

INSPIRE Live Migration

Samuele Kaplun
IUGW2017



INSPIRE as a Graph

Papers + People + Places + Projects... All linked

Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+ \mu^-$ decays

LHCb Collaboration (Roel Aaij (CERN) et al.) [Show all 775 authors](#)

Mar 16, 2017 - 19 pages

CERN-EP-2017-041, LHCb-PAPER-2017-001

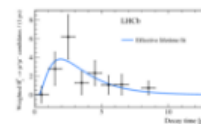
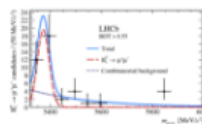
e-Print: [arXiv:1703.05747 \[hep-ex\]](#) | [PDF](#)

Experiment: [CERN-LHC-LHCb](#)

Abstract (arXiv)

A search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed at the LHCb experiment using data collected in pp collisions corresponding to a total integrated luminosity of 4.4 fb^{-1} . An excess of $B_s^0 \rightarrow \mu^+ \mu^-$ decays is observed with a significance of 7.8 standard deviations, representing the first observation of this decay in a single experiment. The branching fraction is measured to be $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$, where the first uncertainty is statistical and the second systematic. The first measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime, $\tau(B_s^0 \rightarrow \mu^+ \mu^-) = 2.04 \pm 0.44 \pm 0.05 \text{ ps}$, is reported. No significant excess of $B^0 \rightarrow \mu^+ \mu^-$ decays is found and a 95 % confidence level upper limit, $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}$, is determined. All results are in agreement with the Standard Model expectations.

Note: All figures and tables, along with any supplementary material and additional information, are available at <https://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbProjectPublic/LHCb-PAPER-2017-001.html>



XME (Extended MarcXML)

Custom INSPIRE specific format where we populate in \$0/\$z... the recid of all referred records.

<https://inspirehep.net/record/1517942/export/xme>

Vs.

<https://inspirehep.net/record/1517942/export/xm>

inspire-schemas

<https://github.com/inspirehep/inspire-schemas>

Just metadata to uniquely identify and to make discoverable items for HEP community

E.g.

<https://qa.inspirehep.net/api/literature/1509304>

Using JSON-REFs

dojson

We've built bi-directional conversion between
INSPIRE XME and inspire-schemas format
Unit tested **every single field(!)**

E.g. see:

https://github.com/inspirehep/inspire-next/blob/master/tests/unit/dojson/test_dojson_hep_bd1xx.py#L35

dojson tests

```
tests/unit/dojson/test_dojson_common.py .....
tests/unit/dojson/test_dojson_conferences.py .x.....X.....
tests/unit/dojson/test_dojson_experiments.py .....
tests/unit/dojson/test_dojson_hep_bd0xx.py .....
tests/unit/dojson/test_dojson_hep_bd1xx.py .....x..
tests/unit/dojson/test_dojson_hep_bd2xx.py .....
tests/unit/dojson/test_dojson_hep_bd3xx.py ...
tests/unit/dojson/test_dojson_hep_bd4xx.py .
tests/unit/dojson/test_dojson_hep_bd5xx.py .....
tests/unit/dojson/test_dojson_hep_bd6xx.py .....
tests/unit/dojson/test_dojson_hep_bd7xx.py .....
tests/unit/dojson/test_dojson_hep_bd9xx.py .....
tests/unit/dojson/test_dojson_hepnames.py .....x.....
tests/unit/dojson/test_dojson_institutions.py .....x.....
tests/unit/dojson/test_dojson_jobs.py .....
tests/unit/dojson/test_dojson_journals.py .....
tests/unit/dojson/test_dojson_model.py ..
tests/unit/dojson/test_dojson_processors.py .
tests/unit/dojson/test_dojson_utils.py .....
tests/unit/dojson/test_dojson_utils_geo.py .....
```

Nightly & Sentry

Every night we reload the 1.5M records on a dedicated test machine and all failures are recorded:

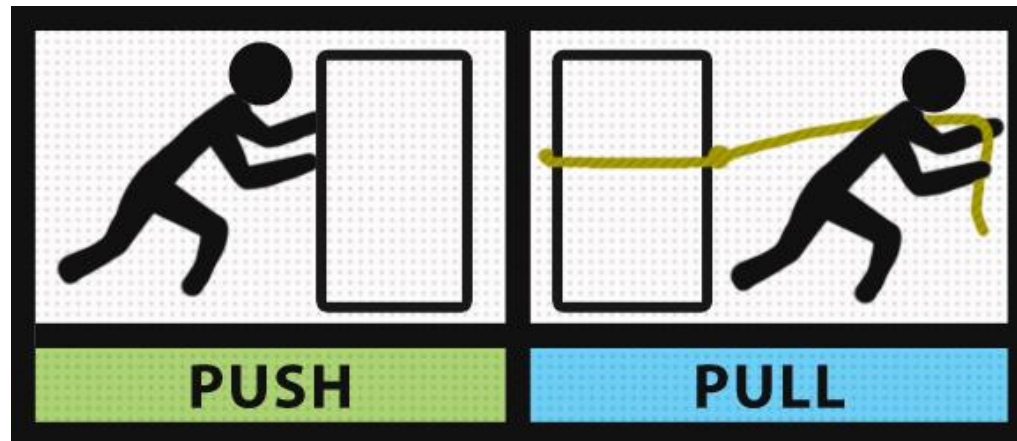
176	inspirehep.modules.migrator.tasks.records in migrate_and_insert_record Migrator Validation Error: properties.persistent_identifiers.items.required on [u'schema', u'valu... users 0 10 hours ago inspirehep.modules.migrator.tasks.records	✓ ★
1	elasticsearch.connection.base in _raise_error TransportError: TransportError(503, u'search_phase_execution_exception') users 1 20 hours ago root	✓ ★
1	elasticsearch.connection.base in _raise_error TransportError: TransportError(503, u'search_phase_execution_exception') users 1 20 hours ago root	✓ ★
1	flask_oauthlib.client in _get_property KeyError: 'ORCID_APP_CREDENTIALS_CONSUMER_KEY' users 1 2 days ago root	✓ ★

Keeping things in sync

Legacy service pushes new and modified record in XME format

Every 5 minute, all new/modified records are queued on a redis server

Every minute labs pull from the redis server and import records



Versioning

After 2 years of fine-tuning our data-model we are aware the continuous need for evolution

A plan has been devised for schema versioning:

- Semantic versioning

- Automatic upgrades of records

Lessons learnt

Designing your data model is the most complex part

Exposing the hidden relations

Continuous migration

Test **EVERYTHING**

Versioning is a must

