Long-term data storage of observation data

Thoughts on the central storage of our individual observation data

Nikolai Wuensche IOTA/ES



Analogue age

- Millions of astronomical photo plates from 150 years are existing and are accessible.
- These historical plates are often used as a reference. Until today.
- Conclusion: Original observation data are valuable!
- What about <u>our</u> observation data today?



Data security

3-2-1 Backup Rule

₽X3-

Maintain at least 3 copies of your data

Keep 2 copies stored at separate locations

••••

Store at least 1 copy at an off-site location

 $-\mathbf{D}X1$

Digital age

- Who among us has a copy of their data on a secure, fire-proof drive...?
- How can we prevent the hard drives with the data from ending up in the rubbish after our demise...?
- Can we store our observation data centrally and securely and keep it accessible?



What data are really important?

- Is it sufficient to store light curves in SODIS? \rightarrow No, because...
- Incorrect results or unrecognised phenomena can be hidden in any of the light curves.
- Without the original observation data, it is no longer possible to verify this.



How big is this task?

- A few numbers from my own observations...
 - Small ROI of 800 px x 400 px, binning x2
 - Usually 3..4 minutes recording
 - Reasonable frame rate
 - \rightarrow \leq 250 MB of data per observation
- Currently, about 3000 observations per year are registered in SODIS.
- In one year, the amount of data adds up to 750 GB ...1 TB. And that every year.
- Data volume optimization makes sense!

Excursus on Data volume optimisation

- Make the region of interest (ROI) as small as possible.
 - You need three or four reference stars.
 - Use a focal reducer if necessary.
 - Small amounts of data also help to prevent dropped frames.
- Use 'Binning mode'.
- Make monochrome recordings
- Always use 12 bit or more for recording.



Excursus on Data volume optimisation

- Do not make unnecessarily long records.
 - ± 2 minutes around the limit of the predicted time is usually enough (?)
 - Longer recording time for slow objects or large error margins.
- How do cameras with analogue output and avi files fit into this scheme?

How can we store this amount of data?

Requirements:

- Central storage (organisational, not physical)
- Long-term storage possibility
- Good protection against data loss or alteration
- Selective retrieval of data must be possible
- Retrieval of individual data happens rarely, of <u>all</u> data almost never
- Restored data does not have to be available immediately.
- European Data Protection Regulation does not play a role

How can we store this amount of data?

- As we do not have our own computer cluster with tape drives: **Cloud storage**.
- Amazon, Google and Microsoft offer products that meet our requirements:
 - Amazon AWS S3 Glacier Deep Archive
 - Google Cloud Coldline storage
 - Microsoft Azure Data Lake Storage Gen. 2
- Glacier Deep Archive is the cheapest storage solution on the market.



If we sink our data into the Amazon 'Glacier'...

- storage costs: 0.00099 US\$ per GB and month,
 ~ <u>12 US\$/TB per year</u> plus 19% sales tax.
- long-term storage possible (for decades)
- highly reliable
- retrieval of data is comparatively expensive, 0.02 US\$/GB.



Before we get started...

• What data is stored?

(Are there exclusions, i. e. negative single observations of small objects?)

• Storage must be organised as well as possible to avoid cost traps.



Even more questions...

- Is Data volume optimization by observers realistic?
 What criteria? Do we have a Observations Cookbook?
- Can all observers upload >150 MB per observation without problems?
- Do we also store the data of non-members of IOTA/ES?
- Wouldn't data storage be an IOTA-worldwide task?
- Who does the work?

Conclusion

- Cloud storage is the most reliable and also cheapest solution for storing all observation data.
- We have the financial ability to permanently store our original observation data in a cloud storage.
- Some regulations for observers would be useful.
- There are still many questions to be clarified.
- And someone has to coordinate / to do the work.

Thank you!

Diskussion?

