

Digital Preservation Costs – Worse than Counting Calories

LIBER conference

Riga

3rd July 2014

Yvonne Friese

nestor working group on costs



Outline

- about nestor and the working group
- digital preservation costs: issues
- use cases (selection)
 - Ingest (automated)
 - preservation planning
 - access & user support
 - legal advice to digital preservation
- summary of findings

about nestor and the working group

nestor

- founded 2004
- several publications, e. g. handbooks and guides
- 10 working groups
- nestor seal for trustworthy digital archives

working group on costs

- 8 members from 6 different institutions
- founded in 2011 – ending 2014

nestor working group on costs: benefits

develop a guideline for cost estimation

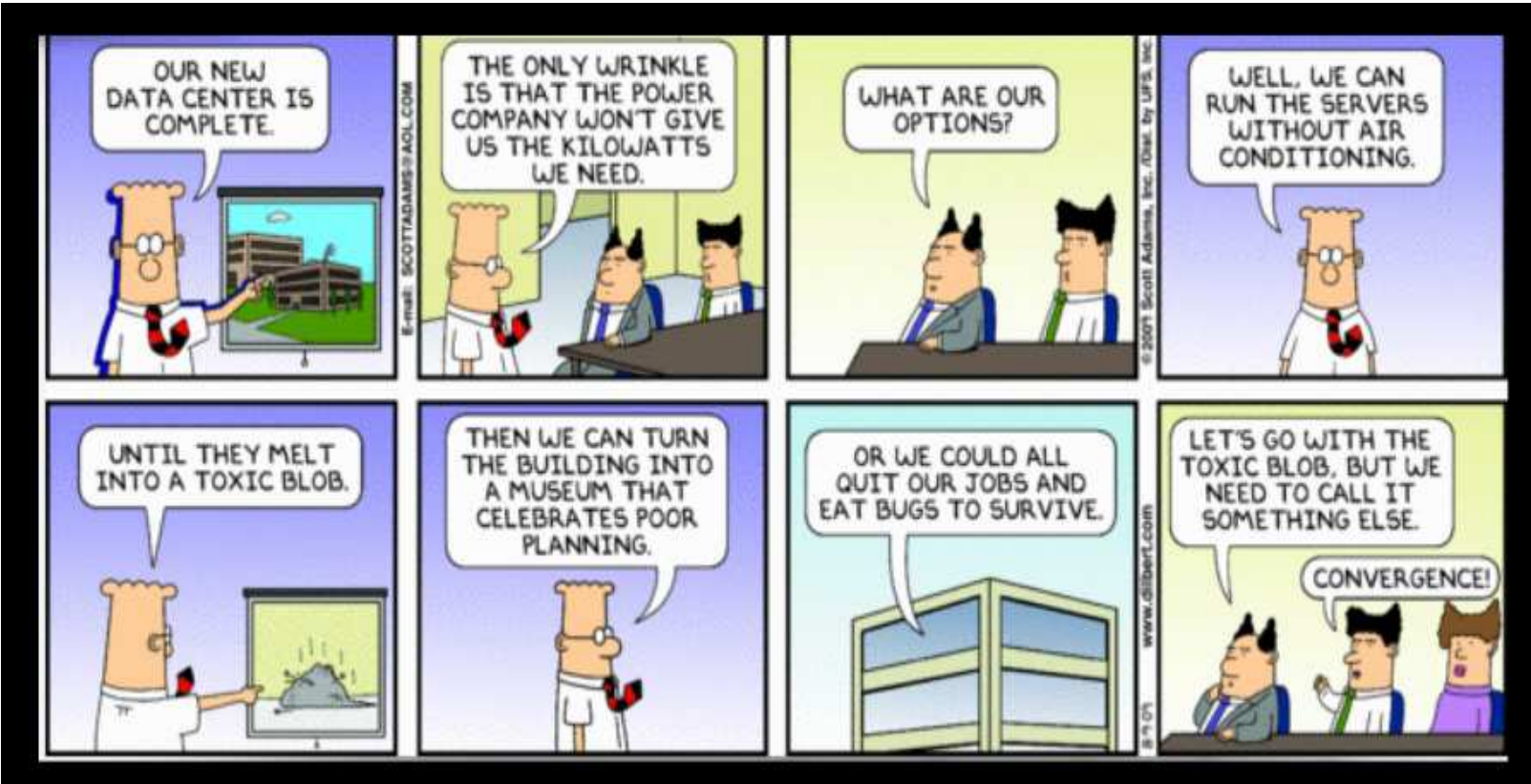
main benefits

1. **budget planning:** hidden and indirect costs included
2. **planning reliability:** ensuring duration of long term preservation
3. **cost control:** for informed decisions it is nessecary to know the costs beforehand
4. **accountability:** e.g. to prove that you cannot guarantee more than a reachability of 99,9% yet

nestor working group on costs: operation methods

- implementation of the DP4Lib cost model (activity-based costing)
- reveal hidden cost factors
- design calculation methods for Ingest, Curation, Access and Administration
- **show examples for real costs (use cases)**
- seeking collaboration (e. g. 4C)
- seeking cost transparency

to be avoided...



Source: <http://www.datacenterknowledge.com/archives/2009/08/10/dilbert-visits-the-data-center/>

Issue: costs can be too difficult to determine

- e.g. process-oriented calculations (cost per ingest per document)
- blind spots because of the lack of experience
- costs can be unpredictable (preservation planning, migration actions)
- cost calculation is costly itself (personnel time)
- difficult comparability of costs because of too many dependencies
- confidentiality (service provider, institution)



use case examples

- automated ingest via Submission Application
- preservation planning with pdf-files
- access to derivate copies and digital master, costs for support
- several use cases to legal advice to digital preservation

plus:

- evaluation + description of known cost models

automated ingest

Development time of Submission Application: 80 hours (in 2011)

- Ingest
- Metadata enrichment

Maintenance: 25 hours/year

But: development time for the metadata mapping to Dublin Core is not measurable in retrospective.

preservation planning (pdf)

Goal: fix broken PDF-files

Preservation action & planning within Rosetta Archive

- automated workflow
- implementation needed less than a week

BUT:

fixing tools test & quality ensurance

access

Hidden costs: support!

Archiving of all derivate copies:

Calculation, in case of data loss, does the storage cost really exceed the cost + effort of the re-creation of all derivate copies?

legal advise

- encrypted data
- migration of data
- support policy documents & contracts
- copyright issues



summary

- reserach is done
- publish our findings in the nestor wiki
 - use cases,
 - knowledge base,
 - cost model comparison





Questions?

Yvonne Friese
ZBW Kiel
Project Manager Digital
Preservation
y.friese@zbw.eu