

INSPIRE and Data Preservation

Travis Brooks

SLAC - INSPIRE

3rd Data Preservation Workshop

7 Dec 2009

INSPIRE

- Next generation HEP information system
- Replaces aging SPIRES system with Invenio technology
- Run by CERN, DESY, Fermilab, SLAC
- Central location for HEP information

T. Brooks
INSPIRE@DPLTA
7/12/09

Release Plans

- INSPIRE Coming Soon to a Computer near You
 - Public Beta Real Soon Now (~1 Month)
 - Full release early 2010 (Feb/March?)
- Already some features released
 - Helpdesk
 - Classification

T. Brooks
INSPIRE@DPLTA
7/12/09

INSPIRE and DPHEP

- Storing documentation
- Storing other objects
- Linking with papers
- Providing simple interface to get information to the external world
- Ensuring that HEP information plays well with others

T. Brooks
INSPIRE@DPLTA
7/12/09

Experiment Documentation

- Notes series
 - Pretty trivial to capture
 - Can be password protected or embargoed
- Other flat file documentation
 - Submit it, no problems

T. Brooks
INSPIRE@DPLTA
7/12/09

Wikis

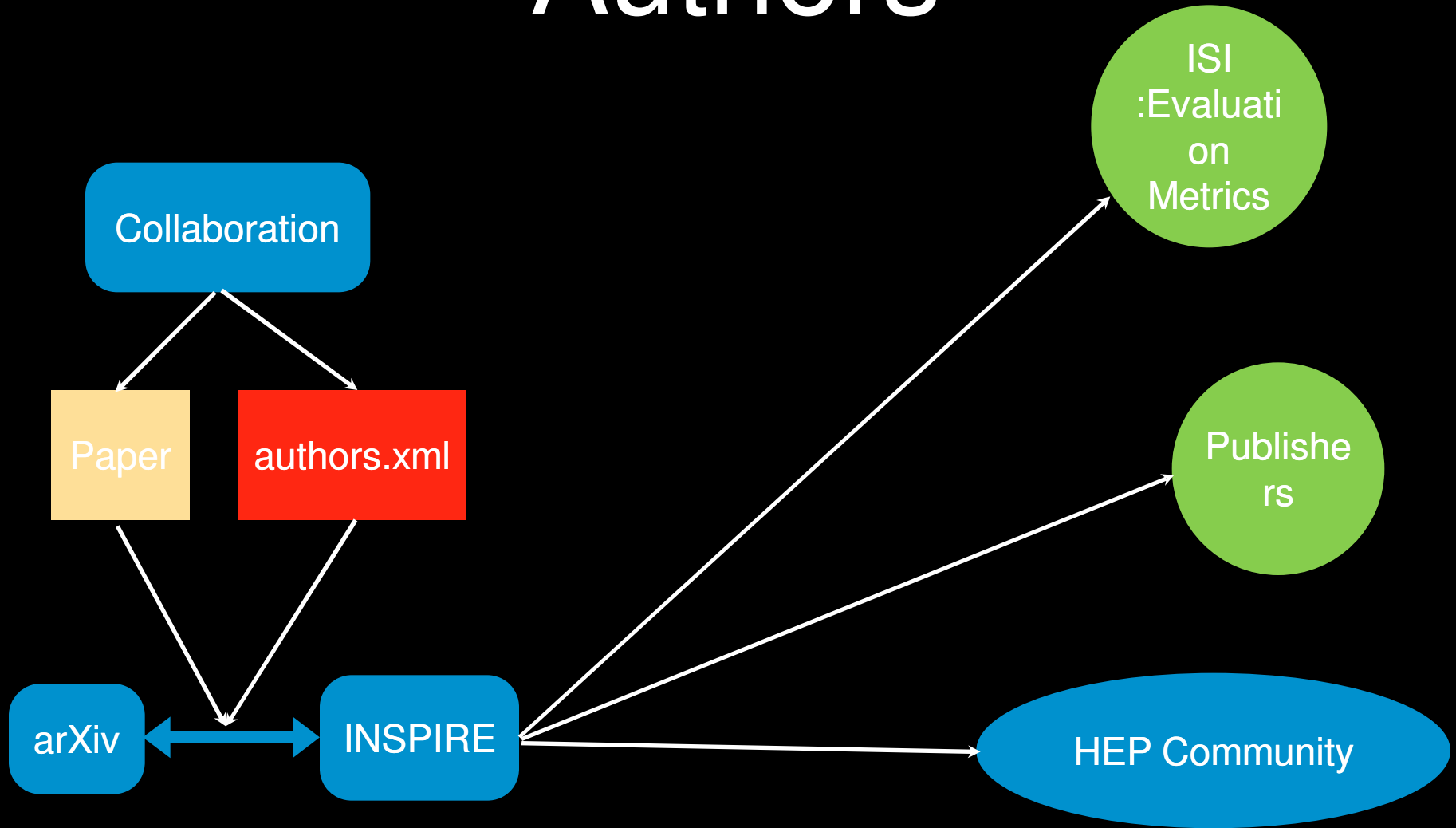
- “Shadowing” internal wikis
 - Backup archive on INSPIRE - 3rd party site
 - Goes live in frozen state when collaboration wishes
- Harder to do, but potentially very useful...
- Some tools available, some need to be written

Other Objects

- Will people actually store stuff?
- Look at a use case that's already happening:

T. Brooks
INSPIRE@DPLTA
7/12/09

Use Case: 2000 Authors



Author Lists

INSPIRE Standardizes Format
Uses IDs to solve name
ambiguities

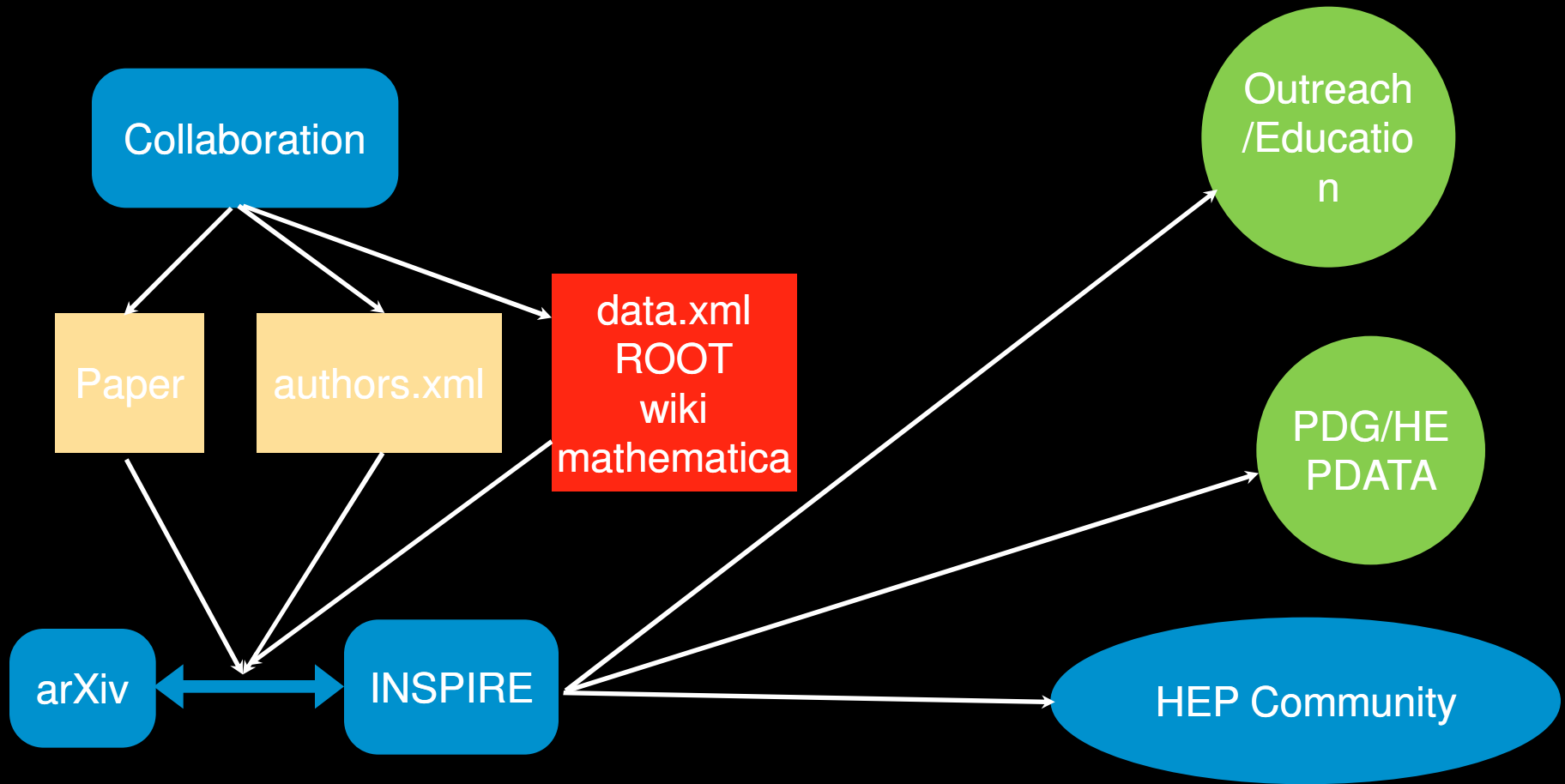
Removes much manual labor
Enables sharing with external
entities from a central source

Authors/collaborations are
already contributing!

CMS, MiniBooNe, IceCube,
CDF, Auger, MIPP already,
others coming soon

```
- <authorList>
  <collaboration>CDF</collaboration>
  - <authorOrganization>
    - <author>
      <name>Aaltonen, T.</name>
      <authorID>INSPIRE-00010138</authorID>
    </author>
    <organization>hip.fi</organization>
  </authorOrganization>
  - <authorOrganization>
    - <author>
      <name>Adelman, Jahred A.</name>
      <authorID>INSPIRE-00041460</authorID>
    </author>
    <organization>uchicago.edu</organization>
  </authorOrganization>
  - <authorOrganization>
    - <author>
      <name>Akimoto, T.</name>
      <authorID>INSPIRE-00041459</authorID>
    </author>
    <organization>tsukuba.ac.jp</organization>
  </authorOrganization>
```

Why stop with authors?



What to store?

- Paper-level data
 - Numbers behind figures, tables
 - XML, ROOT, CSV
- High resolution images
- Mathematica Notebooks
- Code

Depositing Stuff

- Deposit in INSPIRE (or arXiv .. seamless)
- Importance of trusted 3rd party
- INSPIRE would be able to provide embargo period (a la CMS/ATLAS “push to publish”)
- Continue our current philosophy:
 - Accept just about anything, but work with providers to optimize and standardize formats
- INSPIRE already uses OAI-PMH for ingestion

T. Brooks
INSPIRE@DPLTA
7/12/09

Summary

- INSPIRE will soon get experimental note series
- INSPIRE already gets feeds of XML author lists
- A bit of work can re-use this infrastructure for high-level data (Data from figures/tables, ROOT, Mathematica, etc.)
- With work can extend to allow embargoes, and wiki tracking/ingestion

T. Brooks
INSPIRE@DPLTA
7/12/09