BibAuthorID

a brief status update
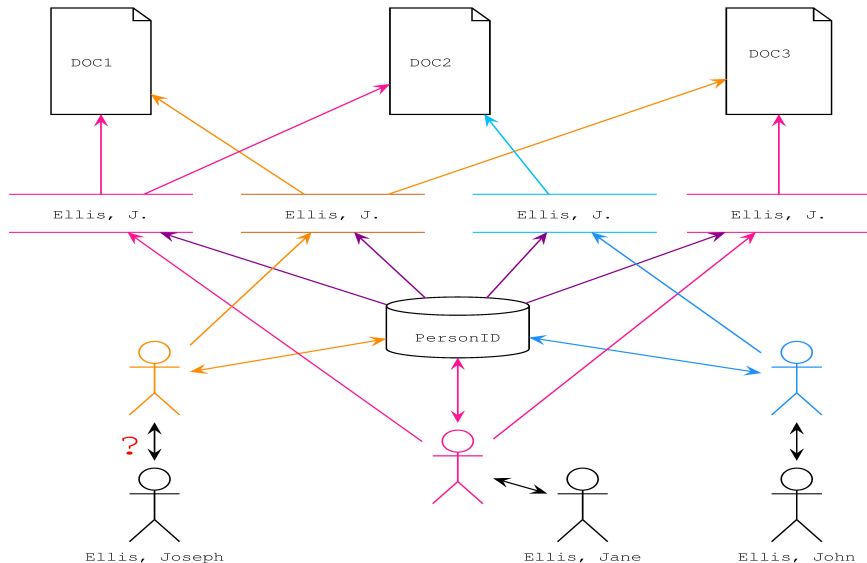
Samuele Carli

19 novembre 2012

Contents

A quick reminder: what are we doing here?

What's next?

A quick reminder: data flows
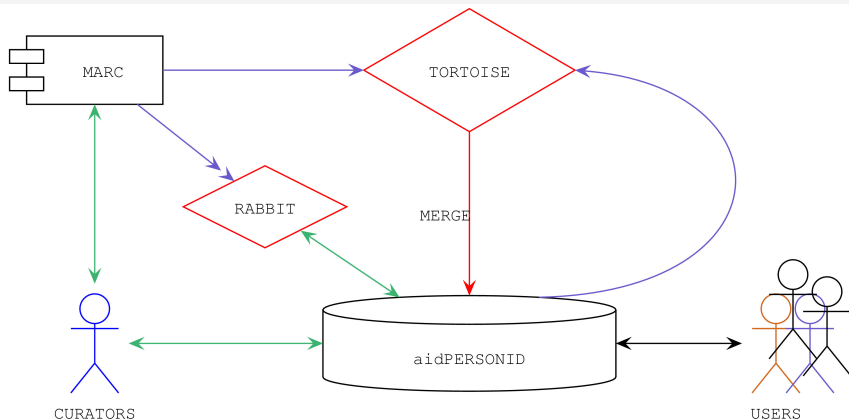
What's next?

What's going on here?

- RABBIT: READY!
  - now can use externally collected IDs for even faster and correct shipping INSPIRE (Tuning suite under development)
- TORTOISE: READY!
  - Disambiguates 1M records in INSPIRE in 10 days
  - Undergoing profiling to increase perfomances
  - Undergoing tuning to identify proper thresholds for best results within INSPIRE (Tuning suite under development)
- MERGER: READY!
  - available in two flavours: Online and Offline
- USABILITY: NOT READY
  - needs constant supervision…
  - …especially when something crashes!
  - missing easy and clear CLI including covering steps asd supervision

What's next?

Are we there yet? And now? Still not there? What about now? Are we there yet?

- Develop tools to allow almost completely automated workflow
- Develop missing features (mostly UI related)
- Strenghten the backend (learn to fail gracefully)
- Optimization
- Fix bugs
- Fix bugs
- Find new bugs
- Fix new bugs
- Ups that actually was an old bug...

Good news: as things are getting more and more stable, it gets easier and faster to spot and fix bugs

What's next?

Press any... nooooo, not that one!

- There is no politic for stability of signatures on records
  - If exact same name twice on paper, bibauthorid sees it as only one author with union of data (different strings for different author works as a workaround)
  - It's a pain to deal with record updates when they involve author data; Some guessing must be involved which leads to (claim) data loss; No easy solution: better tools are needed so that curators can specify if a modified signature is the same person as before or not, but problem remains in case of exported data (cannot keep track with 100% accuracy)
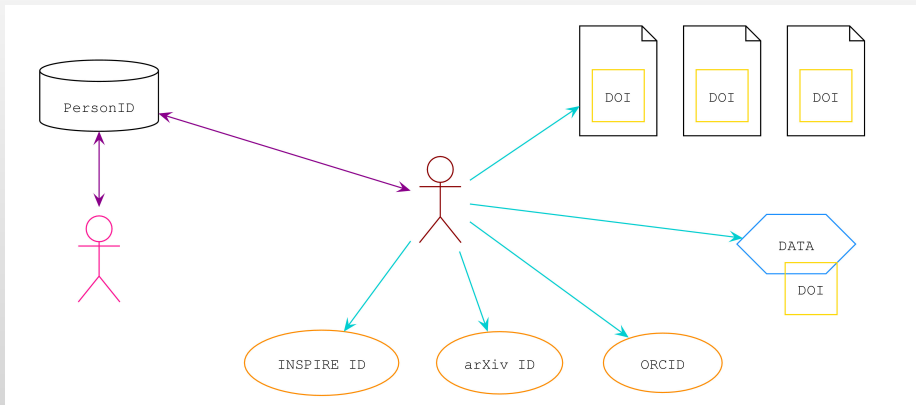
What's next?

...is to be decided!

- List of things to improve is limitless....
- ...especially regarding performances (accuracy, tuning, speed, reliability...)
- Interoperability
- Integration
- Now it's the right time for 'beta testing' and feature tuning / new feature requests!

What's next?

Interoperability

What's next?

Figures as of 12/11/2012

- 7418424 signatures on 966990 papers
- 284470 distict disambiguated PersonIDs
- 5159 PersonIDs with at least one paper manually claimed
- 352571 claimed signatures on 183557 papers
- 4076 physical persons claimed paper directly through arXiv SSO login

Want to know more?

Q&A session!

For further information call 1-800-bibauthorid-help or:

Samuele Carli

E-mail: scarli@cern.ch
Web: www.csspace.net