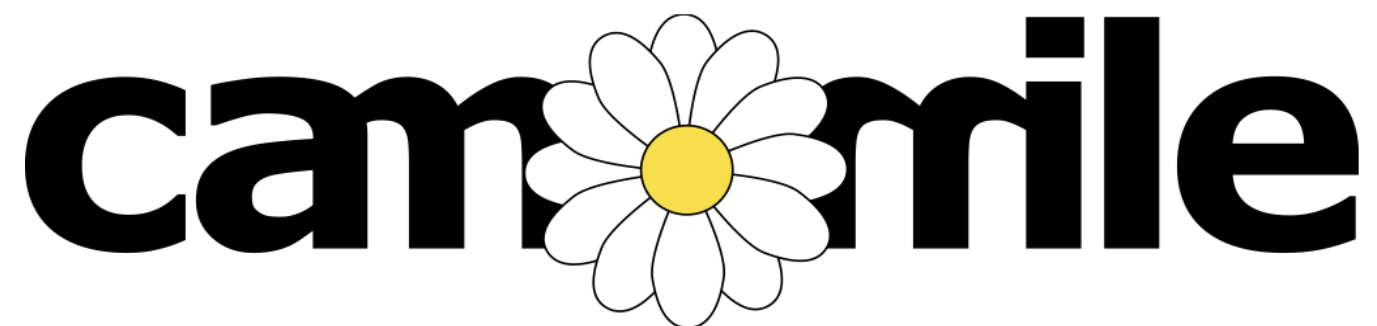
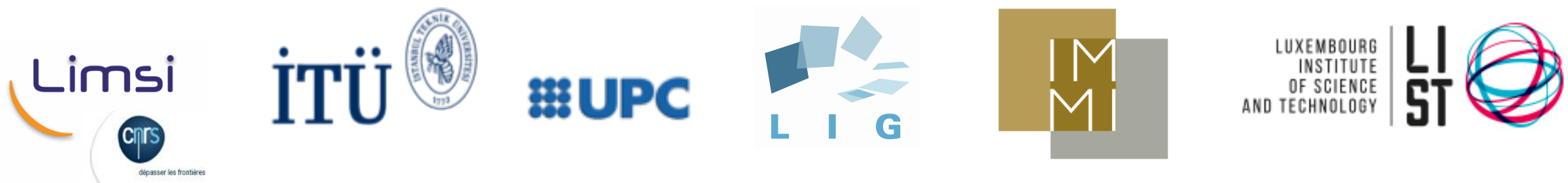




CHIST-ERA Project Seminar

Madrid, March 19, 2015





Project Summary

- Collaborative **A**nnotation of multi-**M**odal, **M**ulti-**L**ingual and multi-m**E**dia documents
- CHIST-ERA call 2011 D2K
Oct. 1, 2012 to March 31, 2016 (42 months)
- 6 participants from 4 countries
France, Luxembourg, Spain, Turkey

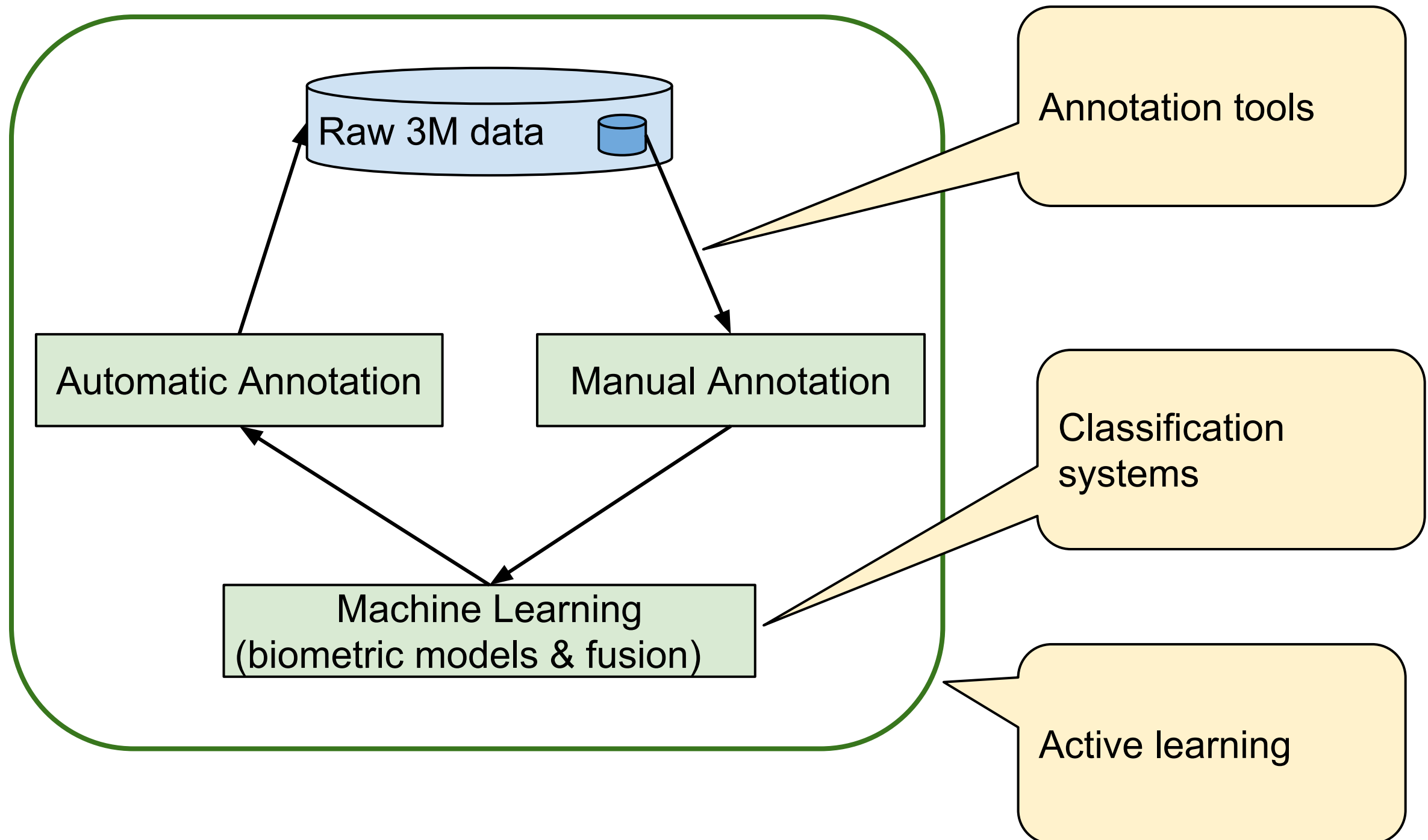


Scientific Background

Scientific Background

- Large amount of heterogeneous, human-generated data available on the web
 - collection and analysis for research (social sciences, linguistics, economics, artificial intelligence & computer science)
- Growing diffusion of embedded, perceptual systems
 - Pattern recognition and machine learning
- Statistical approaches rely on large amount of annotated data
 - human annotation costly
 - several modalities with different levels of abstraction
- Focus on 3M data (multimodal, multimedia, multilingual)
 - share an object of study between scientific domains

Scientific Background





Challenges and Impact

Key challenges and objectives

- Challenges

- Collaborative annotation framework
 - remote annotations combined locally
- Annotation of 3M data
 - improve state-the-art automatic systems
 - integrated framework for the different modalities
- Combine automatic and manual annotations
 - output automatic treatments as help for manual annotation

- Objectives

- Prototype of an experimental platform
- Proof of concept on different use cases

Goals summary

- Go to multilingual / multimodal annotation
 - Annotation for various modalities
 - Distributed environment
 - Collaborative annotation in parallel / in sequence / through crowdsourcing
- Coordination with international initiatives
 - ELDA, LDC, META-SHARE, ISO, OAC...
 - Through an international advisory committee
- Target users for collaborative annotation
 - Researchers in machine learning working on 3M data
 - Institutions aiming at producing manual annotation of 3M data
- Unifying use case: evaluation on multimodal person identification



Consortium



- Spoken Language Processing group
 - 17 academic + support staff, 7 contractuels and post-doct, 15 PhD
- Research topics
 - Speaker characterization in a multimodal context
 - Affective and social dimensions of spoken interactions
 - Perception and automatic processing of variation in speech
 - Robust analysis of spoken language and dialog systems
 - Translation and machine learning
 - Speech recognition
 - Language resources

- Oldest technical university in Turkey (1773)
 - The first faculty on computer and informatics in Turkey
 - Accepts students from top 1% of central university entrance exam
- Involved faculty members:
 - Assoc. Prof. Hazım K. Ekenel (Principal Investigator –Computer Vision)
 - Prof. Zehra Çataltepe (Machine Learning)
 - Asst. Prof. Cenk Demiroğlu (Ozyegin Univ., Speech Processing)



- Universitat Politècnica de Catalunya
- 2 UPC labs
 - Image and video processing group: 11 staff + 1 research assistant + 10 PhD
 - Speech processing: 10 staff + 10 PhD students
- Research lines
 - Video Indexing: Region-based semantic annotation, Face ID, Generic object detection, QbE, Text extraction and recognition
 - 3D scene analysis and coding: Multi-camera reconstruction, body analysis, Tracking, 3DTV coding
 - Speaker, Language, and Emotions recognition, Event detection (audio), Diarization
 - Multilingual and robust speech recognition, Text-to-speech conversion, Speech to Speech Translation



Consortium CNRS - LIG (Grenoble)



- 165 academic and research staff, 260 doctoral and post-doctoral students, 65 support staff
- Two research teams involved (G. Quénot, L. Besacier)
 - GETALP: Study Group for Speech and Natural Language Translation/Processing
 - MRIM: Multimedia Information Modeling and Retrieval Group
- Related research topics (QUAERO consortium & REPERE challenge)
 - Multimodal person recognition in TV broadcast
 - Multimedia semantic indexing
 - Corpus annotation (Collaborative annotation)
 - Evaluation (TRECVID, REPERE)
 - Active learning



Consortium CNRS - IMMI (Orsay)

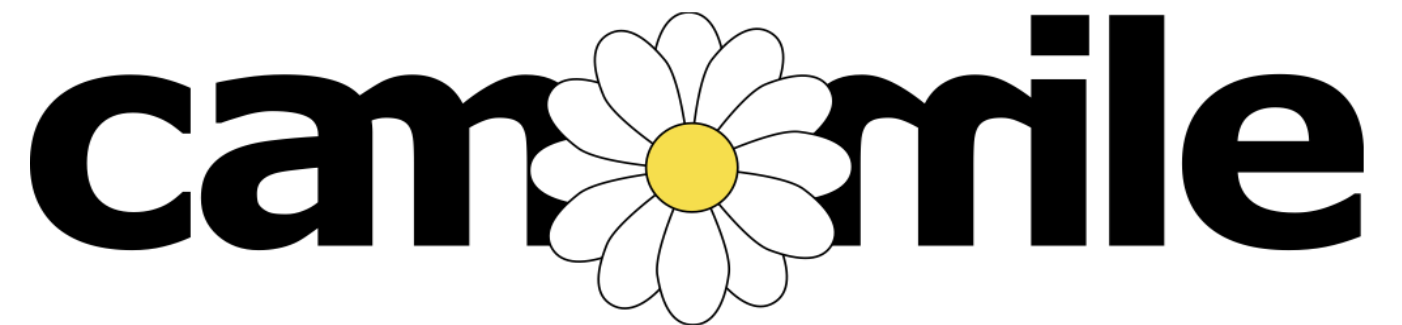


- Institute for Multilingual and Multimedia Information
- French-German Institute
 - Joint venture between LIMSI-CNRS (Paris) + RWTH (Aachen) + KIT (Karlsruhe) (IMMI-Labs)
 - Director : G. Adda, Vice Director : J.J. Mariani
 - Executive Board: A. Waibel, J.L. Gauvain, H. Ney
 - Launched in the framework of QUAERO program
- 10 people at IMMI / 160 people in IMMI-Labs on MMI processing
- Participation in EC projects FLaReNet and META-NET
- Participation in ANR project Bulb and DGA project Papyrus

- Luxembourg Institute of Science and Technologie (LIST)
 - Created 1/1/2015 from Public Research Centers Lippmann & Tudor
 - Applied Research and Technology Transfer: funding from research programs (international, national) & contract research (private, public)
 - Within Environmental Research & Innovation department, eScience Research Unit applies Visual Analytics for scientific discovery and innovative data-driven solutions.
- In Camomile Consortium
 - Expert for a visualization platform that focuses on interpretation, interaction and heterogeneous data
 - Task leader for Cooperative Annotation Monitoring (WP4)
 - Apply the know-how of generic tools developed at LIST: Calluna (advanced visualization) & Cadral (data mining)

Workplan

- **WP1** Management
- **WP2** Multimodal Person Annotation Components
- **WP3** Automatic Processing for Annotation
- **WP4** Collaborative Annotation



WP2 Multimodal Person Annotation Components

WP2 Visual Person Annotation in Video

- ITU

- Improving face tracking performance using confidence and context based cues from face detection outputs
- Face alignment for face matching across pose
- Updating face representation framework & building a new corpus from Harry Potter movies

- UPC

- Research to replace current face tracker by a new method using a region-based particle filter approach.
- Classifiers based on SVM and sparse methods. Robust fusion strategies for video face recognition are being studied.

- i-vector based speaker verification system
- Experiments under noisy environments on WSJ database
- Combination of traditional short-term spectral features with long term both prosodic and voice quality features (jitter, shimmer...) in the speaker diarization front-end.
- Substitution of the traditional generative Gaussian Mixture Models (GMM) of spectral features by novel discriminative Deep Belief Networks (DBN) to classify the i-vectors, the state-of-the-art characterization of speaker.
- An effective global impostor selection method has been designed and now is being transferred from the speaker verification task to the diarization task.

-
- Extended-Multimodal Eigenclassifiers and Selection Criteria for Fusion Methods
 - Active learning on REPERE data set
 - Adapting query by disagreement to unsupervised learning



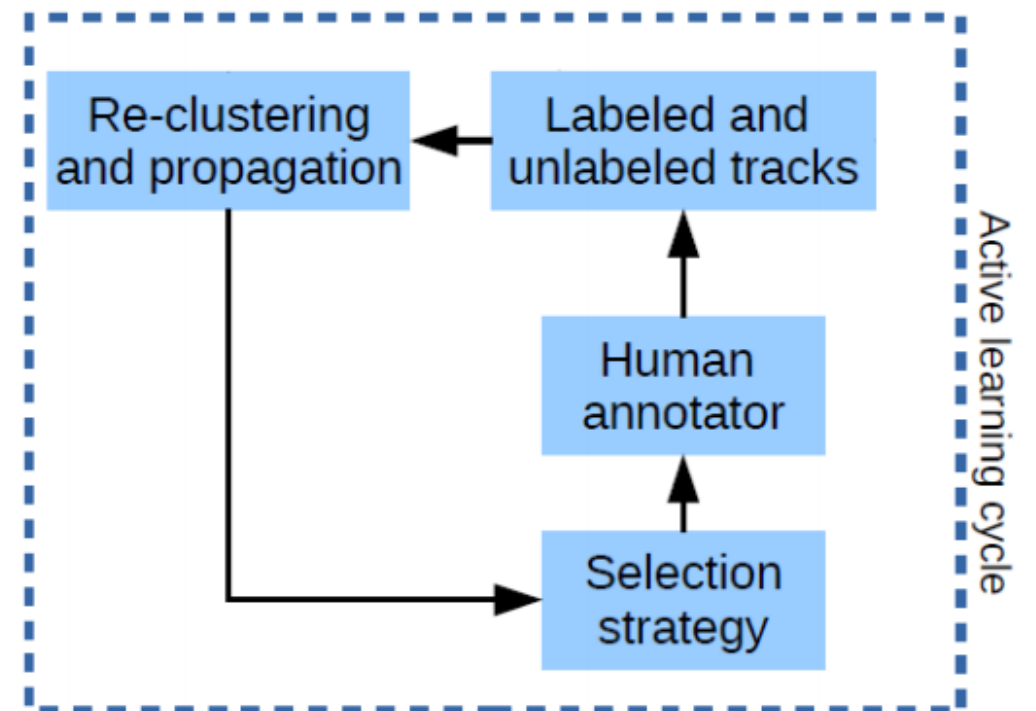
WP3 Automatic Processing for Annotation

WP3 Annotation platform

- Web-based and allows multiple users to work on the same data simultaneously
- Two annotation scenarios are available : speaker annotation and head annotation.
- Use of pre-annotated data: as context or initial annotation
- Dry run:
 - 9 participants
 - Around 1.5 hour spent using the system per person
 - Average time to annotate a single track is 20.2 seconds
 - 716 speech tracks were annotated and 654 were skipped

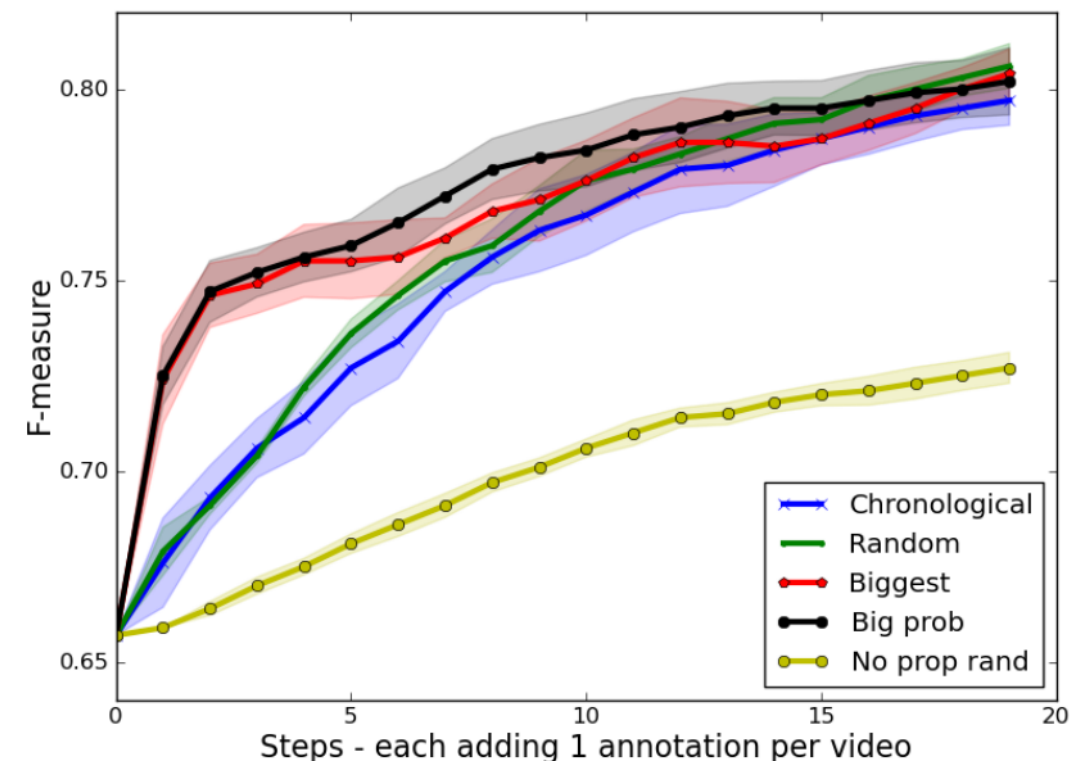
WP3 Retraining and adaptation backend

- The system can be presented as an active learning cycle
- Re-clustering and propagation:
 - Tracks are grouped into individual clusters would correspond to a single person
 - Single or multiple modality cluster
 - Label assigned to a given track is propagated to the entire cluster
- Selection strategy:
 - Based on distance between tracks
 - Size and number of clusters
 - Length of the tracks
 - Purity of the cluster



WP3 Data selection methods for active learning

- Based on different modalities: faces, speech and overlaid names
- Experiments done on the REPERE corpus
- Simulated active learning
- Evaluation done using F-measure
- Several strategies tested, all but one used propagation of labels
- Different strategies for different modalities
- Multimodal effect

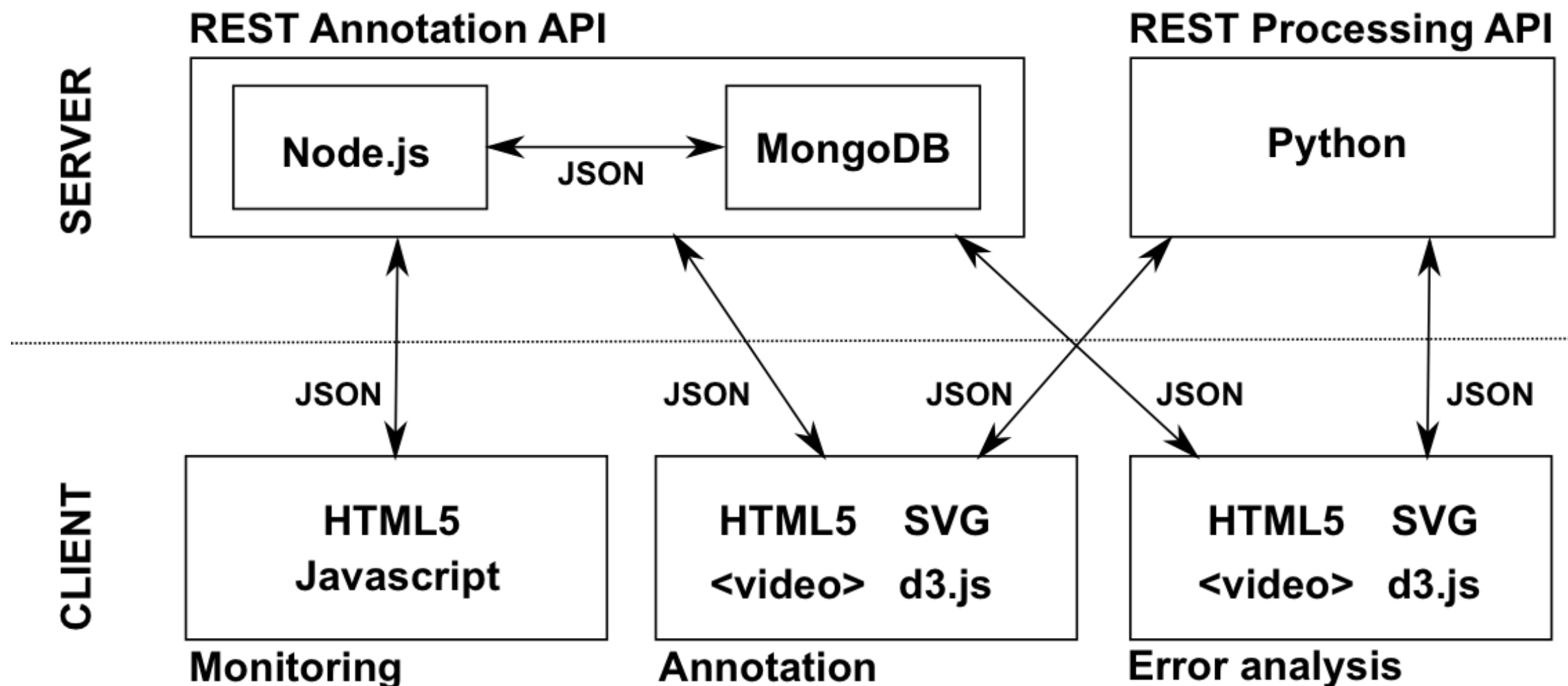




WP4 Collaborative Annotation

WP4 Collaborative Annotation Framework

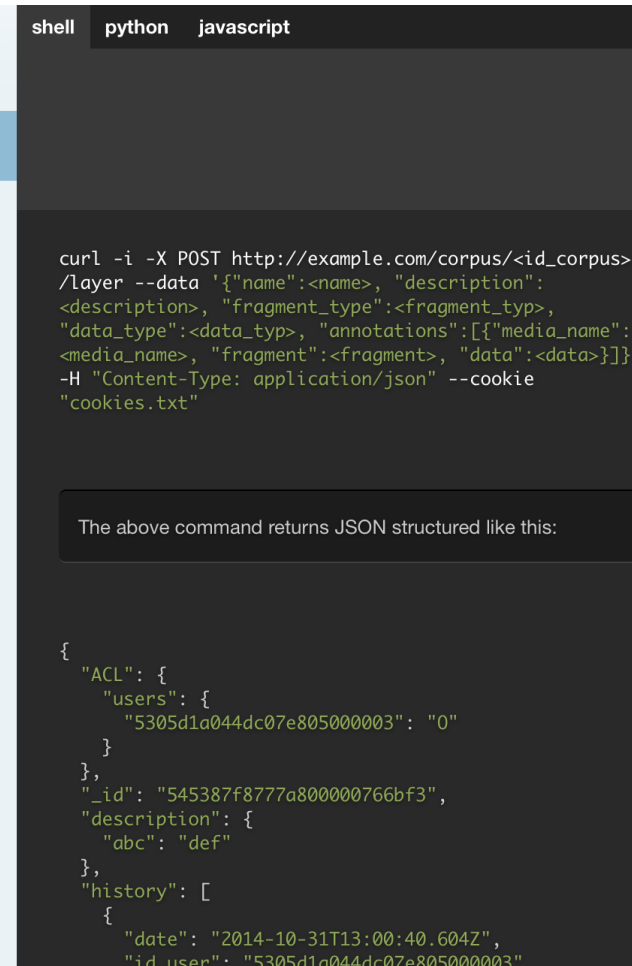
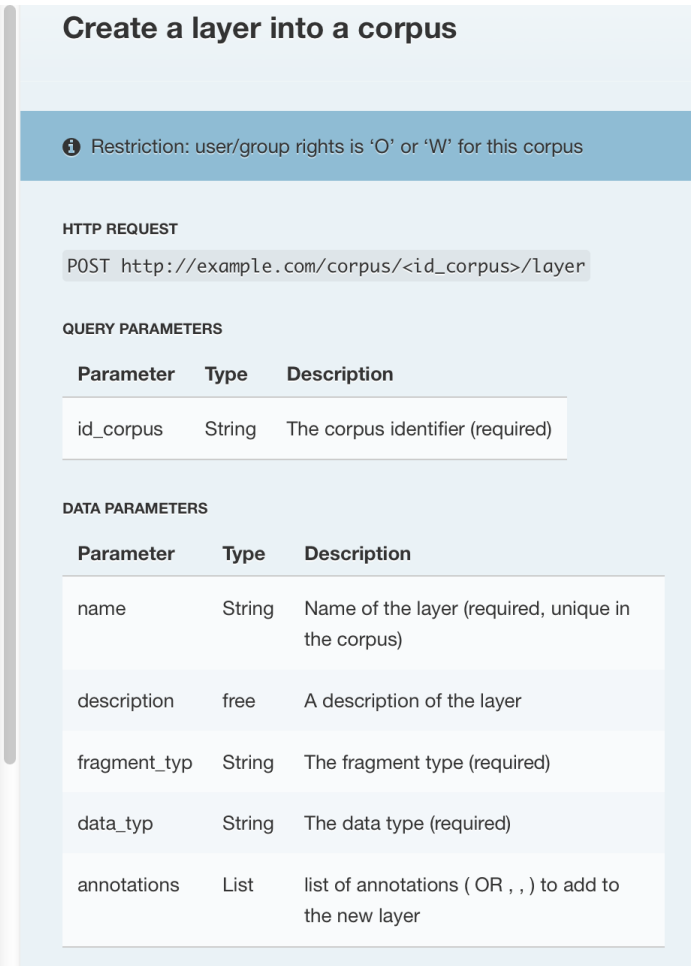
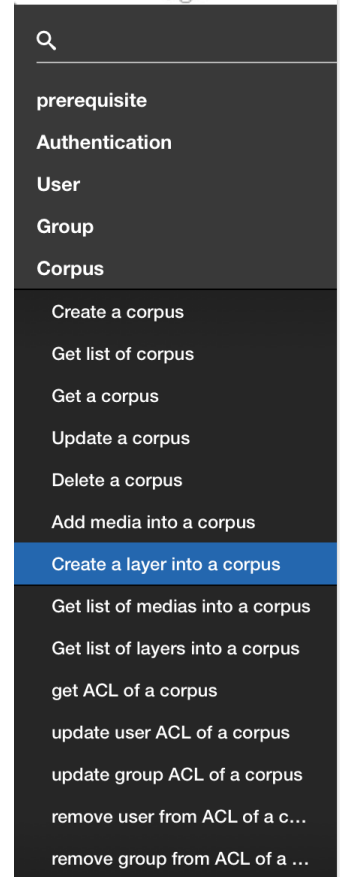
- Interoperability through web services with controlled access to share and perform collaborative annotations



WP4 Collaborative Annotation Framework



- **OPEN-SOURCED** server on GitHub
- **IMPROVED** data model and REST API
- **NEW** access right management
- **NEW** Python and JavaScript clients
- **NEW** online Slate documentation



camomile-project / **camomile-server**

Camomile REST API <http://camomile-project.github.io/camomile-server>

302 commits

3 branches

5 releases



branch: master

camomile-server / +

Merge branch 'hotfix/0.4.1'



hbredin authored 6 days ago

controllers	add filter for get
doc	add spec server 2.0
models	suppression var global + mv fichier
routes	fusion addannotations -> addannotation
.gitignore	ignore doc/build
CHANGELOG.md	fix: version in package.json
Dockerfile	docker: use official node image
LICENSE	chore: MIT license

WP4 Collaborative Annotation Prototype



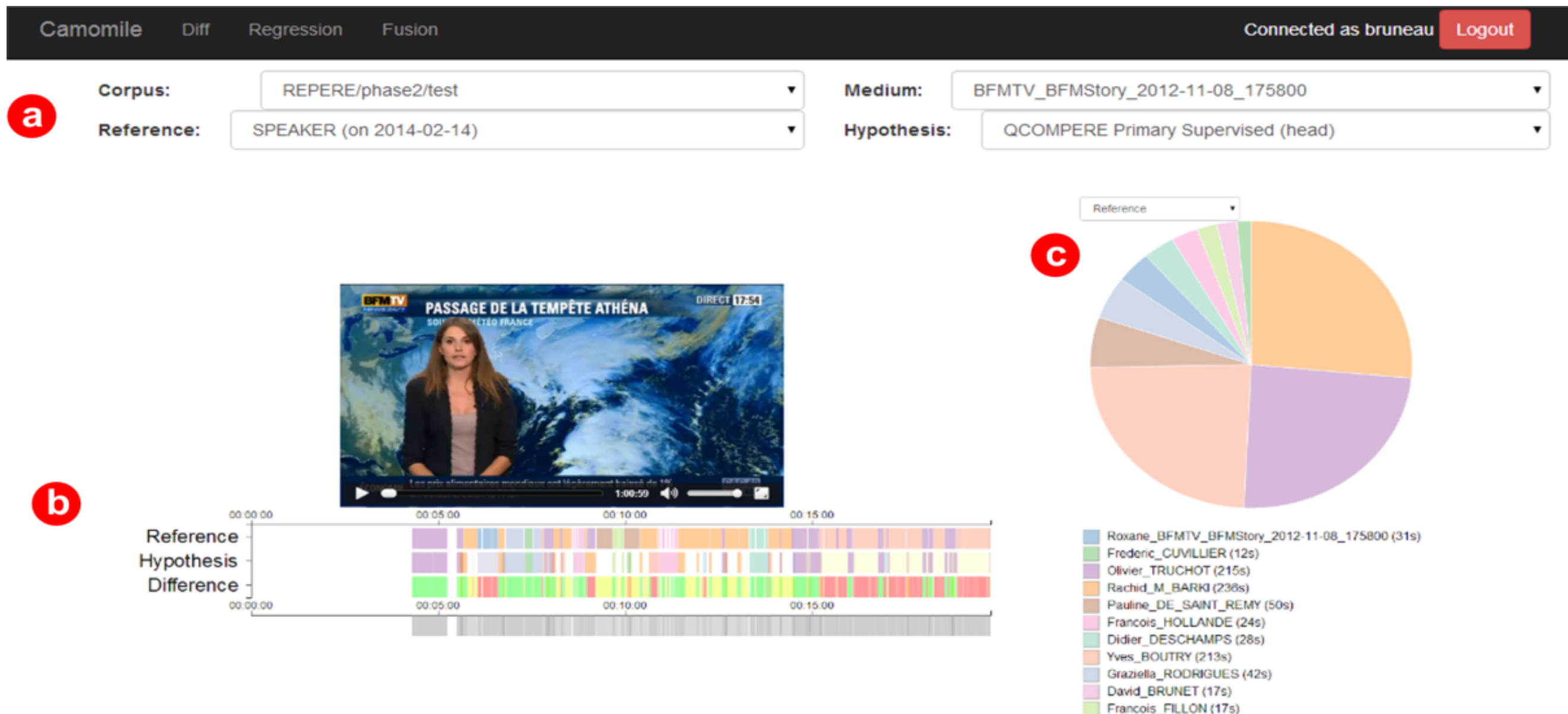
Development + incremental tests

- support a tailored interactive web-based User Interface
- enable interaction, monitoring & evaluation for collaborative annotation
- code modularization in view of re-use of components (e. g. “error analysis” use case adapted to other kinds of annotations)
- manual annotations can be done remotely on many sites, final annotation localized on the main site.

Dissemination

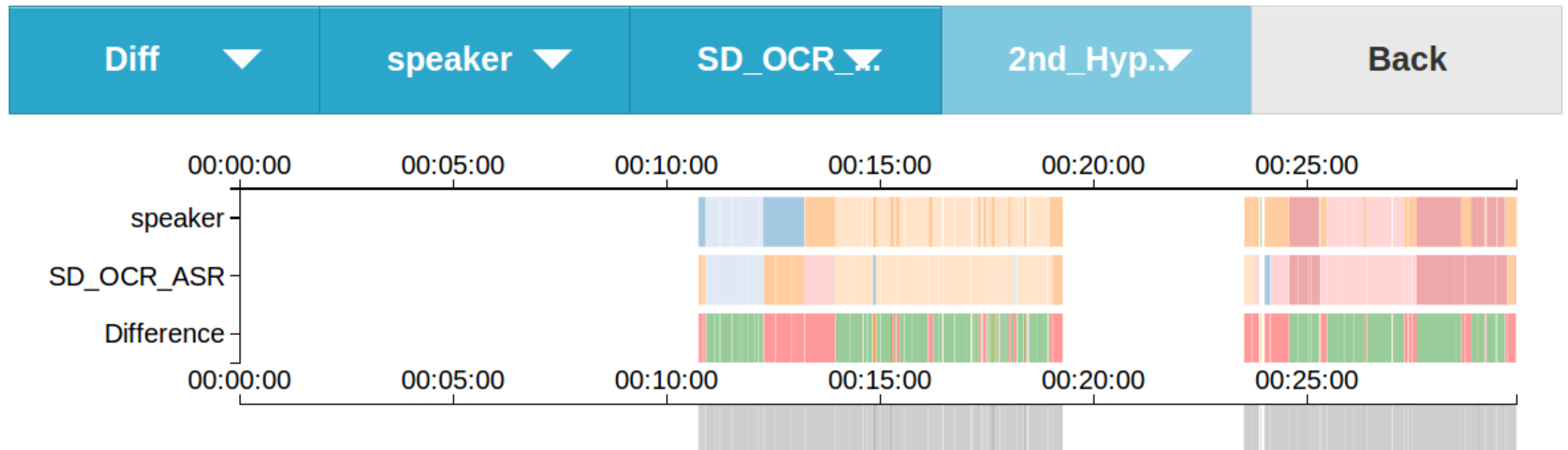
- Close collaboration LIMSI, LIG, and CRP-GL, with physical and remote workshops
- Pilot use of the framework by LIG on uses cases : “speaker recognition” and “face recognition” in MSc course
- Presentations at 2014 international events.

WP4 Collaborative Annotation Prototype



WP4 Annotation Platform Architecture

Full Timeline Component



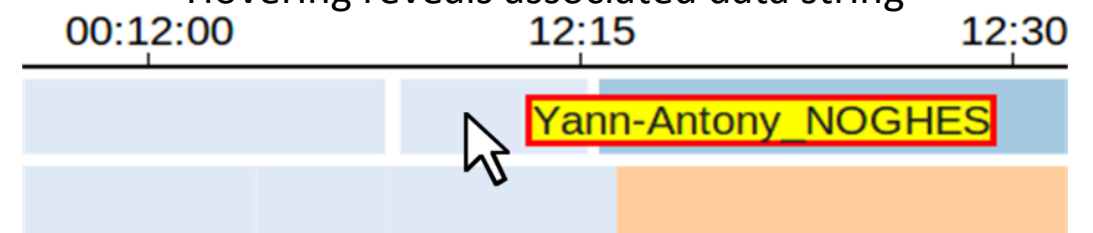
Aggregated View of Annotations

Darker grey = more annotations



Temporal Representation of Annotations

Hovering reveals associated data string



WP4 Guidelines and Guidance

- Annotation Guidelines
 - Definition of annotation guides within a collaborative system
 - Specification of various layers of annotation, in terms of semantic content and formal structure
 - Metadata based on META-SHARE
- Institutional guidance
 - Establish links with ELRA, LNE, LDC, META-SHARE, ISO... and other projects of interest for CAMOMILE
 - Tools and corpora should match needs and requirements of scientific community
 - **Update International Advisory Committee**
 - Khalid Choukri (ELDA/ELRA, France) (+META-SHARE, ISO), Juliette Kahn (LNE, France), **Thomas Schmidt (IDS Mannheim, Germany)**, Marcel Worring (U. of Amsterdam, The Netherlands), Julia Hirschberg (Columbia Univ., USA), Mark Liberman (UPenn and LDC, USA), Liz Shriberg (SRI, USA), **Dominique Estival (MARCS Institute, Australia)**

cammile

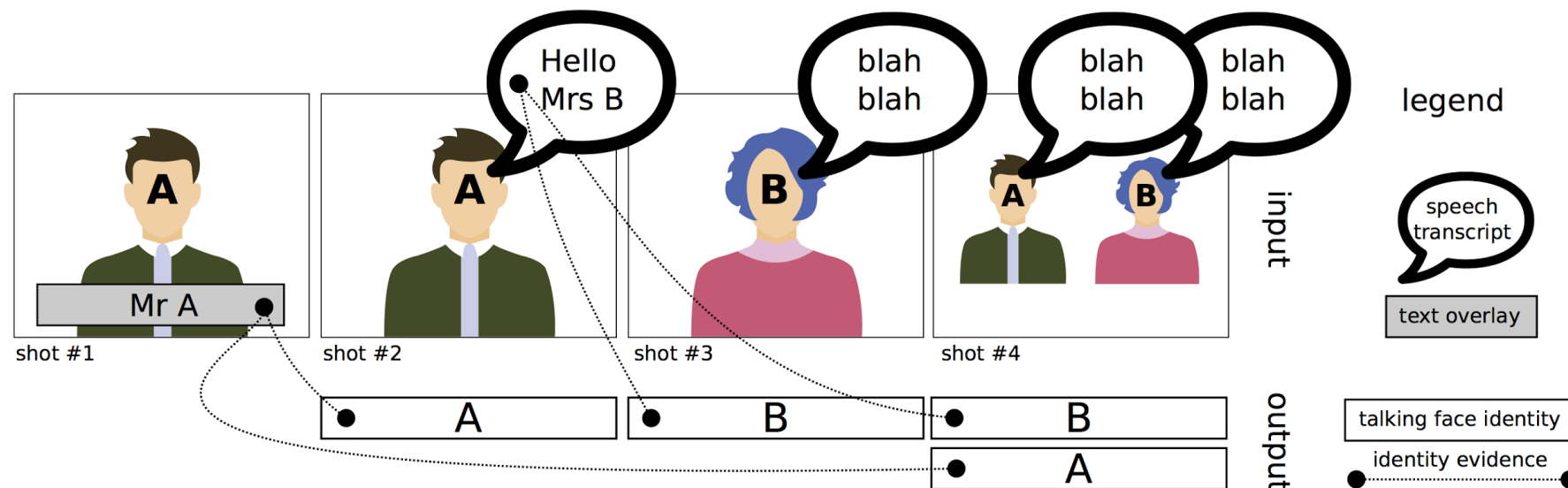


MediaEval transversal use case

New MediaEval task

Multimodal person discovery

- **Multimodal**
 - Identify persons of interest (audible **AND** visible at the same time) within a shot
 - Combine speaker identification, face recognition and speech/overlaid text transcription
- **Person discovery**
 - Discover names in speech and overlaid text transcription
 - **NEW!** Use unsupervised approaches only (no prior biometric models)
- **in TV Broadcast**
 - **NEW!** Extended REPERE corpus (temporally dense annotation into faces by ELDA)
 - **NEW!** INA French TV Broadcast News corpus (100 hours)
 - AGORA Catalan TV Debates corpus (43 hours)



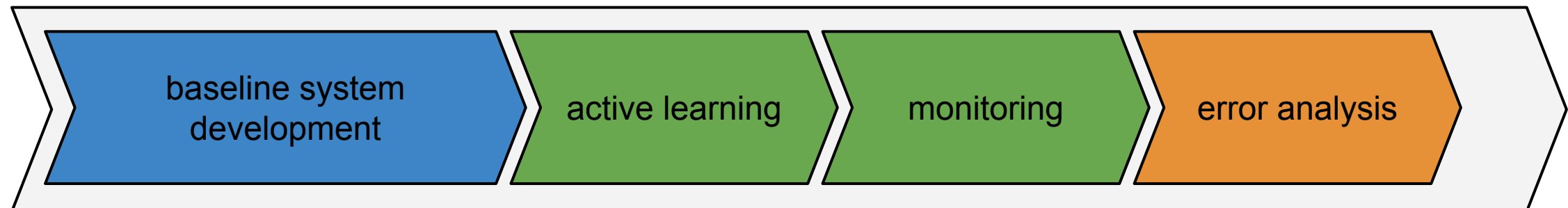
- **NEW!** Information retrieval task
- **NEW!** Distribution of a baseline system (monomodal and fusion components)
- **NEW!** A posteriori collaborative annotation using Camomile framework



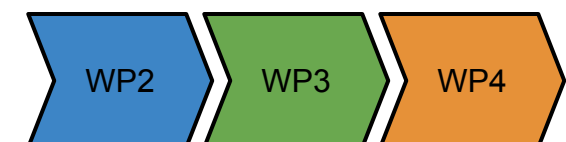
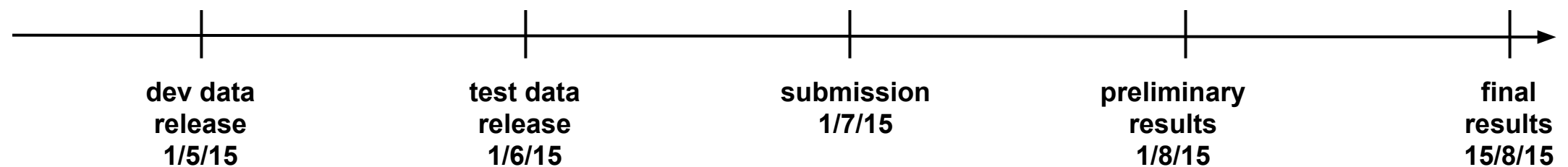
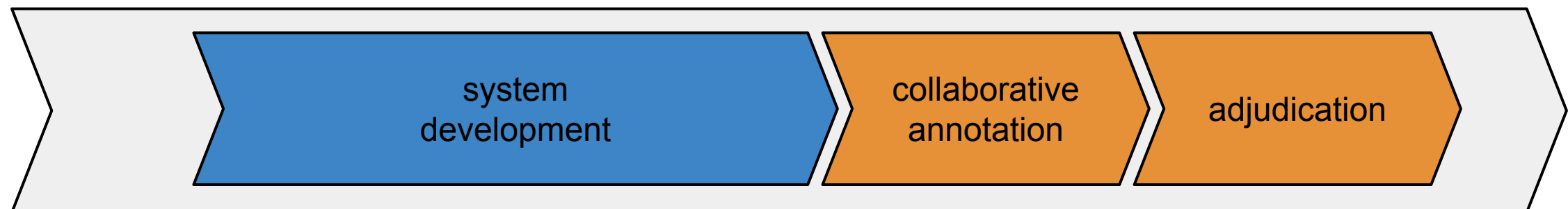
MediaEval

symbiosis

ORGANIZER TIMELINE



PARTICIPANT TIMELINE





Management

Management

- Meetings
 - Kick-off meeting (Feb.'13, Paris)
 - Annual workshops (March'14, Istanbul; March'15, Madrid)
 - CHIST-ERA Project Seminars (March'13, Brussels; March'14, Istanbul)
 - AdCom meetings (Nov. '13, Ermenonville; Feb.'15)
 - Framework architecture meeting (May'13, Luxembourg; Oct'13, Orsay)
 - WP3-WP4 meeting (March'14, Orsay; Oct.'14, Paris)
- Regular Visio/audio conferences
- Wiki and web site + mailing list + server for development

- Meetings
 - 1st AdCom: 20 Nov. 2013 (Ermenonville)
 - 2nd AdCom: 12 Feb. 2015 (visio-conference)
- Main Comments and Recommendations
 - Interest for collaboration (D. Estival)
 - Ahead of initial plans (M. Liberman)
 - Very impressive and interesting report (L. Shriberg)
 - Increase cohesion of partnership (M. Worring)
 - Distribute on various existing platforms
 - Consider proposal submission at next H2020 Call
 - Ensure compatibility with Transcriber
 - Ability to move quickly to different types of data (video, movies)
 - Jointly address at least one common benchmark dataset
 - Evaluate increase of efficiency with human annotators

Planning issues

- Spanish funding was delayed to Jan. 2014
 - with partial cut in budget and duration
- Impact on the project
 - UPC mainly involved in WP2, but whole project slightly delayed (+ 6M)
 - WP2 activity was reduced in 1st year compared to plans
 - UPC could not participate to REPERE challenge in Jan. 2014
- The situation normalized in 2014
 - UPC actively worked in the project and hiring people
 - WP2 deliverables rescheduled

Management

- Financial Reporting

Partner	P.M	Cost	% budget
LIMSI	20	130k€	56%
ITU	68	74k€	64%
UPC	11	20k€	18%
LIG	36	173k€	69%
IMMI	12	79k€	65%
LIST	15	124k€	50%



Scientific Results

Dissemination

- Conferences and workshops

- 10 single-partner communications

ACM ICMR 2015, ICB 2015, IEEE ICIP 2014, IEEE WACV 2014, LREC 2014, SLAM 2014, CBMI 2014, SIU 2013, FUSION 2013

- 5 multi-partner communications

IV 2014, CDVE 2014, VisLR@LREC 2014, WG EGC-BigData 2014, Errare 2013

- Co-organization of ERRARE Workshop series

- “Errors by Humans and Machines in multimedia, multimodal and multilingual data processing”
 - Errare 2013 (Ermenonville, November 2013)
 - Errare 2015 (Sinaia, September 2015)

- International collaboration

- 4-month stay of Pr. Tamura (Japan) at LIMSI/IMMI in 2013

- MediaEval task on person discovery



Sustainability

- Spin-off company from Dr. Ekenel on computer vision in multimedia
- Contacts with content providers: Digiturk, INA, BBC...
- Annotation Guidelines based on META-SHARE
- Collaborative Annotation Framework in Open Source on GitHub



Questions?
