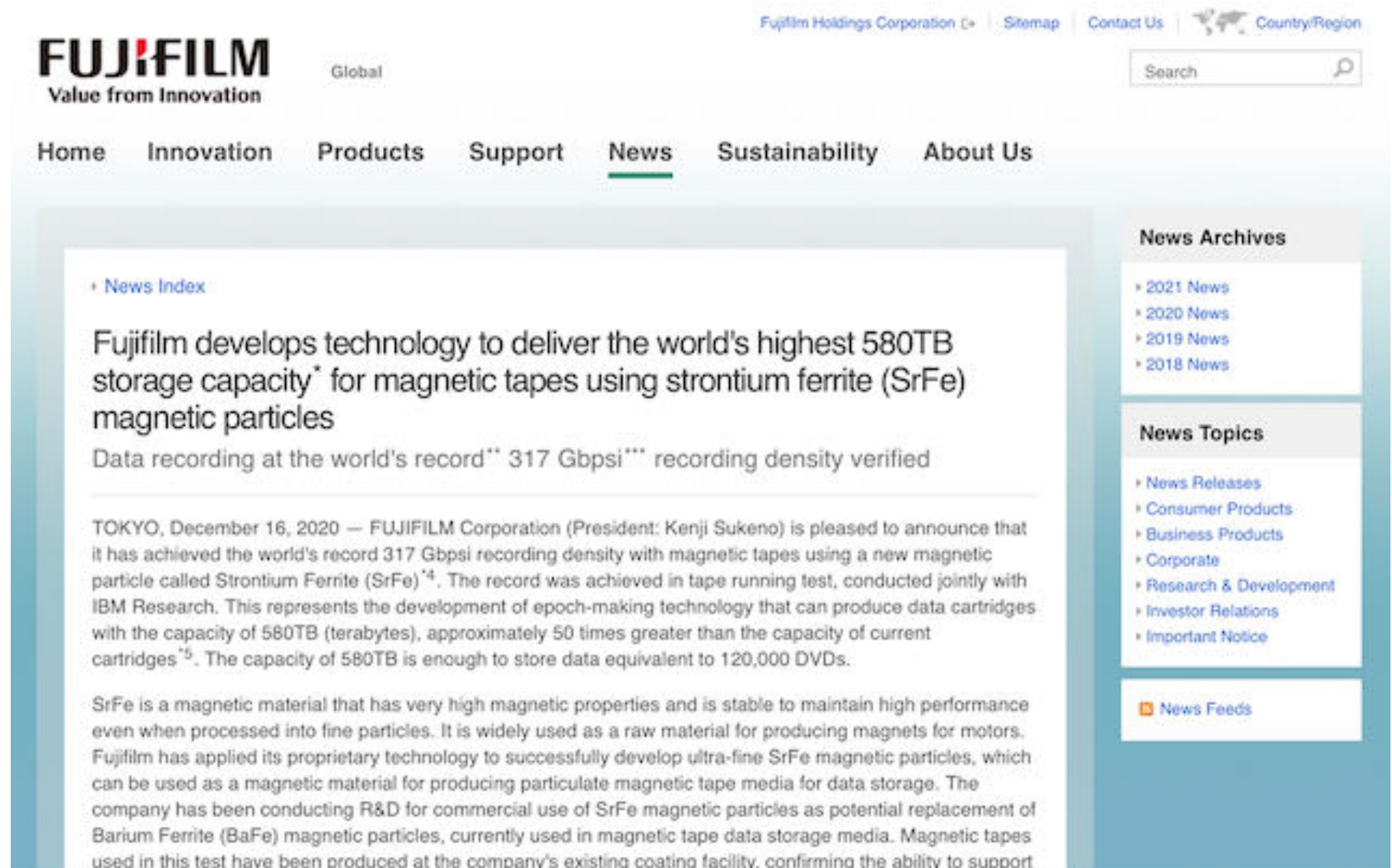
<https://john-popelaars.blogspot.com>



Australian Research Data Commons



The Challenge (all is not lost- technology)



The screenshot shows the Fujifilm Global website's News section. The header includes the Fujifilm logo with the tagline 'Value from Innovation', a 'Global' link, and navigation links for 'Fujifilm Holdings Corporation', 'Sitemap', 'Contact Us', and 'Country/Region'. A search bar is also present. The main navigation menu includes 'Home', 'Innovation', 'Products', 'Support', 'News' (which is underlined), 'Sustainability', and 'About Us'. The main content area features a news article titled 'Fujifilm develops technology to deliver the world's highest 580TB storage capacity* for magnetic tapes using strontium ferrite (SrFe) magnetic particles'. The article includes a sub-headline 'Data recording at the world's record** 317 Gbps*** recording density verified' and a detailed paragraph about the achievement on December 16, 2020. A sidebar on the right contains 'News Archives' (2021, 2020, 2019, 2018), 'News Topics' (News Releases, Consumer Products, Business Products, Corporate, Research & Development, Investor Relations, Important Notice), and 'News Feeds'.

FUJIFILM
Value from Innovation

Global

[Fujifilm Holdings Corporation](#) [Sitemap](#) [Contact Us](#) [Country/Region](#)

Search

[Home](#) [Innovation](#) [Products](#) [Support](#) [News](#) [Sustainability](#) [About Us](#)

[News Index](#)

Fujifilm develops technology to deliver the world's highest 580TB storage capacity* for magnetic tapes using strontium ferrite (SrFe) magnetic particles

Data recording at the world's record** 317 Gbps*** recording density verified

TOKYO, December 16, 2020 — FUJIFILM Corporation (President: Kenji Sueno) is pleased to announce that it has achieved the world's record 317 Gbps recording density with magnetic tapes using a new magnetic particle called Strontium Ferrite (SrFe)^{*4}. The record was achieved in tape running test, conducted jointly with IBM Research. This represents the development of epoch-making technology that can produce data cartridges with the capacity of 580TB (terabytes), approximately 50 times greater than the capacity of current cartridges^{*5}. The capacity of 580TB is enough to store data equivalent to 120,000 DVDs.

SrFe is a magnetic material that has very high magnetic properties and is stable to maintain high performance even when processed into fine particles. It is widely used as a raw material for producing magnets for motors. Fujifilm has applied its proprietary technology to successfully develop ultra-fine SrFe magnetic particles, which can be used as a magnetic material for producing particulate magnetic tape media for data storage. The company has been conducting R&D for commercial use of SrFe magnetic particles as potential replacement of Barium Ferrite (BaFe) magnetic particles, currently used in magnetic tape data storage media. Magnetic tapes used in this test have been produced at the company's existing coating facility, confirming the ability to support

News Archives

- [2021 News](#)
- [2020 News](#)
- [2019 News](#)
- [2018 News](#)

News Topics

- [News Releases](#)
- [Consumer Products](#)
- [Business Products](#)
- [Corporate](#)
- [Research & Development](#)
- [Investor Relations](#)
- [Important Notice](#)

[News Feeds](#)

Data Retention Project Strategy

How to get...

Coherent business intelligence

&

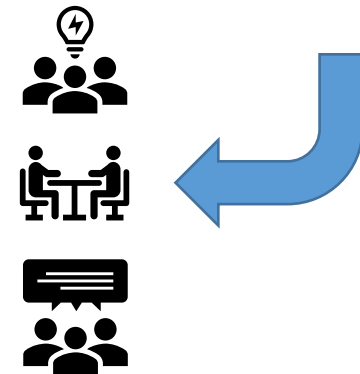
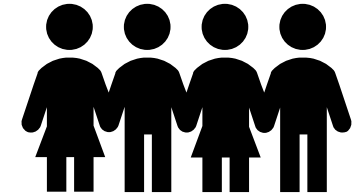
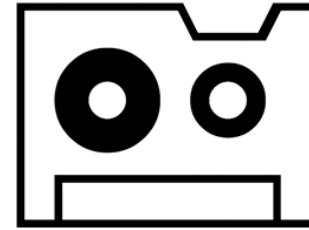
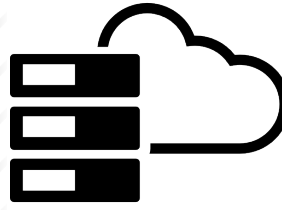
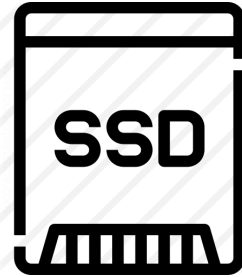
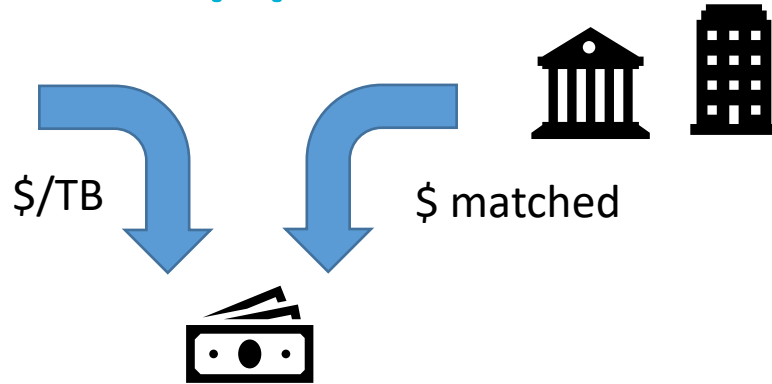
Consistent characterisation of content

when....

Fragmentation remains across a) providers and b) practice



Project Approach



Areas of Focus

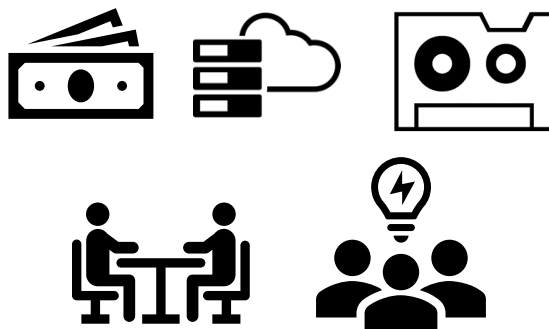
Business pressures

Realistic & Sustainable
investment strategies



Technology Pressures

De-Risk innovation in
the solution space

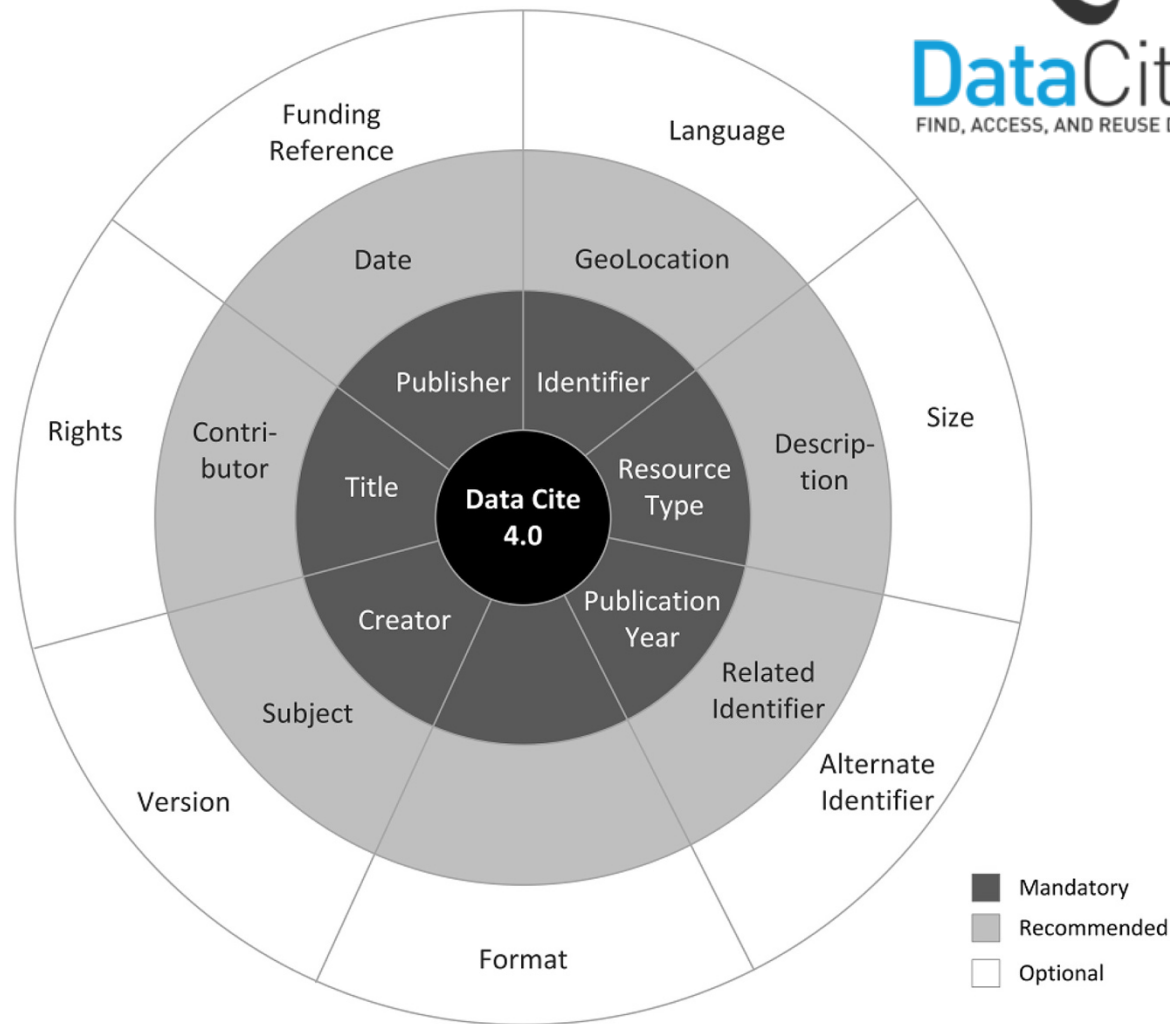


Cultural Pressures

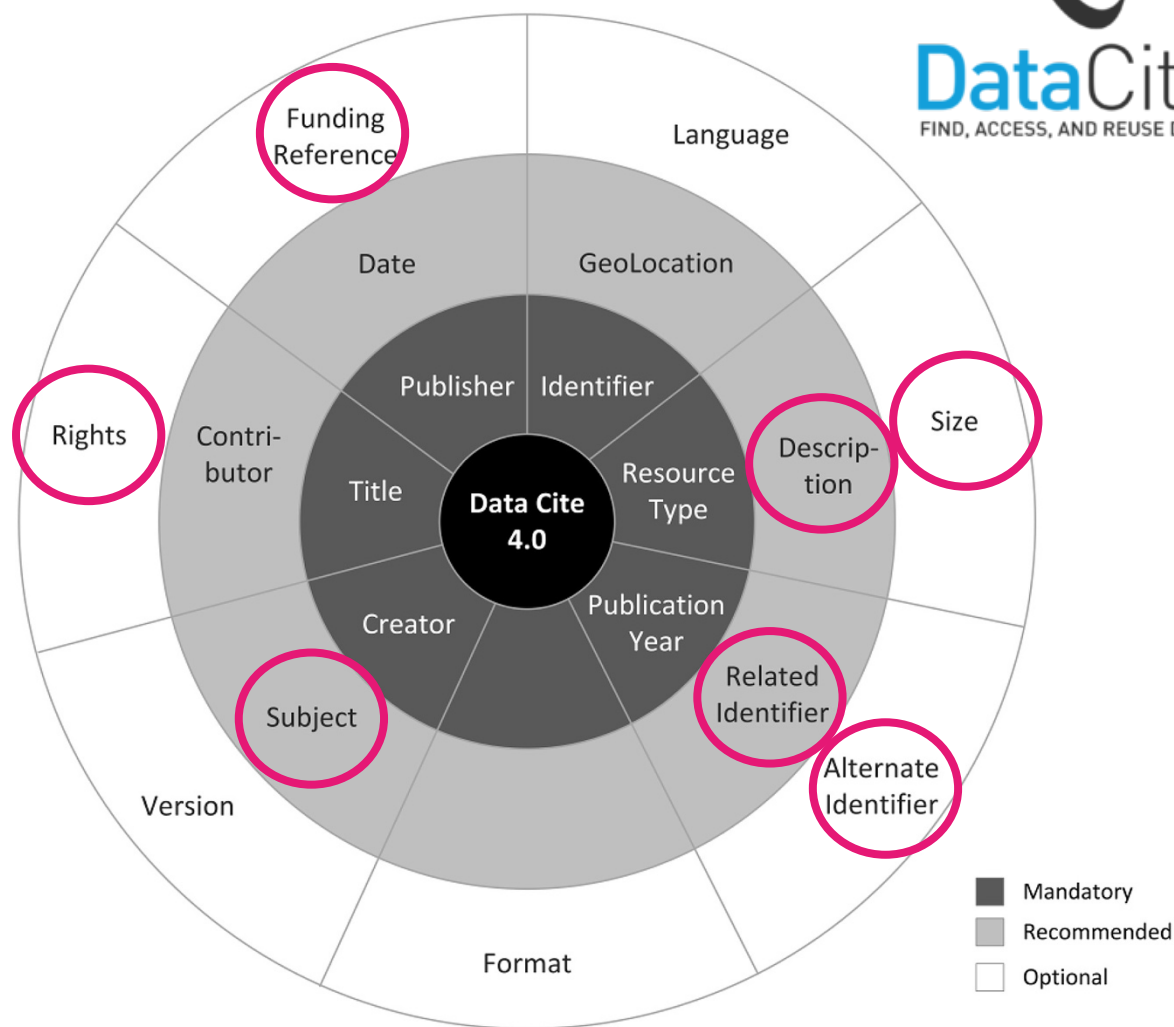
Achievable RDM
practice



RDM metadata



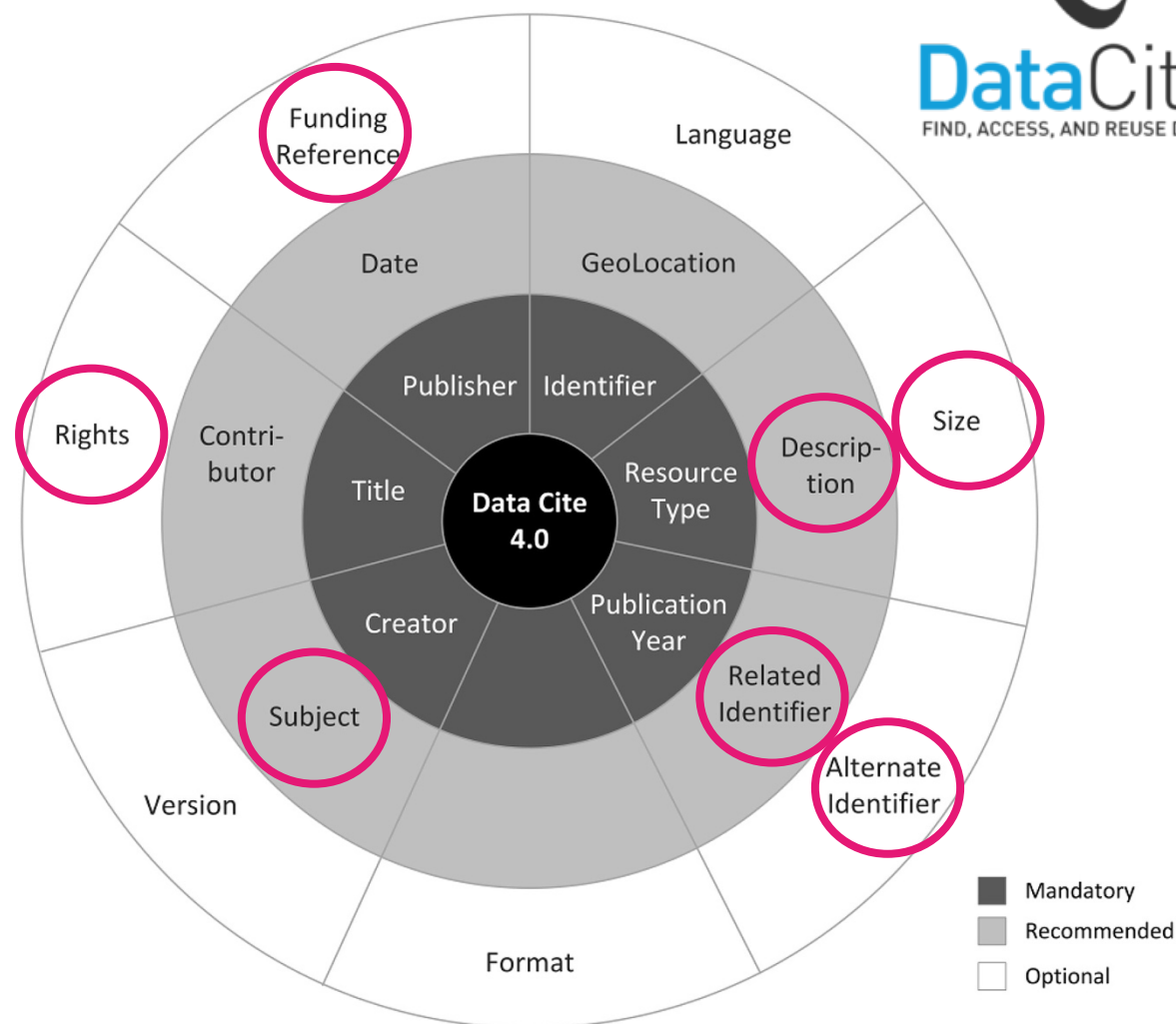
RDM metadata



DataCite	ARDC DRP
Resource Type	Collection type
Creator	Owner
Title	Title
Description	Description
Size	Capacity
Publisher	Controller
Related Identifier	URL
Subject	Discipline/s
Rights	License
Funding Ref	Merit
Year	Year
Alternate ID	Collection ID
Identifier	DOI



RDM metadata



DataCite	ARDC DRP
Resource Type	Collection type
Creator	Owner
Title	Title
Description	Description
Size	Capacity
Publisher	Controller
Related Identifier	URL
Subject	Discipline/s
Rights	License
Funding Ref	Merit
Year	Year
Alternate ID	Collection ID
Identifier	DOI



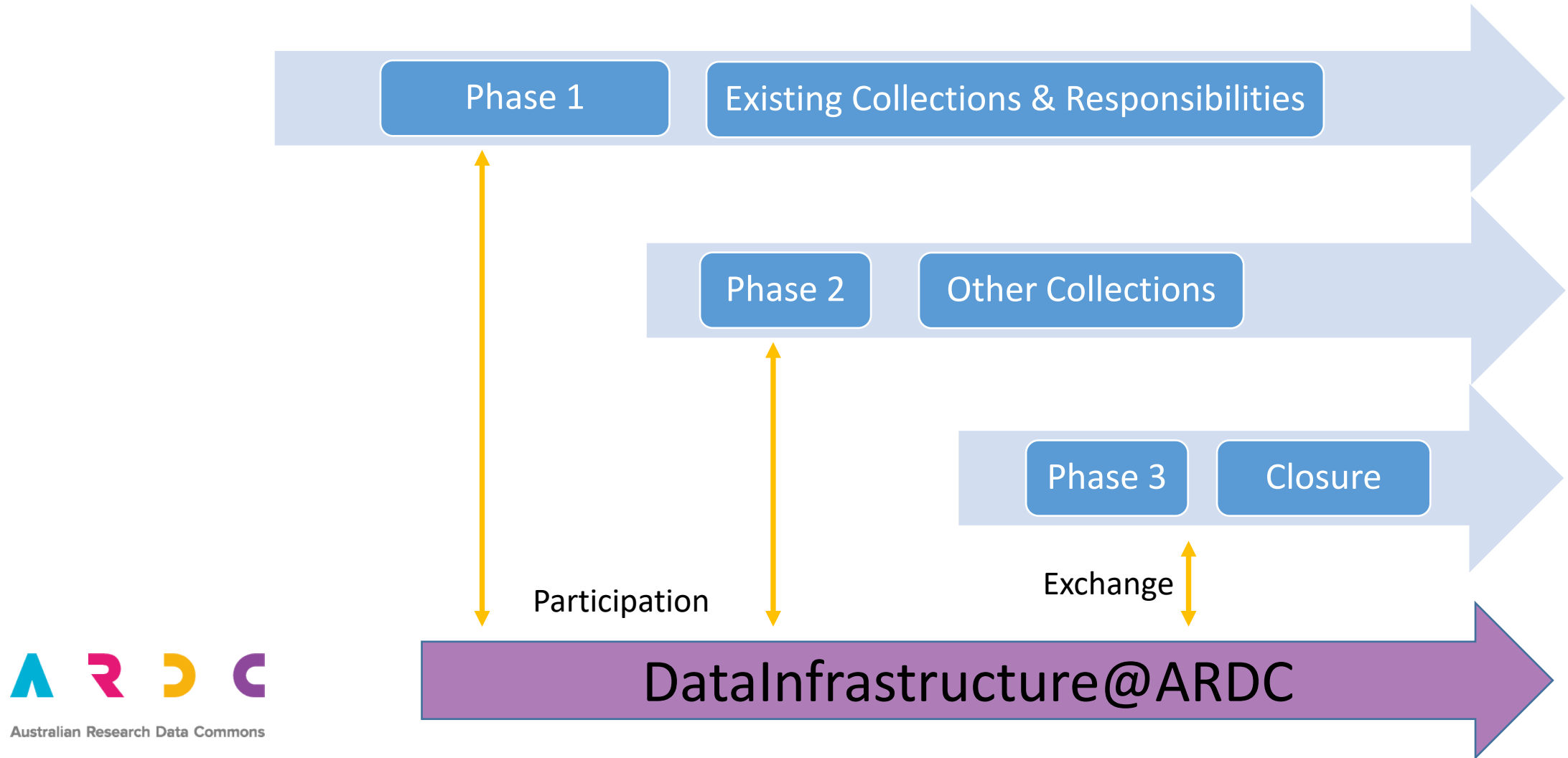
Phased Approach

2020

2021

2022

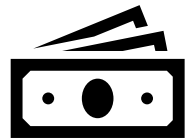
2023



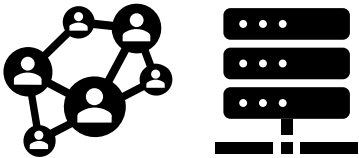
Outcomes



Coherent and consistent intelligence on research data content



Accurate strategy to plan investment



Together more sustainable models for significant and important research data



Australian Research Data Commons

max.wilkinson@ardc.edu.au

Thank you

The Challenge (summary)

Data growth accelerating

Cost decrease decelerating

“To be an effective and sustainable service, research data architectures require active, accurate and informed business planning”



Metadata Key Points

- Third party controlled schema
- Independent PIDs
- We require more than mandatory
 - Intelligence on content
- ...but less than complete
 - which is often best managed by creators/curators anyway
- Transaction returning a DOI
 - supporting data citation

The Challenge (part 2)

