



SEAGATE

GOVERNMENT  
SOLUTIONS

---

# STATE OF INDUSTRY 2016

Henry Newman September 19, 2016

# Hi from Dave Anderson 9-16-2016 7:05AM



**IS IT TIME TO DO  
A COMPLETE  
RETHINK FOR  
ARCHIVES?**



# Picture from an article in 2008

Egyptian obelisk Central Park, NYC



- You can actually see the hieroglyphics
- Thanks to the [Rosetta Stone](#) you can actually decipher what it says into a modern language
  - But until finding the translation on the Rosetta Stone between ancient Egyptian hieroglyphs, Demotic script, and Ancient Greek no one knew what these hieroglyphs meant for ~1800 years!
- **NOT GOOD** or as I say **non-optimal**

# Picture from Aix-en-Provence 2016

Egyptian obelisk



- You cannot read the hieroglyphics
- Everything is lost
- All of the data is gone

**As we have said, some data loss is inevitable.**

**But even if we had the hieroglyphs there is long term trouble ahead...**

# Hieroglyphs are pretty easy compared to what we have

As the saying goes, Houston we have a problem

- **Something as simple as Microsoft file formats**
- Here are a few of the As
  - There are 246 for A alone
  - 153 for B etc etc
- **How 20 years from now much less 1000+ is any one going to figure it out**

Ext.	Description	Used by
A	<a href="#">ADA source code file</a>	
A	Library	<a href="#">Unix</a>
A##	AIZip Compressed Files' pieces	
A11	Graphics AIIM image file	
A2W	World file for Alice programming language, version 3.1	
A80	Assembler language source for 8080	
AA, AB, ...	Split parts of a single whole file	<a href="#">split (Unix)</a>
AAC	<a href="#">Advanced Audio Coding, a standardized, lossy compression and encoding scheme for digital audio</a>	
AAF	<a href="#">AutoPatcher file</a>	
AAF	<a href="#">Advanced Authoring Format, an advanced media wrapper file for professional applications</a>	
AAS	Animation Play Script	AAPlay
AB	<a href="#">Android adb Backup</a>	<a href="#">adb</a>
AB6	Datafile	ABStat
AB8	Datafile	ABStat
ABC	Alembic - 3D geometry / models	<a href="#">Many Computer Graphics tools, see Alembic (Computer Graphics) for a list.</a>
ABC	Document	ABC FlowCharter 1.0
ABC	Music in ABC format	Mainly Video Games
ABP	Data	<a href="#">ABCComp compressed</a>
ABP	AVS Barcode Profile	<a href="#">AVS Barcode Source</a>
ABF	Adobe Binary Screen Font	<a href="#">Adobe Systems Software</a>
ABK	Automatic backup file	<a href="#">CorelDRAW</a>
ABM	Image PALS album file	
ABM	Album file (various programs)	
ABR	<a href="#">Adobe Brush file for PhotoShop</a>	PhotoShop
ABS	Abstracts (info file)	
ABS	Data file	Abcissa
AC	Project backup	Audit Command Language (ACL)

# Other format types

It is not just Microsoft

**There are many other per industry format types**

**Examples include:**

**Medical imaging: Digital Imaging and Communications (DICOM)**

**Health Level Seven International (HL7)**

**Geospatial file format (lots)**

**There are more file formats that we can reasonably manage in long term archives**

# Why are file formats in State of Industry Talk?

Bringing things up a level

**Storage, whether it is tape, disk, SSD or some other media really does not matter 100 years from now – change – and data migration – are inevitable**

**Only medium we use today that has lasted over 100 years is paper**

**What matters long term is how well people in the future will be able to use our preserved data to form a better future**

**We have usable open formats like tiff, PDF and others, how long will they be compatible backwards?**

**If not for King Ptolemy V. decree in 196 BC requiring it written in 3 languages much would be lost! What can we learn from this?**

**What will our future fate be even though we are in a very different societal situation?**