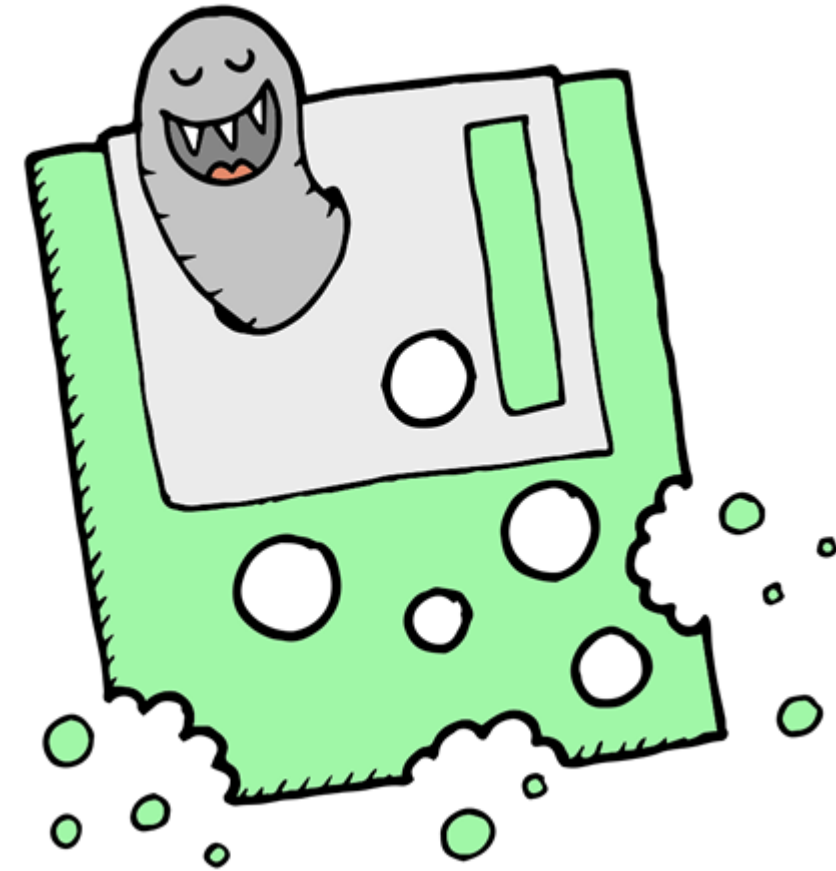# File formats for long-term archiving

SCCER/Swisstopo-Workshop, Bern, 22.05.2018

Roland E. Suri, ETH Library

Problem of long-term preservation

- File formats change over time: old files looks different or do not open

    - Old software may only run on old hardware

- Some archives migrate files to new formats (difficult, visual checks needed)

- Migration of 20 year old files is usually challenging

- Will our files be readable in 100 years??? («digital dark ages»)

File formats for long-term preservation

- Formats should be popular and follow open and non-proprietary standards

- Avoid password protection, encryption or compression.

- Use standard features of format (no fancy features)

Recommended file formats

Data type	Recommended format
Text, source code	ASCII text files (*.txt)
Tables	*.csv (Comma-separated values)
Images (raster graphics, bitmap)	Baseline TIFF (*.tif), Tagged Image File Format
Publications (text and images)	PDF/A if possible (PDF ok)

- See https://documentation.library.ethz.ch/display/DD/File+formats+for+archiving

Which archive?

ETH researchers: research-collection.ethz.ch (migration of common formats)

Other options:
- re3data.org to find specialized archives for special file formats that may need migration
- figshare.com, scribd.com (easy, but no migration, no guarantees)
- github.com for sharing code
- CodeOcean.com: archive and free simulation environment !!!

Conclusion

- File formats should be popular and follow non-proprietary standards

- ASCII for text, CSV for tables, TIFF for images, PDF/A (PDF ok) for publications

- ETH researchers: research-collection.ethz.ch