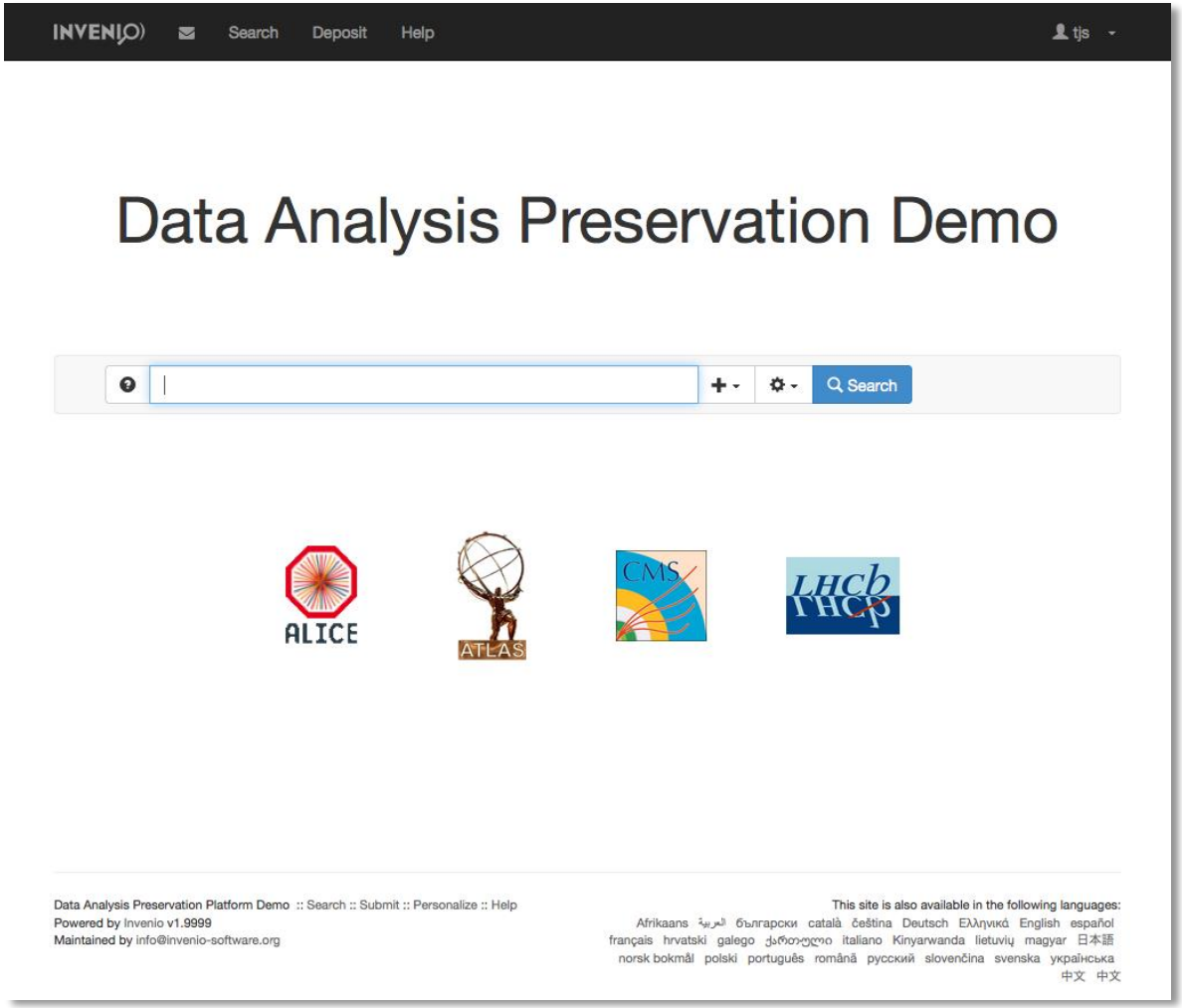
CERN Analysis capture Portal (CAP)

Tim Smith

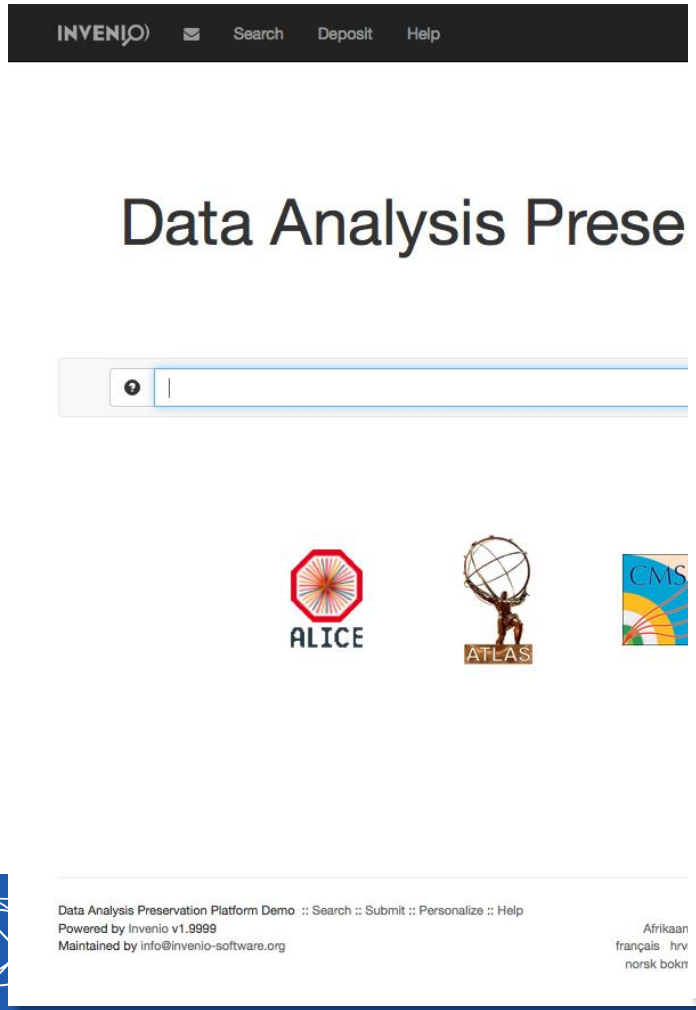
On behalf of Patricia, Tibor,
Sünje, Frank, and many more

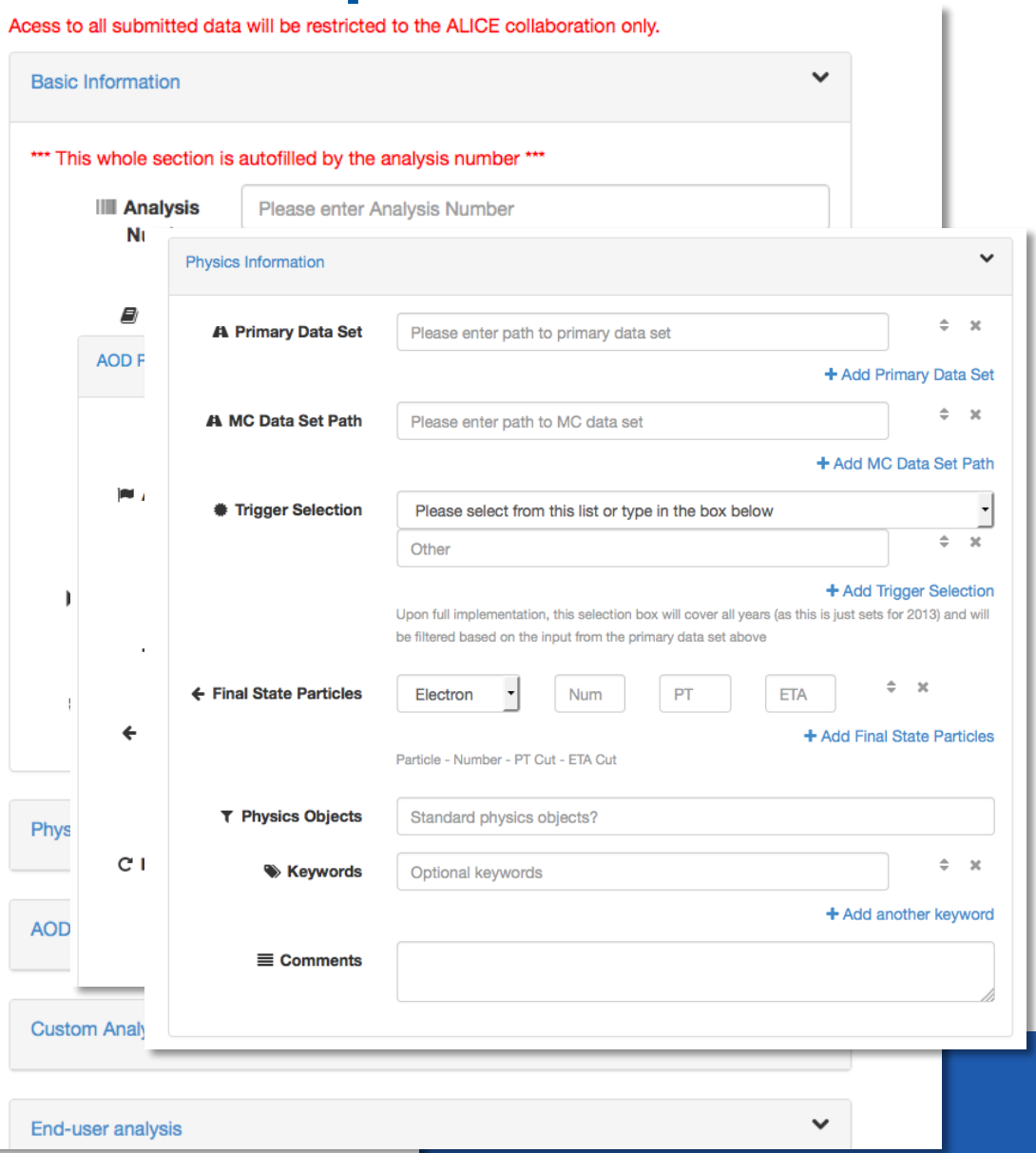
Pilot


INVENIO
Search
Deposit
Help
tjs
Data Analysis Preservation Demo
Search
ALICE
ATLAS
CMS
LHCb
Data Analysis Preservation Platform Demo :: Search :: Submit :: Personalize :: Help
Powered by Invenio v1.9999
Maintained by info@invenio-software.org
This site is also available in the following languages:
Afrikaans العربية български català čeština Deutsch Ελληνικά English español
français hrvatski galego ქართული italiano Kinyarwanda lietuvių magyar 日本語
norsk bokmål polski português română русский slovenčina svenska українська
中文 中文

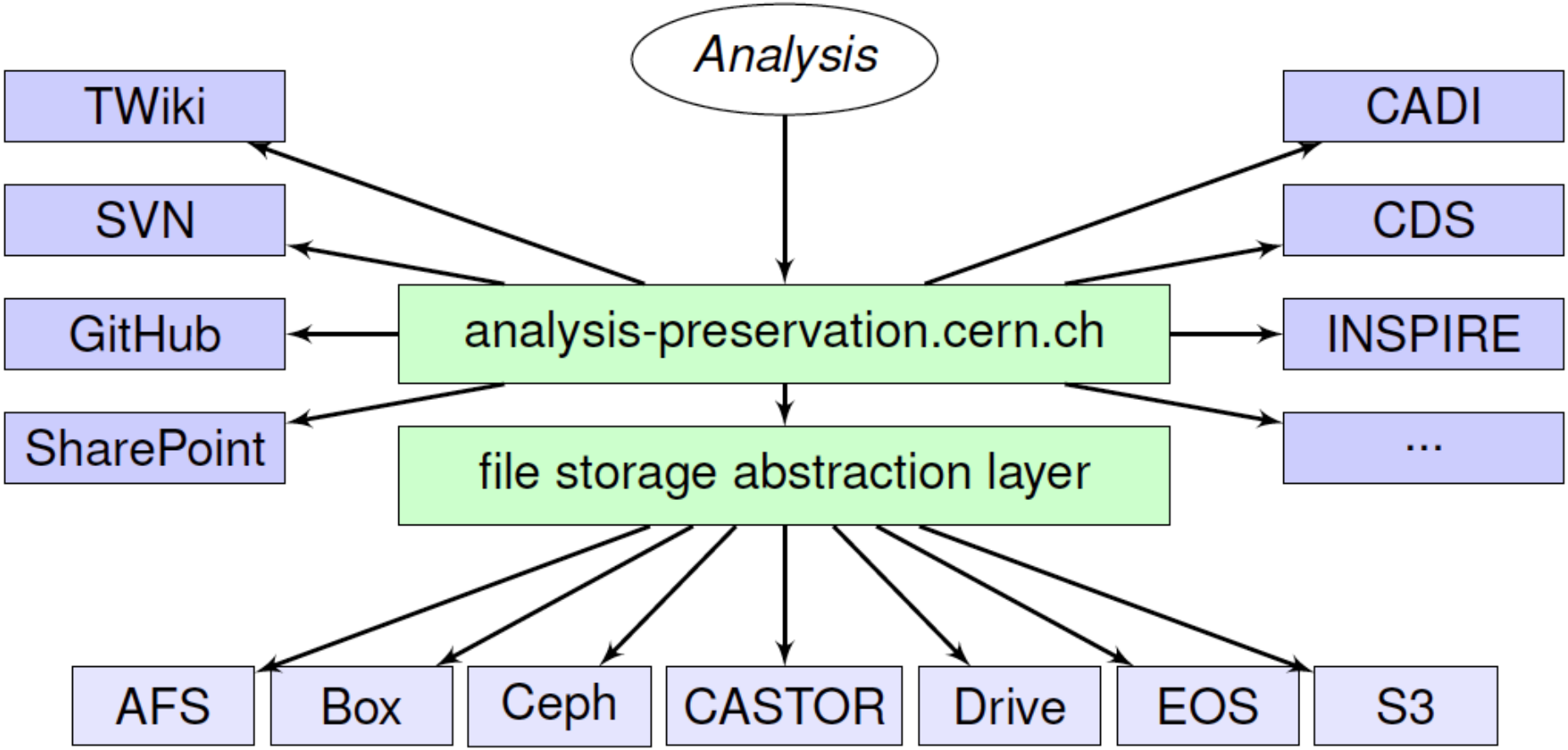
Submission Mock-ups

- Workflow, code, statis environment





System Architecture



Timelines

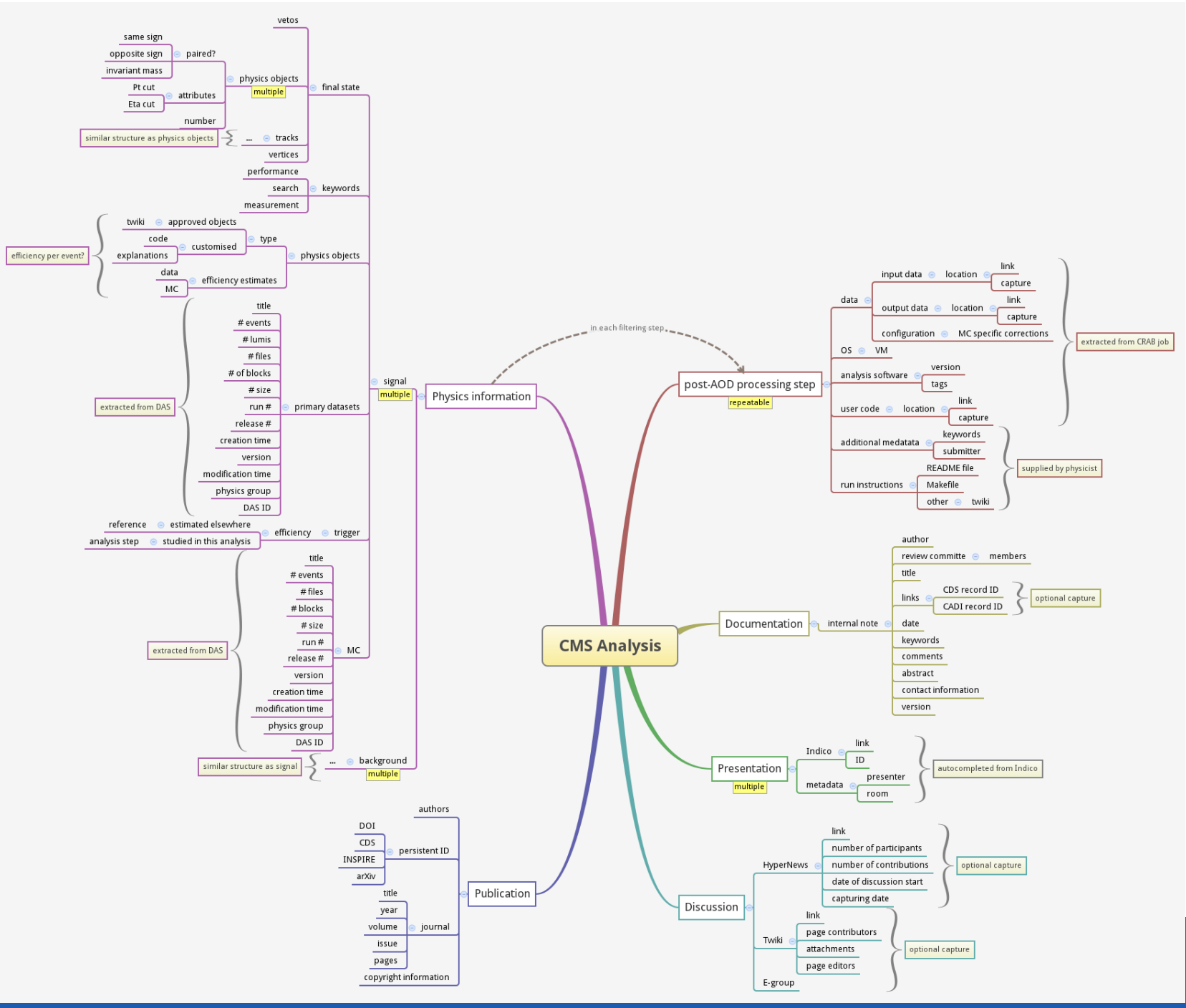
- **Summer 2014** Pilot studies
  - example analysis
  - submission forms
- **Winter 2014** Infrastructure updates
  - Invenio 2
  - EOS
  - OAuth
- **Spring 2015** Restarting the work
  - requirements and use cases
  - data model
  - RECAST
  - statistics committee questionnaire
  - synergies with DASPOS (VoCampND2015)

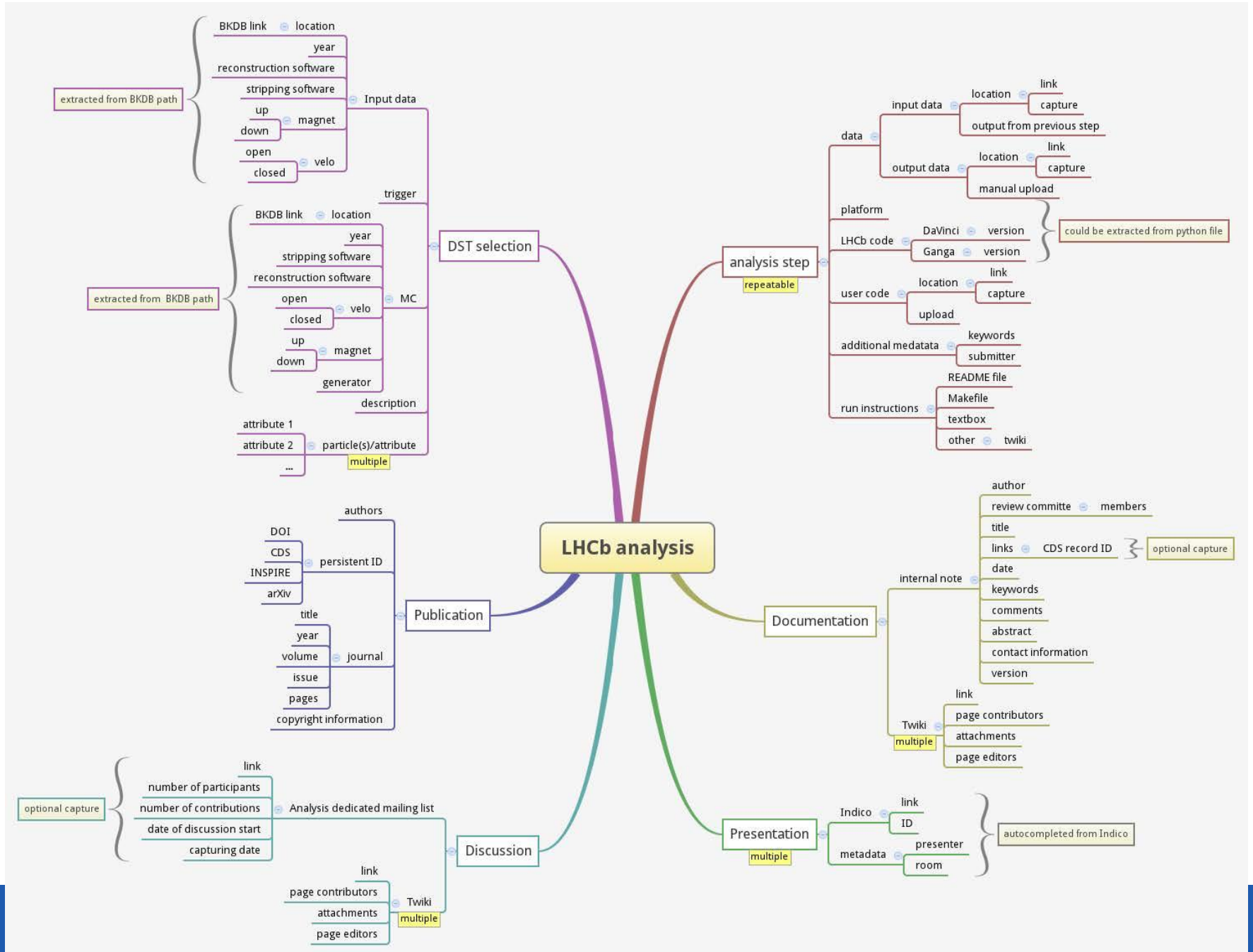
CAP Use Cases (I)

1. The person having done (part of) an analysis is leaving the collaboration and has to hand over the know-how to other collaboration members.
2. A newcomer would like join a group working on some physics subject
3. In a large collaboration, it may occur that two (groups of) people work independently on the same subject
4. There is a conflict between results of two collaborations on the same subject

CAP Use Cases (II)

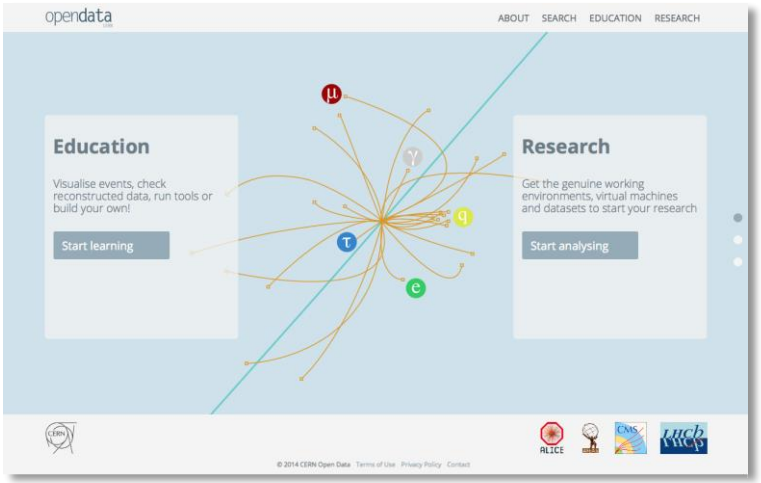
5. A previous analysis has to be repeated
6. Data from several experiments, on the same physics subject, have to be statistically combined
7. A working group or management member within a collaboration wishes to know who else has worked on a particular dataset, software piece or MC
8. Presentation or publication is submitted for internal/collaboration review and approval: lack of comprehensive metadata
9. Preparing for Open Data Sharing

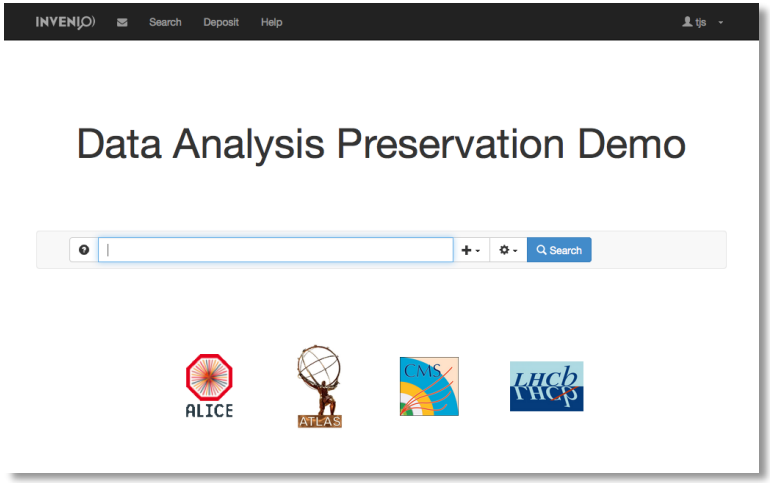

CMS Analysis
Physics information
in each filtering step
signal
multiple
final state
vetos
physics objects
multiple
paired?
same sign
opposite sign
invariant mass
attributes
Pt cut
Eta cut
number
similar structure as physics objects
...
tracks
vertices
keywords
performance
search
measurement
physics objects
type
approved objects
customised
code
explanations
efficiency per event?
efficiency estimates
data
MC
primary datasets
extracted from DAS
title
events
lumis
files
of blocks
size
run #
release #
creation time
version
modification time
physics group
DAS ID
efficiency
trigger
reference
estimated elsewhere
analysis step
studied in this analysis
MC
extracted from DAS
title
events
files
blocks
size
run #
release #
version
creation time
modification time
physics group
DAS ID
background
multiple
similar structure as signal
...
post-AOD processing step
repeatable
data
input data
location
link
capture
output data
location
link
capture
configuration
MC specific corrections
extracted from CRAB job
OS
VM
analysis software
version
tags
user code
location
link
capture
additional medatata
keywords
submitter
supplied by physicist
run instructions
README file
Makefile
other
twiki
Documentation
internal note
author
review committe
members
title
links
CDS record ID
CADI record ID
optional capture
date
keywords
comments
abstract
contact information
version
Presentation
multiple
Indico
link
ID
metadata
presenter
room
autocompleted from Indico
Discussion
HyperNews
link
number of participants
number of contributions
date of discussion start
capturing date
optional capture
Twiki
link
page contributors
attachments
page editors
E-group
optional capture
Publication
authors
persistent ID
DOI
CDS
INSPIRE
arXiv
journal
title
year
volume
issue
pages
copyright information


LHCb analysis
DST selection
Input data
BKDB link
location
year
reconstruction software
stripping software
magnet
up
down
velo
open
closed
extracted from BKDB path
MC
trigger
BKDB link
location
year
stripping software
reconstruction software
velo
open
closed
magnet
up
down
generator
extracted from BKDB path
description
particle(s)/attribute
attribute 1
attribute 2
...
multiple
analysis step
repeatable
data
input data
location
link
capture
output from previous step
output data
location
link
capture
manual upload
platform
LHCb code
DaVinci
version
Ganga
version
could be extracted from python file
user code
location
link
capture
upload
additional medatata
keywords
submitter
run instructions
README file
Makefile
textbox
other
twiki
Publication
authors
persistent ID
DOI
CDS
INSPIRE
arXiv
journal
title
year
volume
issue
pages
copyright information
Documentation
internal note
author
review committe
members
title
links
CDS record ID
optional capture
date
keywords
comments
abstract
contact information
version
Twiki
multiple
link
page contributors
attachments
page editors
Discussion
Analysis dedicated mailing list
optional capture
link
number of participants
number of contributions
date of discussion start
capturing date
Twiki
multiple
link
page contributors
attachments
page editors
Presentation
multiple
Indico
link
ID
metadata
presenter
room
autocompleted from Indico

Conclusion

Capture, Share, Preserve





In close collaboration with:

Kati Lassila-Perini, Tom McCauley, Achintya Rao

Alicia Calderon, Ana Rodriguez-Marrero, Adam Huffman, Jonatan Piedra

Silvia Amerio, Ben Couturier, Ana Trisovic

Mihaela Gheata, Costin Grigoras

Kyle Cranmer, Lukas Heinrich, Felix Socher, David Rousseau

Mike Hildreth, Jakob Blomer, Luca Mascetti

Frank Berghaus, Jamie Shiers

DPHEP

DASPOS

www.cern.ch