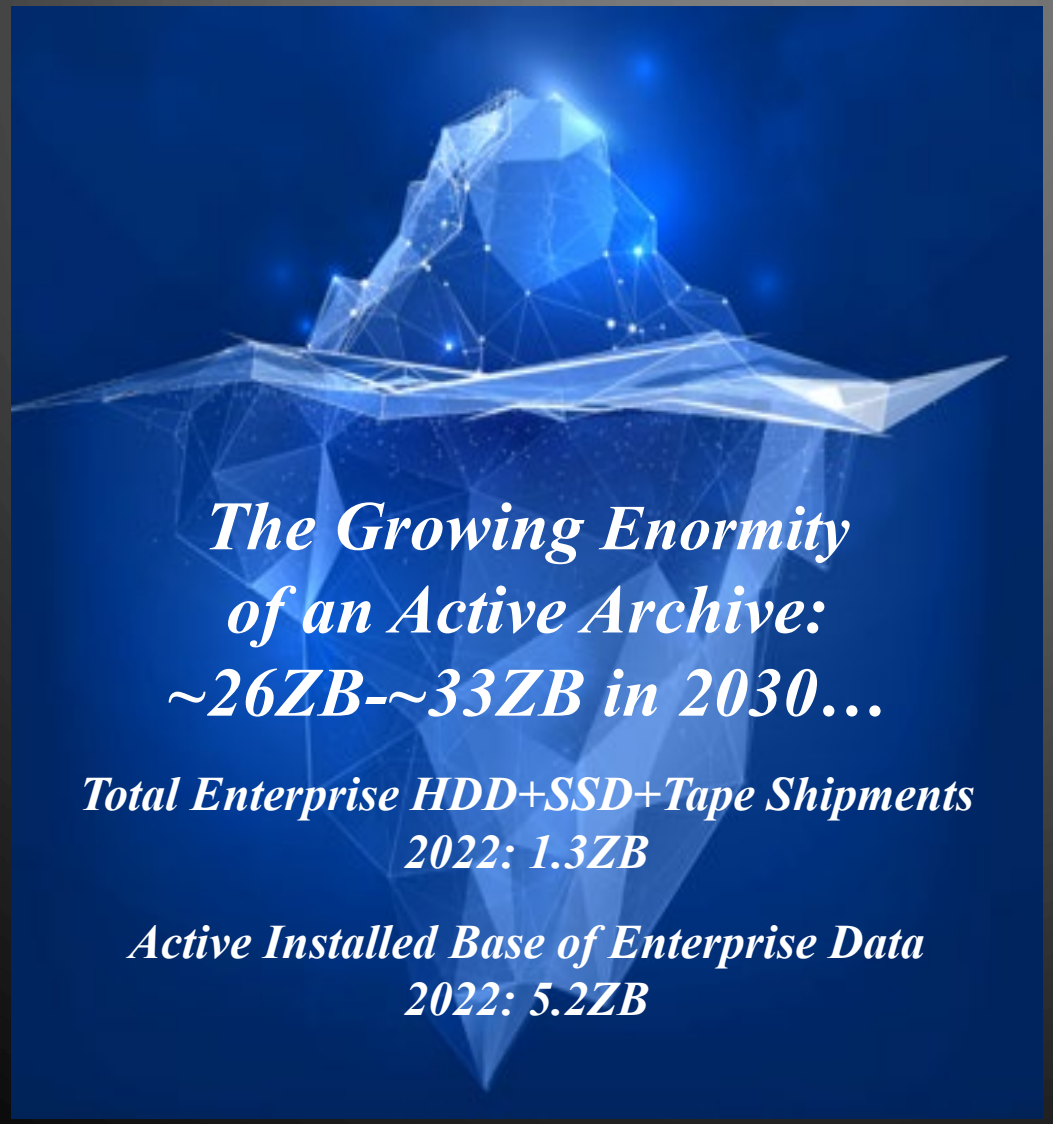
PRESERVATION OR DELETION: ARCHIVING AND ACCESSING THE DATAVERSE



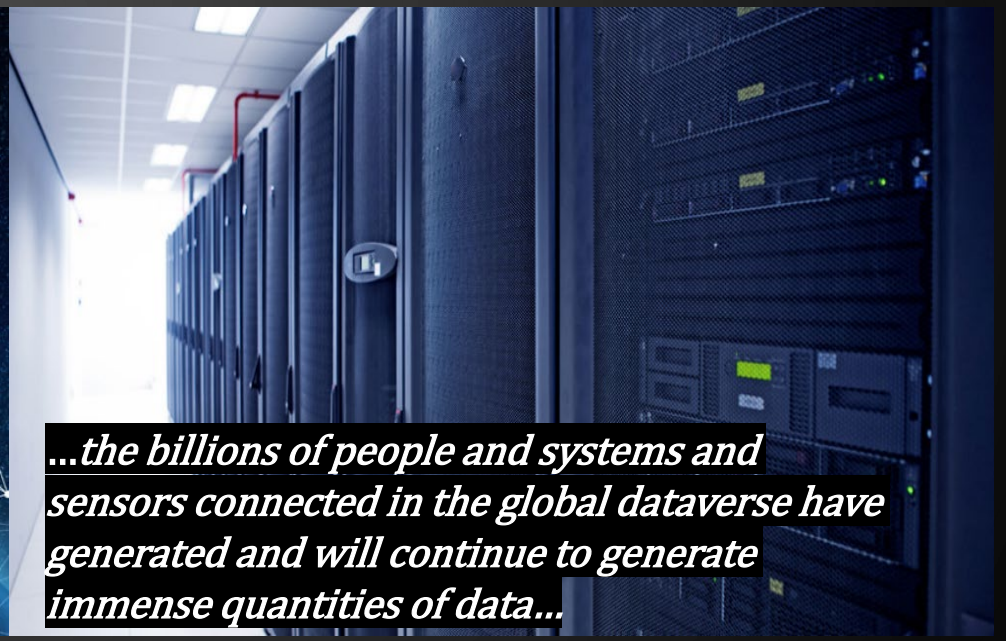
- Most of the data in the cold/frozen layers may be JIC (Just in Case) or WORN (Write Once Read Never), which may never be accessed at all—*nor in most cases, will it ever be deleted…*

Library of Congress
DSA Conference, March 27, 2023

John Monroe
torstones@gmail.com
Mobile Office: +1 (408) 828-3967
www.Furthurdata.com

Furthur
Market Research

THE EXPANDING DATAVERSE



...the billions of people and systems and sensors connected in the global dataverse have generated and will continue to generate immense quantities of data...

- The enterprise "data pools" of the early 2000s became "data lakes" by 2010 and grew in recent years to become "data oceans" which have already begun to morph into a vast multiform "dataverse." And because "data is the new oil," we are loathe to delete any data.
- Increased storage at any point in the World Wide Web—bear in mind that a mobile phone is a point in the Web—increases the possibilities for storage in every part of the Web.
- We are only beginning to see the enormous implications of that simple fact.

AGENDA

- Shipment History and New Forecasts
- Fresh Delineations of Enterprise Data
- Recent Surveys
- Inconclusive Conclusions

Enduring Question: *Will the Past be Prologue, or Will History Be Bunk?*

Note: My forecasts are always devised with these precautionary adages in mind:

- *The only thing we know with certainty about any forecast is that it will be wrong. — Anonymous*
- *He who foretells the future lies, even if he tells the truth. —Arab Proverb*

2010-2022 HISTORICAL SHIPMENTS AND FORECAST OF ENTERPRISE PETABYTES DELIVERED 2023-2030 WITH ACTIVE INSTALLED BASE ESTIMATES

	2010	2015	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	"CAGR 2023-2030"
Enterprise SSD PB	187	26,154	130,766	178,972	186,588	230,664	334,195	498,792	690,849	925,047	1,098,030	1,604,222	2,050,196	36.6
Annual Growth %	–	223.9	64.1	36.9	4.3	23.6	44.9	49.3	38.5	33.9	18.7	46.1	27.8	
Enterprise HDD PB	45,216	157,093	679,887	959,011	941,749	1,251,584	1,610,789	2,177,786	2,735,300	3,605,125	4,589,324	6,021,194	8,062,378	30.5
Annual Growth %	–	34.7	39.9	41.1	–1.8	32.9	28.7	35.2	25.6	31.8	27.3	31.2	33.9	
Enterprise Tape PB	30,208	98,432	136,119	189,938	206,842	259,794	303,440	366,252	464,773	574,460	763,457	966,537	1,258,431	25.3
Annual Growth %	–	15.0	1.4	39.5	8.9	25.6	16.8	20.7	26.9	23.6	32.9	26.6	30.2	
Total Compressed Shipments PB	75,611	281,679	946,772	1,327,921	1,335,179	1,742,043	2,248,424	3,042,830	3,890,922	5,104,631	6,450,812	8,591,952	11,371,005	30.7
Annual Growth %	–	33.9	33.6	40.3	0.5	30.5	29.1	35.3	27.9	31.2	26.4	33.2	32.3	
Active Installed Base PB	91,000	819,949	2,923,201	3,950,945	5,232,405	6,447,587	7,985,007	10,081,065	12,644,067	16,413,518	21,122,288	27,465,816	35,793,990	

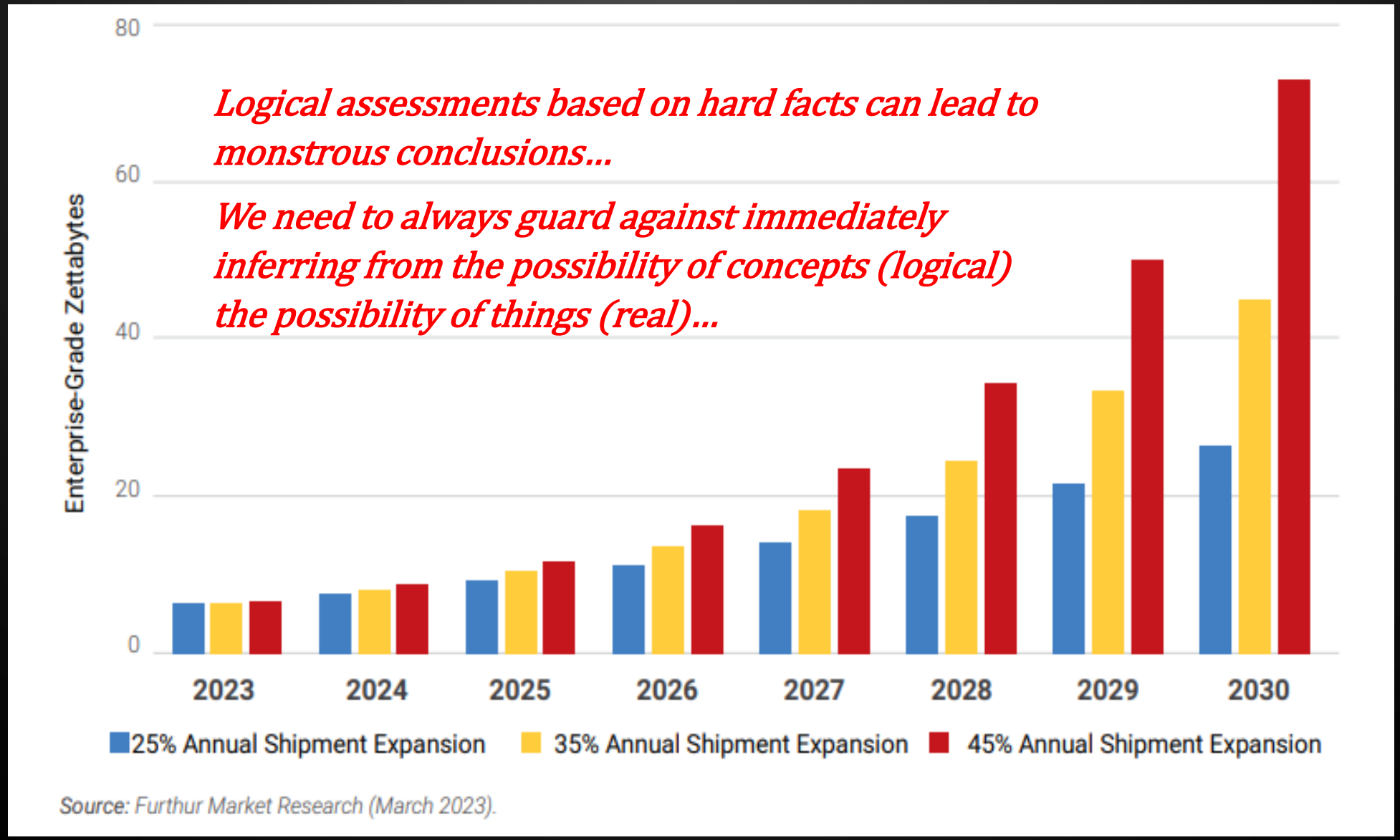
Source: Furthur Market Research (March 2023)

ALTERNATE 2023-2030 GROWTH SCENARIOS

	2023	2024	2025	2026	2027	2028	2029	2030
Alternate 2023-2030 Shipment Scenarios								
Total Shipped Enterprise PB Expanding at 25%/Year 2023-2030	1,668,974	2,086,218	2,607,772	3,259,716	4,074,644	5,093,306	6,366,632	7,958,290
Total Shipped Enterprise PB Expanding at 35%/Year 2023-2030	1,802,492	2,433,365	3,285,042	4,434,807	5,986,990	8,082,436	10,911,288	14,730,239
Total Shipped Enterprise PB Expanding at 45%/Year 2023-2030	1,936,010	2,807,215	4,070,462	5,902,169	8,558,145	12,409,311	17,993,501	26,090,576
Alternate 2023-2030 Active Installed Base Scenarios								
Active Installed Base PB at 25% Annual Shipment Expansion	6,374,518	7,749,733	9,410,733	11,342,528	14,081,993	17,433,256	21,551,464	26,466,924
Active Installed Base PB at 35% Annual Shipment Expansion	6,508,036	8,230,398	10,568,668	13,675,554	18,327,364	24,667,757	33,330,622	45,018,031
Active Installed Base PB at 45% Annual Shipment Expansion	6,641,554	8,737,766	11,861,455	16,435,704	23,658,670	34,325,938	50,071,015	73,118,761

Source: Furthur Market Research (March 2023).

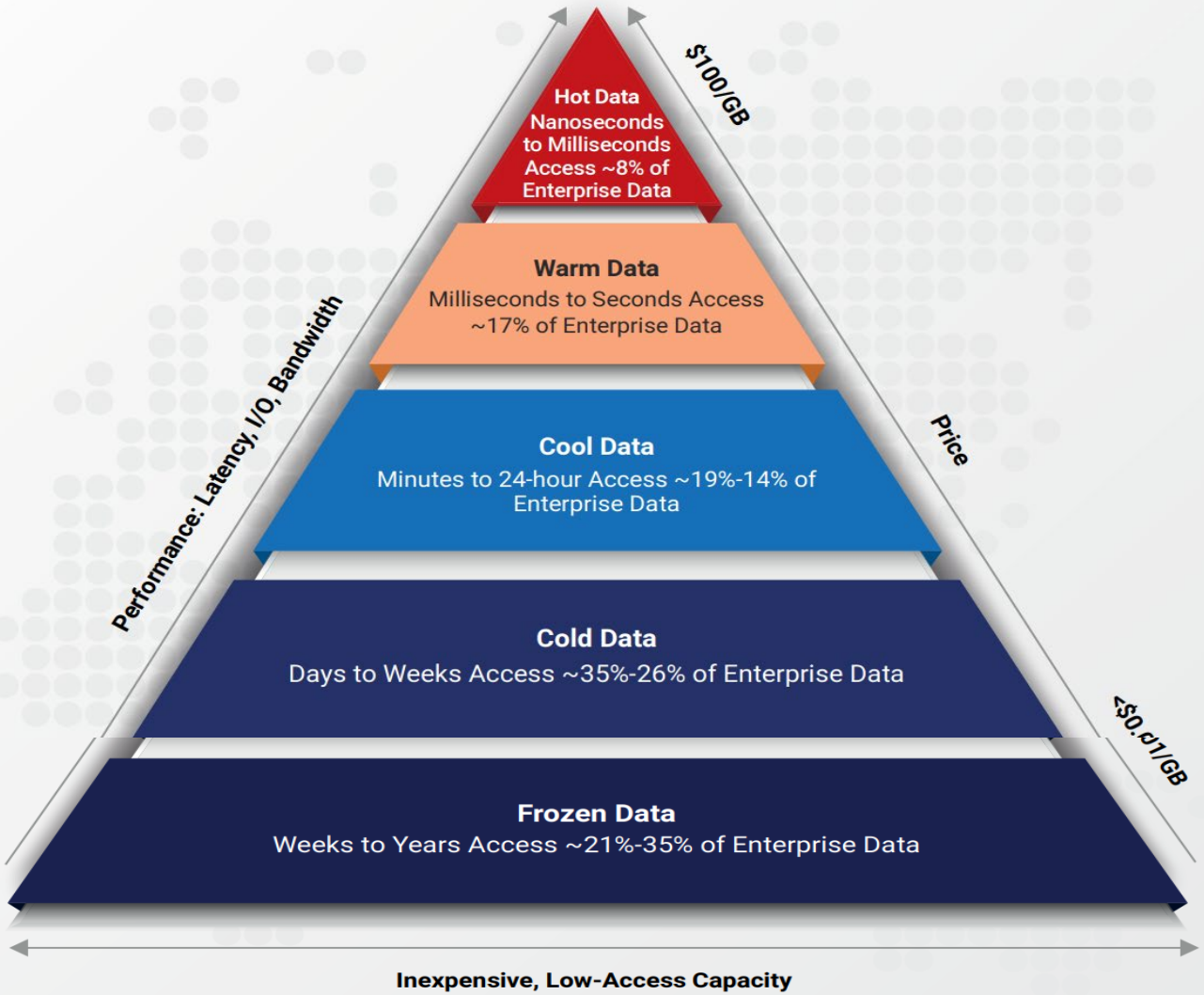
ALTERNATE 2023-2030 INSTALLED BASE GROWTH SCENARIOS



Source: Furthur Market Research (March 2023).

NEW DELINEATIONS OF ENTERPRISE DATA BASED ON ESTIMATED ACCESS FREQUENCY

Differing degrees of storage temperature, differing technologies in the layers... With an ever-increasing base of cold/frozen data...



A majority of the cold/frozen layers may be JIC (Just in Case) or WORN (Write Once Read Never) data, which may never be accessed at all—nor, in most cases, will it ever be deleted.

Source: Furthur Market Research (March 2023).

SUSTAINABILITY

Total Shipments, Interfaces	Active/Idle Watts	% of Shipments	% Active/Idle Usage	Total Power Watts		Total Power Megawatts
Enterprise-Grade HDDs ~ 428M Shipped 2020-2025						
SATA Idle	5.5W	85%=364M	65	1,301,300,000		1,301
SATA Active	7.7W		35	980,980,000		981
SAS Idle	5.8W	15%=64M	65	24,128,000		24
SAS Active	9.8W		35	21,952,000		22
					Total for HDDs	**2,328**
Enterprise-Grade SSDs ~ 399M Shipped 2020-2025						
SATA/SAS Idle	1.5W	20%=90M	40	48,000,000		48
SATA/SAS Active Read/Write	2.1W/3.2W		60	100,800,000		101
NVMe Idle	3.5W	80%=319M	40	446,600,000		447
NVMe Active Read/Write	11W/13.5W		60	2,105,400,000		2,105
					Total for SSDs	**2,701**
HDD+SSD Power Consumption vs Enterprise Tape						
					Estimated Total 2020-2025 Megawatt Power Consumption for New Shipments of HDDs + SSDs	**5,029**
					Estimated Total 2020-2025 Megawatt Power Consumption for HDDs + SSDs in the Active Installed Base (~ 3x New Shipments)	**15,087**
					Estimated Total 2020-2025 Megawatt Power Consumption for Enterprise Tape in the Active Installed Base	**18**
					Ratio HDD+SSD:Tape	**838**

Source: Furthur Market Research (March 2023)

- While the dataverse expanded (on average) by more than 30% per year, tape shipments comprised only ~15.5% of total enterprise petabytes delivered in 2022, down from 34.9% in 2015.
- It is blindingly blatant that HDDs and perhaps a significant number of SSDs are handling far too much of the cold/frozen workloads at far too great a cost/GB while consuming an inordinate share of available energy.

RECENT SURVEY RESULTS

Interviews with IT managers of 50PB-500PB databases revealed the following:

- Data retention period: "Indefinite"
- Data security, immutability and sustainability (SIS) increasingly crucial concerns
- Growing majority of the data was "cold" but could become "hot" at any time…in other words, 100% of their data is an "active archive."
- The core problem with data deletion: no agreed-upon ground rules for 5-year, 7-year, 10-year extinction periods for any data…
  - *There was always the lingering fear that after 5 years or 7 years or 10 years and 1 day, they would absolutely need that old data for some unspecified, but critical, future purpose.*

INCONCLUSIVE CONCLUSIONS

- The data centers of the future will need everything the SSD, HDD, and tape industries can manufacture and deliver, as well as requiring new DNA and optical and perhaps other enterprise storage technologies, to cost-effectively and reliably preserve the priceless artifacts of our personal, corporate, and cultural history.
- Availability and sustainability challenges will create a global need for "autonomic" data systems that can provide intelligent "active archive" management and seamless migrations of hot-to-warm-to-cool-to-cold-to-frozen data and back again, from core to edge to cloud.

INCONCLUSIVE CONCLUSIONS

- The costs of managing our multi-millionfold-petabyte dataverse over increasingly lengthy time periods will create new use cases for old storage technologies and demand the creation of new, more cost-effective, and power-efficient storage technologies.
- *Inevitably and inescapably, richly varied computing technologies will come and go, but the DATA we create will remain, and will grow to unimaginable immensity.*

An enlargement of the library of forms in which DATA, unleashed in fresh dimensions, can come to profitable life…