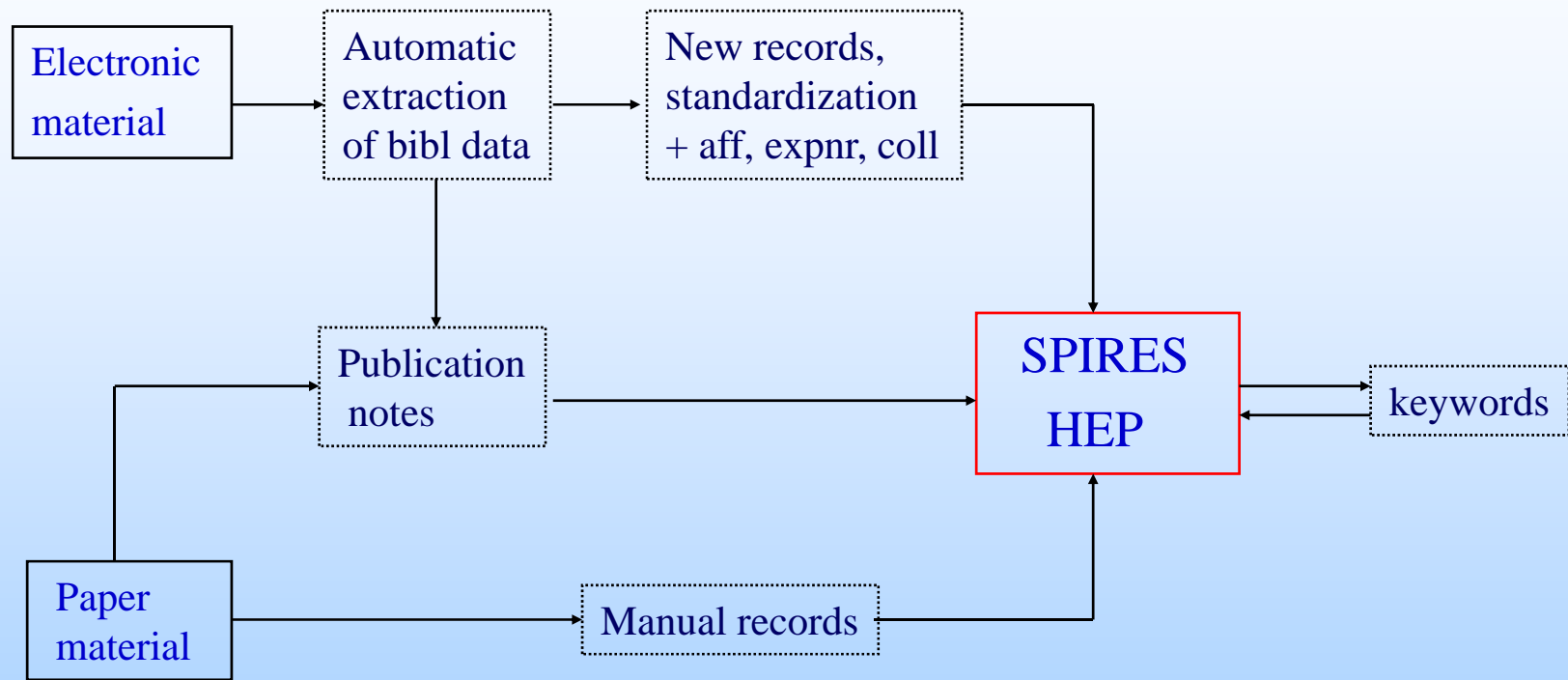


DESY Documentation: Status + projects

- SPIRES input
- Merits / deficiencies
- Current collaborations
- Future projects

SPIRES input



Bibliographic records for journal articles, conference proceedings, books, non-arXiv preprints



SPIRES input

D
E
S
Y

D
O
C

Bibliographic data:

- title, authors, publication note, rep nr, url, doi
- affiliations, collaboration + experiment nr (link to exp database)
- link to conf database
- content classification (analogous to arXiv) + standardized keywords
- formal classification (talk, review, book, jnl article...)

Acquisition of material:

- toc alerts, xml data delivery, regular scans of known sites
- active web search (conf procs, theses...)



Merits - deficiencies

D

E

S

Y

Merits:

- comprehensive: preps + jnls + conf procs + books
- content based selection
- extensive data: aff, coll, exp nr, author kw, pacs
- human quality control, standardization
- content classification + standardized keywords

Deficiencies:

- no citations for journals and conference proceedings
- lack of automatization (e.g. no automatic recognition of affiliations)
- time-delayed input (no temporary entries)
- meagre exploitation of keyword potential

D

O

C



Current collaborations

D
E
S
Y

D
O
C

- SPIRES (SLAC, Fermilab)
- Data delivery SPIRES \Leftrightarrow CDS
 - CERN articles, recently added or updated
 - recent non-arXiv preprints
- HEP taxonomy + automatic keywording (CERN)
- SCOAP³ (CERN + ...)



Future projects

- Journal citations (SLAC/CERN)
 - HEP taxonomy + automatic keywording (CERN + EPFL?)
 - Marie Curie training network
-
- Journal abstracts ⇔ abstracts database (only for searches)
 - Automatic recognition of affiliations
 - Unique author – affiliation relation
 - Scanning of paper material, extraction of bibliographic data
 - Mapping of author keywords onto standardized keywords

D
E
S
Y
D
O
C

