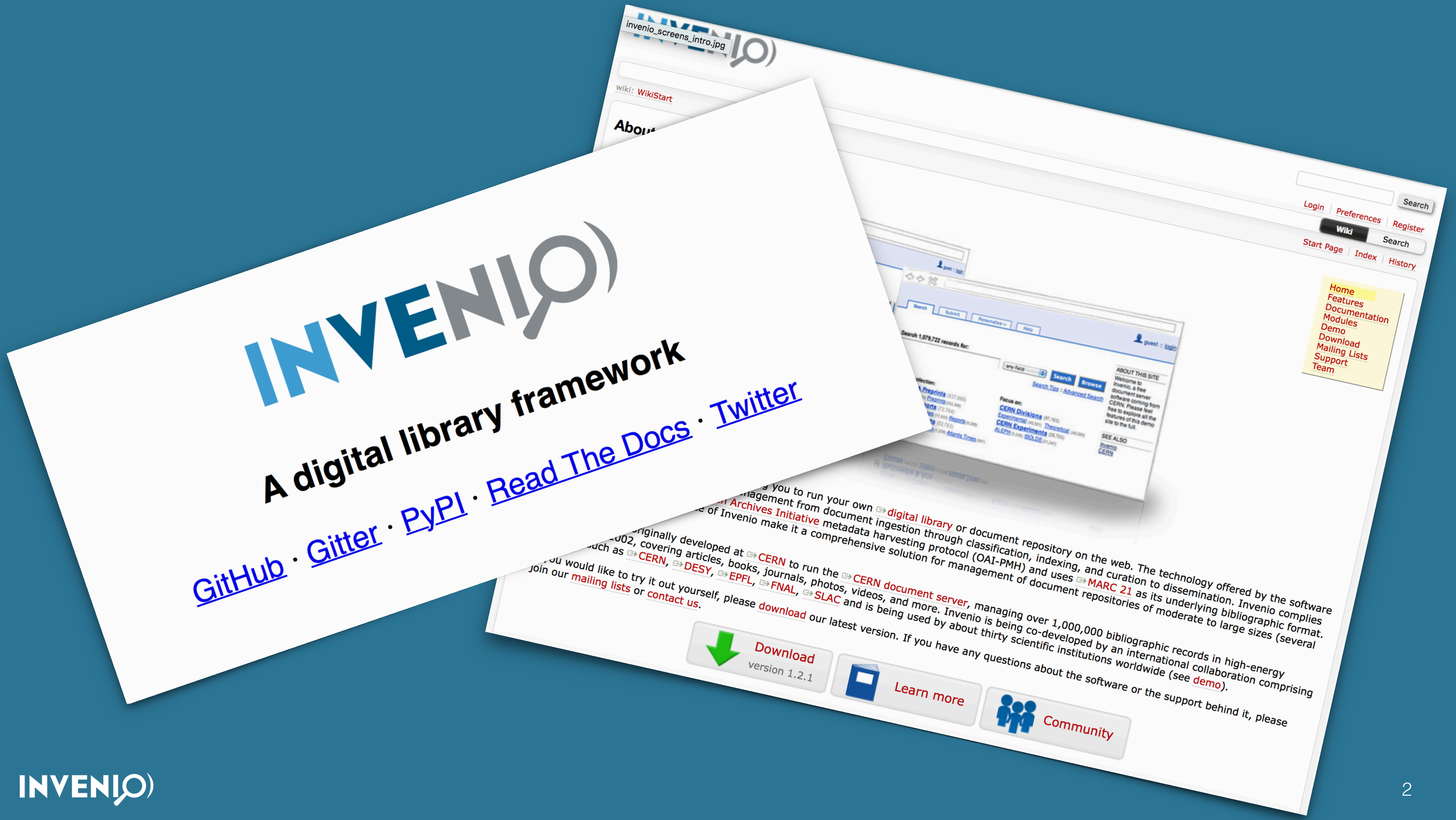
Invenio
State of the Union

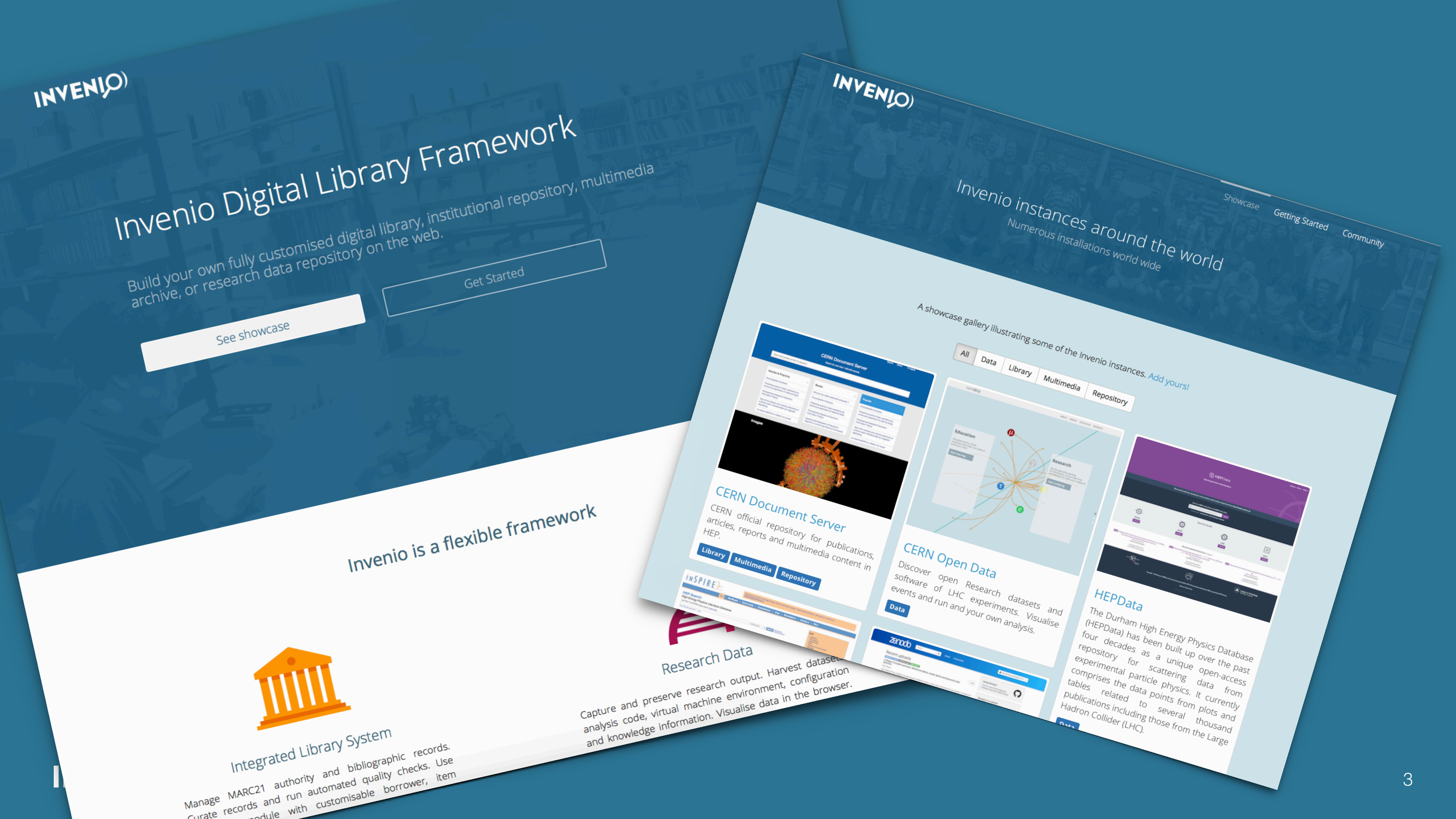
Lars Holm Nielsen
CERN

Invenio User Group Workshop 2017, Garching bei München

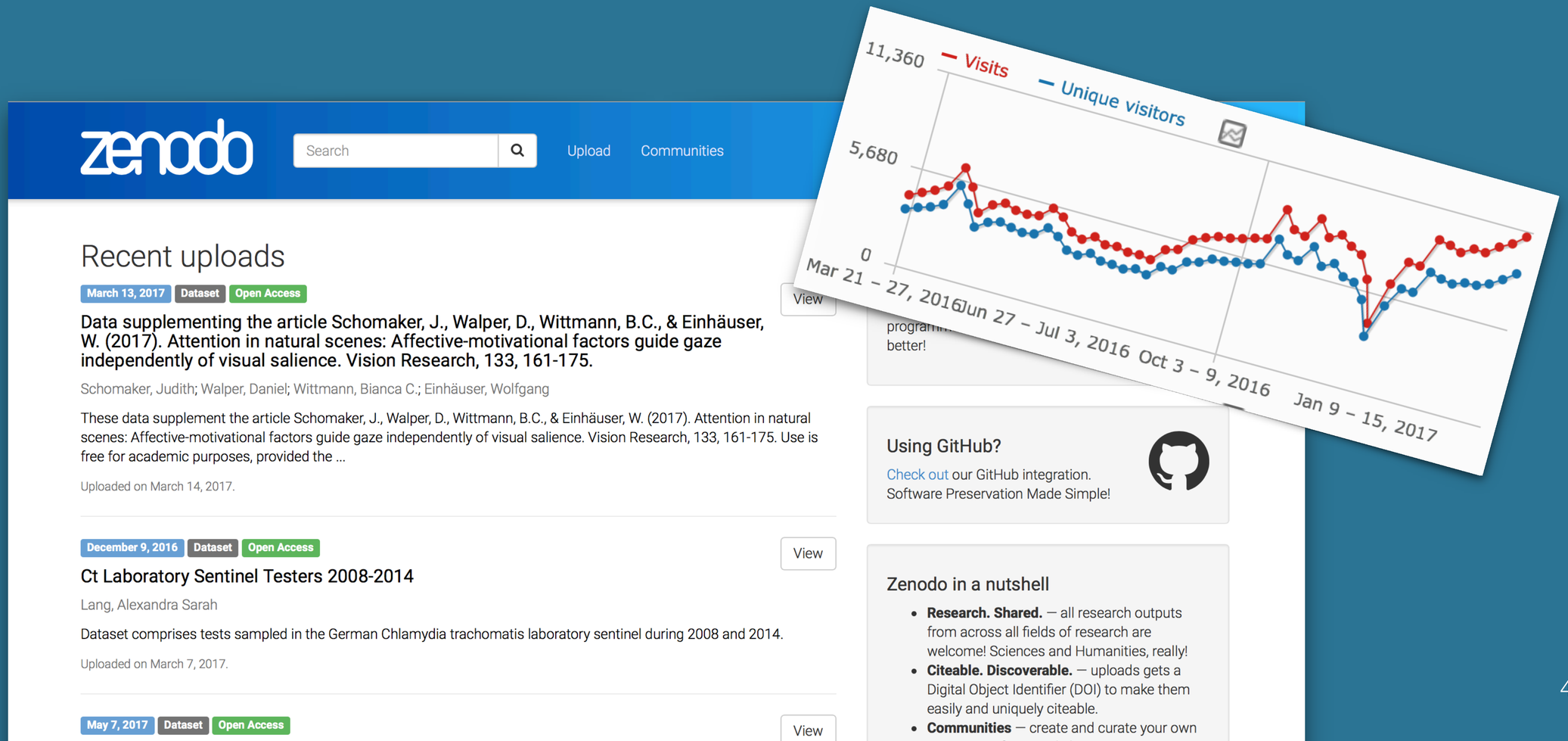
INVENIO

MLZ
Heinz Maier-Leibnitz Zentrum

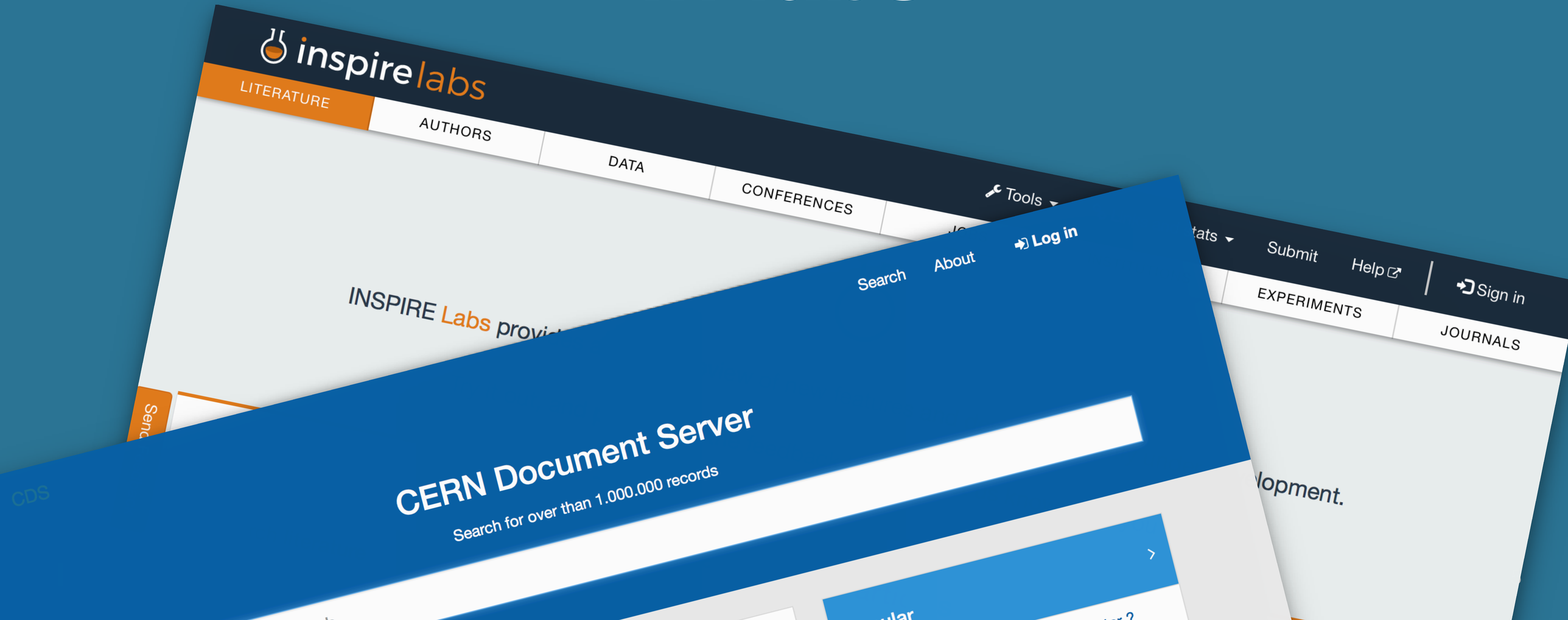

INVENIO
A digital library framework
GitHub · Gitter · PyPI · Read The Docs · Twitter
invenio_screens_intro.jpg
wiki: WikiStart
Search
Login | Preferences | Register
Wiki
Search
Start Page | Index | History
Home
Features
Documentation
Modules
Demo
Download
Mailing Lists
Support
Team
you to run your own digital library or document repository on the web. The technology offered by the software
nagement from document ingestion through classification, indexing, and curation to dissemination. Invenio complies
Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as its underlying bibliographic format.
e of Invenio make it a comprehensive solution for management of document repositories of moderate to large sizes (several
riginally developed at CERN to run the CERN document server, managing over 1,000,000 bibliographic records in high-energy
002, covering articles, books, journals, photos, videos, and more. Invenio is being co-developed by an international collaboration comprising
uch as CERN, DESY, EPFL, FNAL, SLAC and is being used by about thirty scientific institutions worldwide (see demo).
ou would like to try it out yourself, please download our latest version. If you have any questions about the software or the support behind it, please
join our mailing lists or contact us.
Download
version 1.2.1
Learn more
Community


INVENIO
Invenio Digital Library Framework
Build your own fully customised digital library, institutional repository, multimedia archive, or research data repository on the web.
See showcase
Get Started
Invenio is a flexible framework
Integrated Library System
Manage MARC21 authority and bibliographic records. Curate records and run automated quality checks. Use
Research Data
Capture and preserve research output. Harvest datasets analysis code, virtual machine environment, configuration and knowledge information. Visualise data in the browser.
INVENIO
Showcase
Getting Started
Community
Invenio instances around the world
Numerous installations world wide
A showcase gallery illustrating some of the Invenio instances. Add yours!
All
Data
Library
Multimedia
Repository
CERN Document Server
CERN official repository for publications, articles, reports and multimedia content in HEP.
Library
Multimedia
Repository
CERN Open Data
Discover open Research datasets and software of LHC experiments. Visualise events and run and your own analysis.
Data
HEPData
The Durham High Energy Physics Database (HEPData) has been built up over the past four decades as a unique open-access repository for scattering data from experimental particle physics. It currently comprises the data points from plots and tables related to several thousand publications including those from the Large Hadron Collider (LHC).

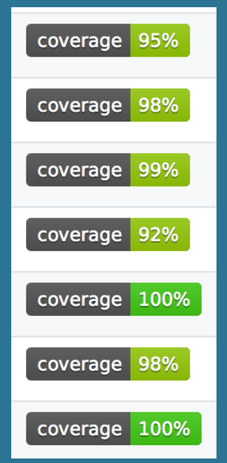
v3
in production

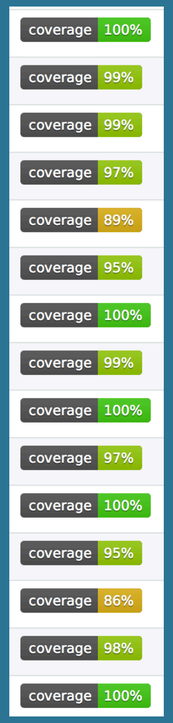


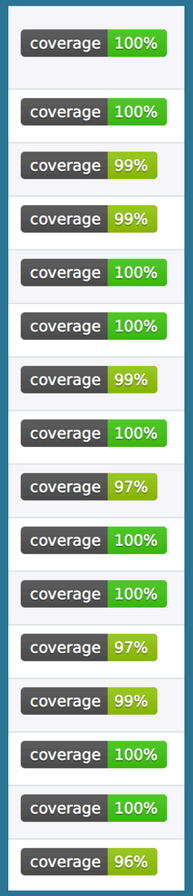
v3
in labs

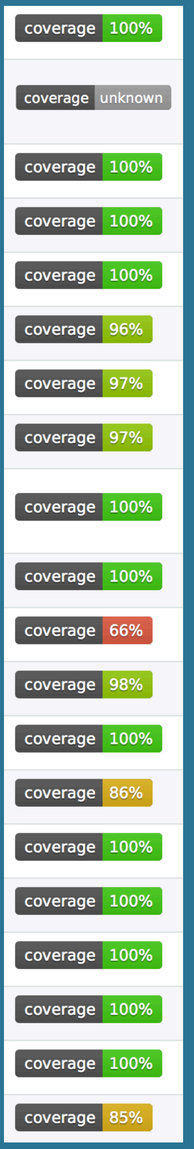

inspire labs
LITERATURE
AUTHORS
DATA
CONFERENCES
Tools
INSPIRE Labs
Submit
Help
Sign in
EXPERIMENTS
JOURNALS
Search
About
Log in
CERN Document Server
Search for over than 1.000.000 records
CDS

v3
in test coverage









INVENIO

Where are we?

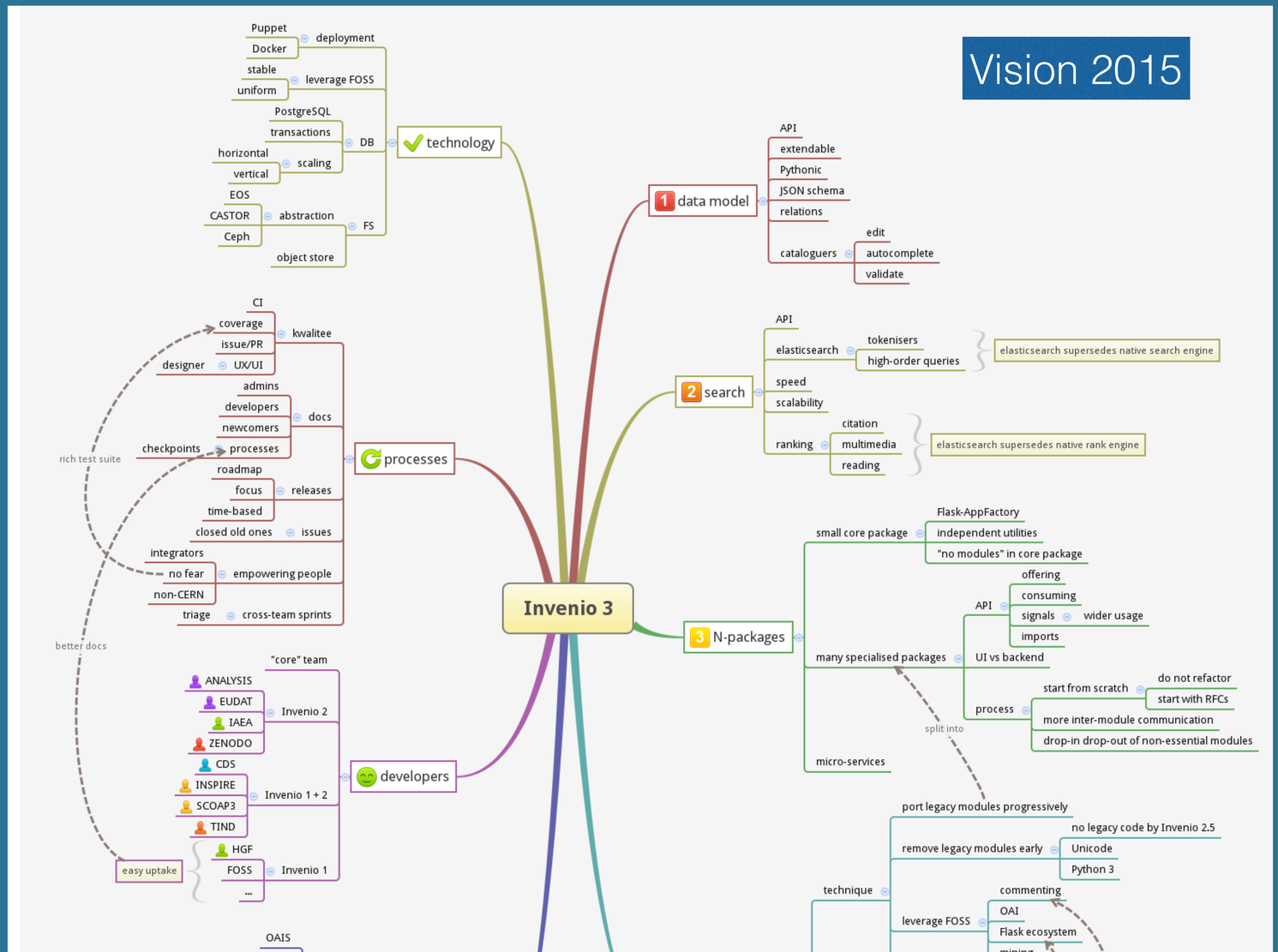
Invenio v1

- Releases: 1.2.2; 1.1.7; 1.0.10
- Feature freeze
- Bug and security fixes only
- v1 -> v3 migration path (revisions + files)
- v1.3? Maintainers and input needed

Invenio v2

DISCLAIMER:
Don’t try this at home!

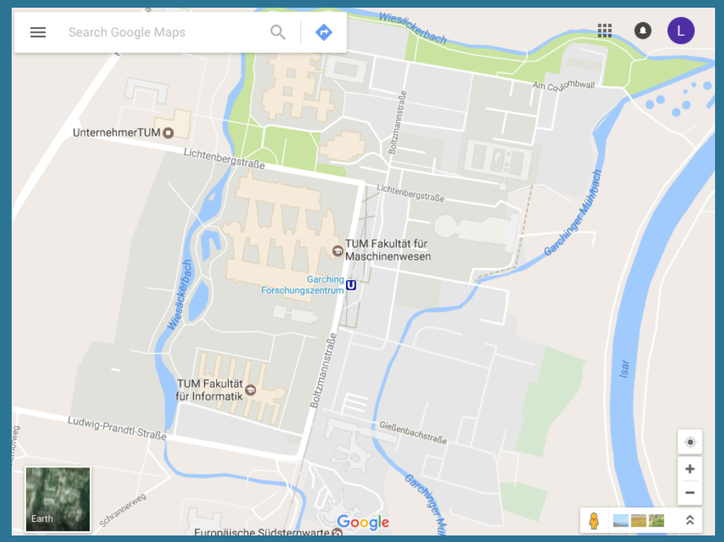
Invenio v3


Vision 2015
Invenio 3
technology
Puppet
Docker
deployment
stable
uniform
leverage FOSS
PostgreSQL
transactions
DB
horizontal
vertical
scaling
EOS
CASTOR
Ceph
abstraction
FS
object store
1 data model
API
extendable
Pythonic
JSON schema
relations
cataloguers
edit
autocomplete
validate
2 search
API
elasticsearch
tokenisers
high-order queries
elasticsearch supersedes native search engine
speed
scalability
ranking
citation
multimedia
reading
elasticsearch supersedes native rank engine
3 N-packages
small core package
Flask-AppFactory
independent utilities
"no modules" in core package
many specialised packages
API
offering
consuming
signals
wider usage
imports
UI vs backend
micro-services
split into
process
start from scratch
do not refactor
start with RFCs
more inter-module communication
drop-in drop-out of non-essential modules
port legacy modules progressively
no legacy code by Invenio 2.5
remove legacy modules early
Unicode
Python 3
technique
leverage FOSS
commenting
OAI
Flask ecosystem
processes
CI
coverage
kwalitee
issue/PR
designer
UX/UI
admins
developers
newcomers
docs
checkpoints
processes
rich test suite
roadmap
focus
releases
time-based
closed old ones
issues
integrators
no fear
empowering people
non-CERN
triage
cross-team sprints
better docs
developers
"core" team
ANALYSIS
EUDAT
IAEA
Invenio 2
ZENODO
CDS
INSPIRE
SCOAP3
TIND
Invenio 1 + 2
HGF
FOSS
Invenio 1
...
easy uptake
OAIS

10.1594/WDCC/EH4_OPYC_SRES_A2

Google Scholar

2002



facebook

2017

v3 code

- Before:
  - Software
  - Monolith
  - Python 2
  - Mix
  - 450k SLOC
  - Low test coverage
- After:
  - Framework (+ Software)
  - Independent packages
  - Python 2.7/3.3+
  - Unicode
  - 160k SLOC
  - 95%+ test coverage

v3 metadata

- Before:
  - MARC21 (almost)
  - 1-5M records
- After:
  - JSON + JSONSchema
  - MARC21 (full)
  - …
  - 10-100M records?

v3
soft spots

Deposit **Access** **SSO**

I18N **JavaScript**

How and when?

Bundles + Flavours

- Base
- Auth
- Metadata

} Minimal v3

- Files
- Statistics
- …

- Integrated library system
- Research data
- Institutional repository
- Media archive

Community

Break the learning curve

Empower the collaboration

Focus the discussions

Where do we go from here?

Credits

- This works is licensed under Creative Commons Attribution 4.0 International except as otherwise noted here:
  - iPod image: CC-BY-SA 2.0 Austria. Link
  - Logos are property of their respective owners.