

HERA, DPHEP, ADMPs and EU principles

some personal remarks

(addendum to talk on DESY/HERA data preservation)

A. Geiser, DPHEP Collaboration meeting
CERN, 14.3.2017

Data management in particle physics

this part is about plans for long term

Active Data Management ! (-> talk S. Jones)

actual ZEUS experience (also other experiments):

- passive data management (just storing the data somewhere) will not work long term,
Active data management is crucial
- **Data** must include metadata, and preservation of software, knowledge, and useability
- **Management** must include manpower needed for long term management, both at IT and user level

ZEUS

“Active Data Management Plan”

- wasn't called like that at the time, but a three page “bottom-up ZEUS ADMP” can be found in the 2012 DPHEP study group document (see previous talk)

... and we conceptually implemented more or less exactly what we planned 😊 with some practical variations (of course at that time it was already half way done)

Challenge: How to measure the success?

personal measure used:

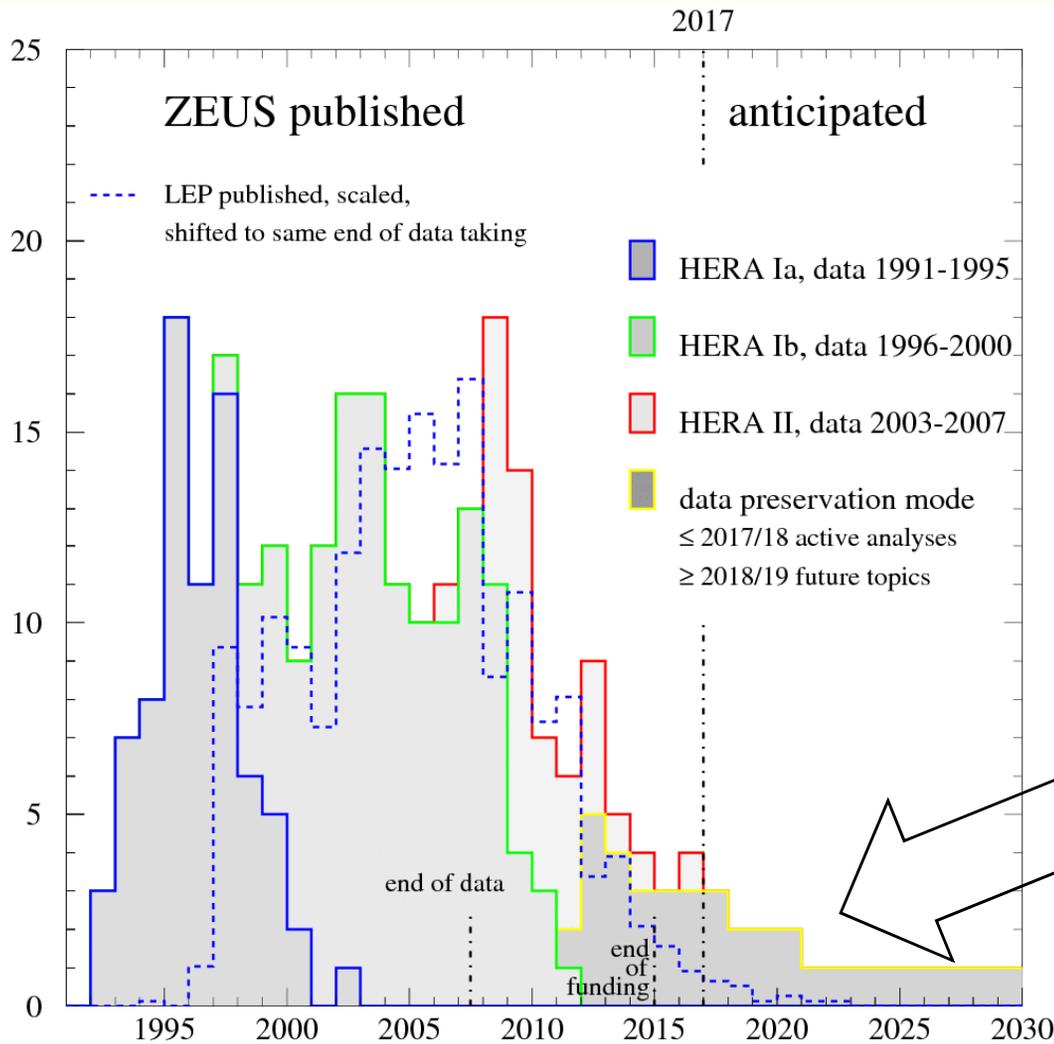
$$\frac{\text{expected \# of additional scientific papers}}{\text{total \# of scientific papers}}$$

compared to

$$\frac{\text{estimated integrated cost of data preservation}}{\text{estimated integrated total cost of project}}$$

arguable - but is there a better one?

of ZEUS papers vs. time



scientific benefit
of long term
data preservation:

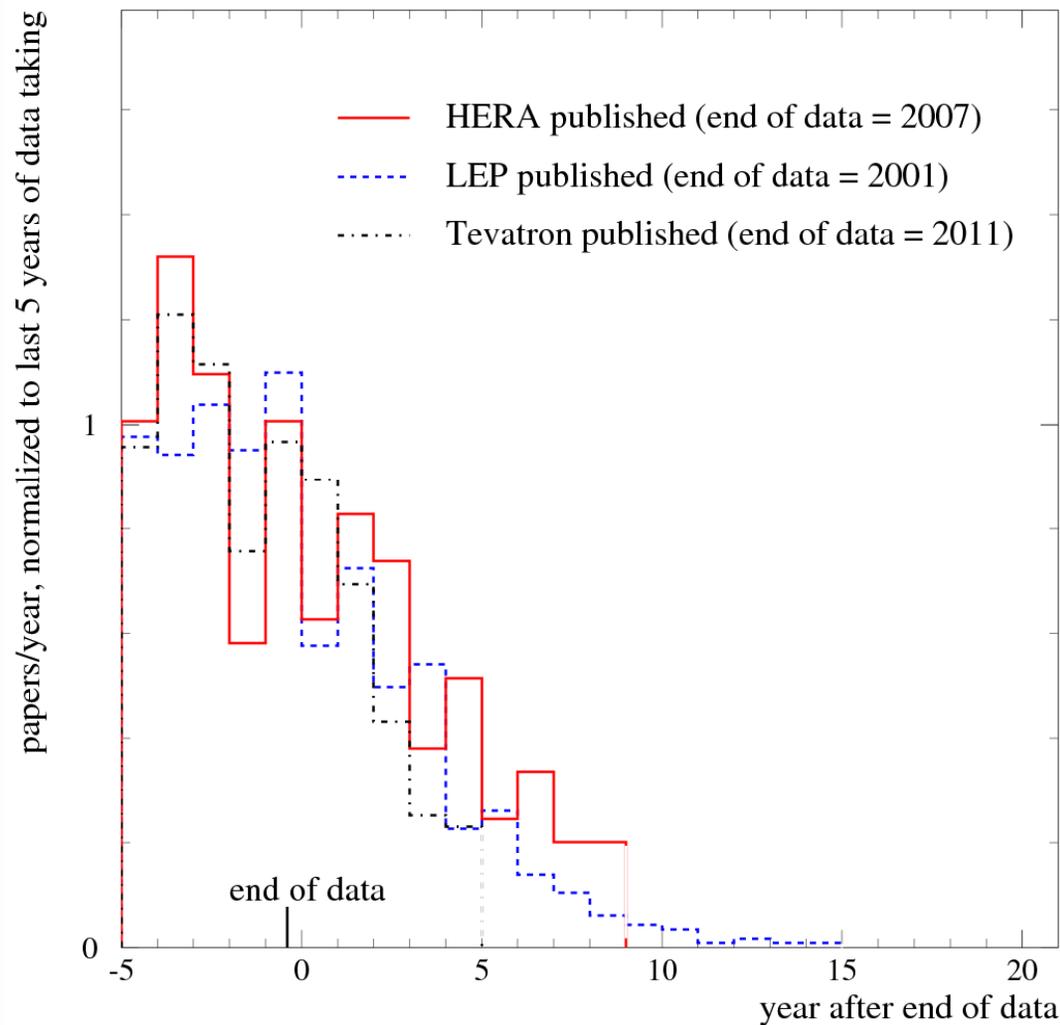
~10% of total benefit
(for <1% of total cost,
of which ~90% during
active phase)

(my personal estimate
- not official numbers)

difference between
having/following a plan,
or not having one?



HERA vs. LEP vs. Tevatron



normalized to last five
years of data taking

algorithm:

INSPIRE,

find (CN XX or CN YY or ...)

and collection:published

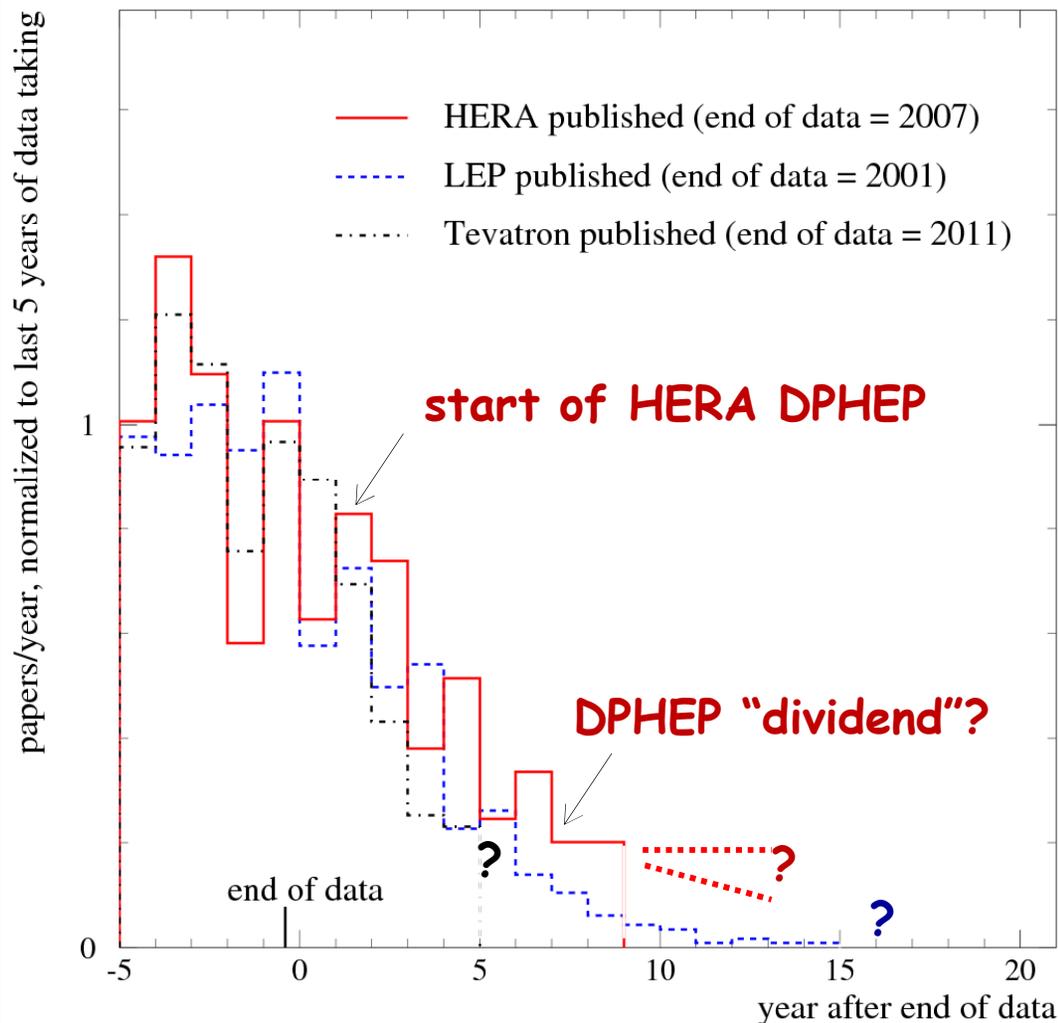
and date XXXX

and not date XXXX-1

HERA vs. LEP vs. Tevatron

normalized to last five
years of data taking

algorithm:
INSPIRE,
find (CN Exp1 or CN Exp2 or ...)
and collection:published
and date "year"
and not date "year"-1
(to avoid double counting)



“EU data principles”

- Discoverable: ZEUS and DPHEP web pages, conferences, workshops, ...
 - Accessible: ZEUS data are not (yet) open data
(would need more manpower/funding)
but “Free Access to ZEUS Data” programme for PhD students and physicists (e.g. EIC),
data accessible at DESY, + on data grid via MPI
 - Intelligible: **bottleneck!** currently OK, but would strongly profit from more manpower (keep experts involved!)
 - Assessable: quality/reproduceability is ensured by the ZEUS collaboration
 - Useable: Yes! (papers based on these data continuously being published)
all recent ZEUS papers are open access (DESY rule)
- proposal: add**
- Sustainable: **bottleneck!** Can't do without some **funding**, in particular for **long term manpower!**
“librarian” attitude to preservation could be useful!