
SEAGATE
GOVERNMENT
SOLUTIONS
DESIGNING STORAGE ARCHITECTURES SEPTEMBER 18-19 2017
Henry Newman | September 18, 2017

Different types of archives

Same word different meanings

Library of Congress, NARA, and many others here

- Archive=preservation of the data forever, protection of each element from corruption, and tampering

Weather Forecasting around the world

- Archive=ensuring that news models can be validated with the data that was collected
- Corruption of some of the inputs is an issue to know, but not a crisis

IOT (Internet of Things)

- Lots and lots of data to collect to turn into actionable information
- Buying stuff, driving a car, or loading your refrigerator
- Do not worry you can recollect data and generate updated information sensors are always on and always getting better

Self-driving car data collection

My opinion is liability will drive data collection and archive

The time to get a radar, camera or LIDAR detection, send it to a processing center get it back and act is likely over 1 second

- 50ms to the cloud site+data transfer time of 500ms for the data at 12 gb/sec, 500 ms to process with security controls etc. etc., and 60ms back with the decision with the data
- This amount of time is not going to work
- So the data must be processed locally and decisions must be local
- But how do you get to a point where decisions are local

LOTS OF DATA that needs to be turned into actionable information

Self driving car=lots of archive data

But what needs to be kept and why

Programming and decision support must be therefore collected, tracked and migrated to all vehicles.

- The raw data collected has information that must be anonymized
  - License plate numbers, people's face etc.
    - **But what if someone hits you then you need the data and there are all of the privacy issues around who can see what when**

Most of the issues will be seen continually so a single missed collection, corrupted collection,, and lost collection does not really matter over time

- You might say but what about the N-cases
  - And I agree just like weather those are important cases to keep but:

The issue is that preservation archives are very different

IoT problem and archives

Everything is the same but totally different

People use the same term archive to mean many different things

- Often for IoT it means collection and processing data for “deep learning”, but the data is often called an archive

Raw data in IoT world might be kept but if something is lost it is not a big deal given the results of the data are processed

- Yes there are specific cases that might want to be kept that are far out of the norm
  - In the weather world input data is often kept and used when there is a significant missed prediction

Some data does need to be kept but that data is few and far between for IoT

- It is more likely the decision process needs to be kept not the raw data 10,000s of cars that cross intersection X every single day and a single collection likely does not matter that much

Final Thoughts

Things are getting worse not better

Though data is king, it is the sum of the data not the individual single collection

- In the preservation archive world all data matters and much of it is equally important

The example I used is true for many IoT, even your home refrigerator

- Though there is not the liability except if the refrigerator gets warm and someone gets salmonella

IoT cares about aggregation of data for the most part not a specific collection

- Sure as some say we do not know what we do not know about the data, but IoT does not necessarily mean we have to keep all of the raw data

What will drive the IoT to keep raw data is liability

- Until law are passed or the courts rule on responsibility everyone will try to protect their own interests