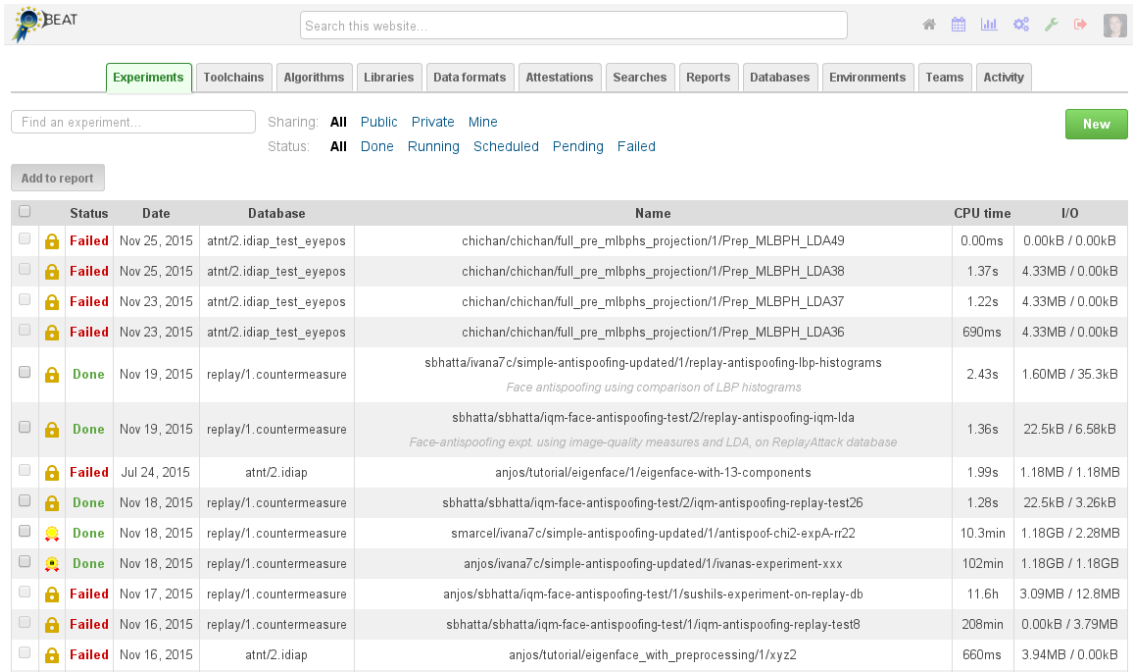
BEAT - Reproducible Open Science

André Anjos & Sébastien Marcel - Biometrics Group

December, 4th. 2015

The BEAT Platform

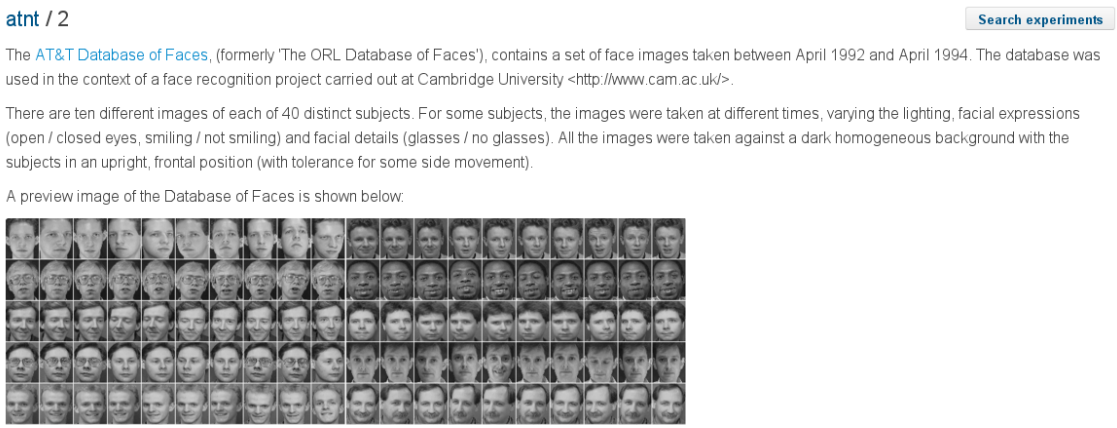


`https://www.beat-eu.org/platform/`

An open-source web-based platform to:

- conduct **reproducible** software-based computer experiments
- **certify** results and preparing publication-ready assets
- coordinate **challenges** with precise objectives
- share and develop code **socially**
- allow access to data, but ensure **privacy** and confidentiality

Workflow



`https://www.beat-eu.org/platform/databases/atnt/2/`

1. Databases and specific usage protocols are input into the computing backend by platform administrators
2. Users create workflows and algorithms to address a particular problem
3. The platform runs user toolchains and algorithms **respecting usage protocols**
4. Results are stored and indexed, so you can create leader boards and keep track advances

Databases

- **train**

image:	**system/array_2d_uint8/1**	Two-dimensional array of 8 unsigned integer values:
file_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:
client_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:

- **templates**

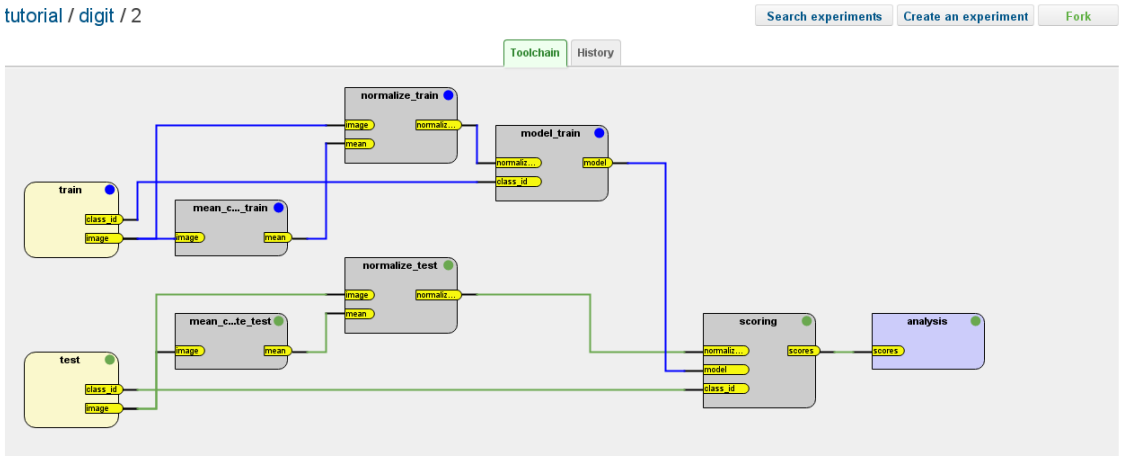
image:	**system/array_2d_uint8/1**	Two-dimensional array of 8 unsigned integer values:
file_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:
client_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:
template_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:

- **probes**

image:	**system/array_2d_uint8/1**	Two-dimensional array of 8 unsigned integer values:
client_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:
file_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:
probe_id:	**system/uint64/1**	Basic format containing only a single unsigned integer (64 bits) value:
template_ids:	**system/array_1d_uint64/1**	Basic format containing a one-dimensional array of unsigned integer (64 bits) values

- Composed of any type of raw data and (at least one) usage protocol
- A protocol defines:
  - how to use the data: e.g. train, validation, test
  - associations between raw and meta data (e.g. *this label belongs to that image*)
- Data is **sequestered** inside the platform (only directly accessible by admins)
- Following the usage protocol, data is fed into user toolchains for processing at a backend
- Present limitation: raw data must be Python-readable

Toolchains



`https://www.beat-eu.org/platform/databases/atnt/2/`

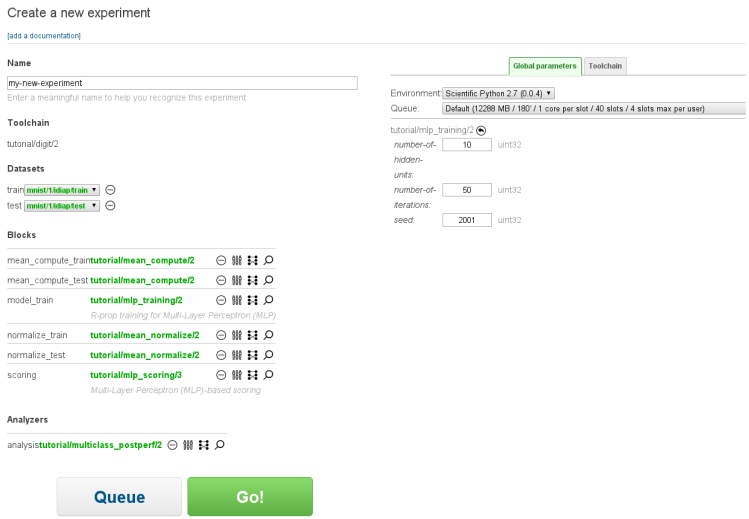
- Express data flow, from the raw data to analysis
- Define possible places where **algorithms** may be implemented
- Versionable, trackable (start from the work of others)
- Easy to draw **at the platform**

Algorithms

`https://www.beat-eu.org/platform/algorithms/tutorial/pca/2/`

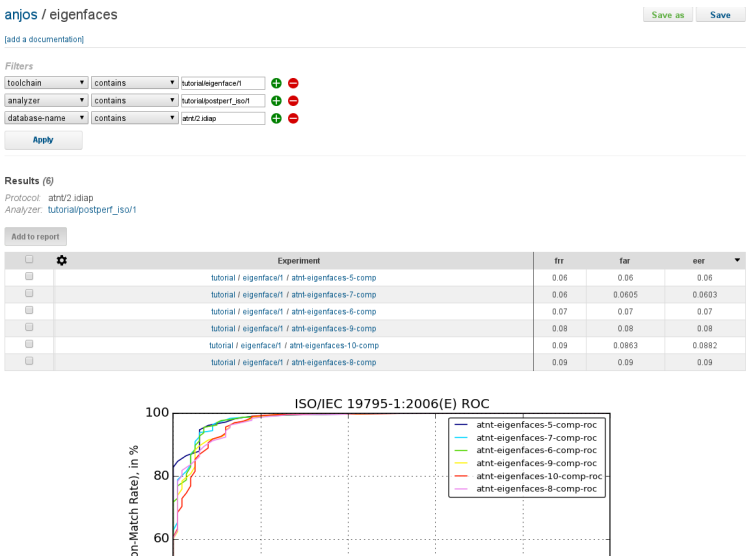
- Implement functionality (a.k.a. *the meat*) of toolchains
- Versionable, trackable, confidential
- Can be *potentially* written in any language
  - Current limitation: A python backend exists
  - Future collaborations: Matlab, Binary, Julia, ...

Experiments



- Where everything comes together: database, protocols, toolchains and algorithms
- Fully automated: only components that fit together can be chosen
- Hybrid by design: block X may run on a Python backend, block Y on another!
- Automatic parallelisation
- Configurable e-mail notification
- Automatic data caching
- Easily reproducible (no hidden parameters, fully tracked)

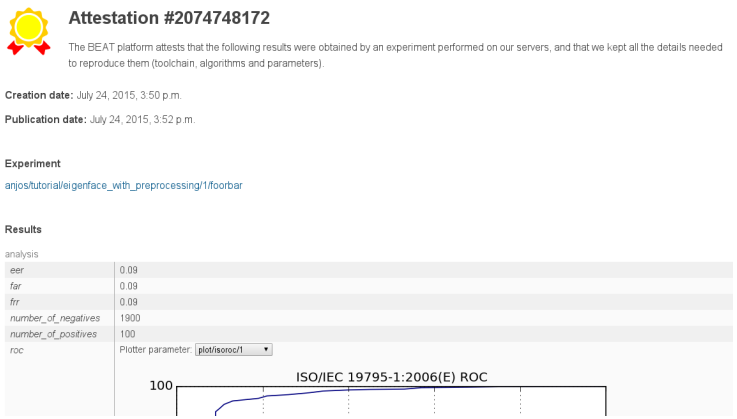
Search and Leaderboards



`https://www.beat-eu.org/platform/search/anjos/eigenfaces/`

- Start searching from anywhere in the site
- User filters to refine your search
- Display plots and rank tables
- Save and re-run a search when necessary
- Upcoming: leaderboards and automatic notifications

How does it compare to others?



Attestation #2074748172

The BEAT platform attests that the following results were obtained by an experiment performed on our servers, and that we kept all the details needed to reproduce them (toolchain, algorithms and parameters).

Creation date: July 24, 2015, 3:50 p.m.

Publication date: July 24, 2015, 3:52 p.m.

Experiment

anjos/tutorial/eigenface_with_preprocessing/1/foorbar

Results

analysis

eer	0.09
far	0.09
frr	0.09
number_of_negatives	1900
number_of_positives	100
roc	Plotter parameter: plot/isoroc/1

BEAT features the following:

- Data: sequestered inside the platform (participants may access the data publicly if owner whishes so)
- Unique certification procedure for results (paper reviewing supported)
- Reproducibility by design: all details of certified experiments are kept (code, parameters, toolchains, datasets, ...)
- Privacy by design: all interactions are kept private until the user wishes to share them
- Open source: create your own platform

The near future

Some features we're currently working on:

- Leaderboards
- Built-in tutorials and user-guides
- First public release: February/2016

The near future

Some features we're currently working on:

- Leaderboards
- Built-in tutorials and user-guides
- First public release: February/2016

Can I try it? **absolutely YES**

- sign up now: `https://www.beat-eu.org/platform/signup/`
- discuss: `https://groups.google.com/forum/#!forum/beat-devel`