



# PANACEA WP7

## Evaluation of components integration and resources produced

WP leader: CNR-ILC

WP participants: UPF, ILSP, UC, LG,  
DCU, ELDA



# Overview



- WP7 Objectives, Tasks and Deliverables
- Major Strengths of WP7
- Major Challenges
- Addressing the Challenges
- Fall-back strategy
- Workplan 1<sup>st</sup> semester
- Discussion



# WP7 Objectives



Internal evaluation of the platform:

- Evaluation of the integration components
  - Evaluation of components
  - Validation of the integration
- Validation and evaluation of resources produced/information acquired
  - Validation of formats and documentation according to agreed standards/interfaces formats
  - Validation/Evaluation of acquired data



# WP7 Tasks



- Tasks:
  - Task 1 (T1-6): Analysis and definition of the requirements and criteria for evaluation (resp. CNR-ILC)
  - Task 2 (T13-30): Evaluation of the integration of components (resp. ELDA)
  - Task 3 (T13-30): Evaluation of the Resources produced (resp. CNR-ILC)
  - Task 4 (T13-30): Evaluation of the resources for MT (resp. DCU)



# WP7 Deliverables



- Deliverables:
  - D7.1 (t6) Criteria for the evaluation of resources, technology and integration
  - D7.2 (t14) First Evaluation report. Evaluation of PANACEA v1 and produced resources
  - D7.3 (t22) Second Evaluation Report Evaluation of PANACEA v2 and produced resources
  - D7.4: Third Evaluation Report. Evaluation of PANACEA v3 and produced resources



# Major Strengths of WP7



- continuous evaluation → evaluation of each integration cycle
- All partners are involved in the WP
- Most partners have long-standing experience in evaluation and validation → access to evaluation data
- Partners are currently involved in evaluation tasks/campaign
- Direct involvement in other networking and infrastructure projects (FLaReNet, CLARIN, T4ME)



# Major Challenges



- Definition of good criteria for validating integration
- Identification of suitable gold-standards
- Definition of evaluation datasets, baselines and test for the variety of different resources produced
- Set up baselines and evaluation tests for resources in absence of standards of reference
- Integration of evaluation components in the platform



# Addressing the Challenges



- Task 1 specifically dedicated to the definition of relevant evaluation tasks, criteria, methods, and metrics for all the different components and resources produced
- Create datasets and tests for resources if no evaluation tasks exists (->in synergy with other initiatives)
- Propose and organise new evaluation tasks at major evaluation conferences





# Fall-back strategy



- In case of problems, or in the light of new advances in the field, together with WP 3,4,5,6 evaluation tasks may be revised/improved
- Involve communities in joint preparation of new evaluation tasks: FLaReNet, CLARIN?, T4Me,...
- Promote/perform human evaluation, possibly through collaborative means over the web



# Workplan

## 1<sup>st</sup> semester (1/2)



Analysis and definition of the requirements and criteria for evaluation:

### Coordinate with WP 3,4,5,6

- Formats of tools and of interfaces: fix standards and/or interchange formats
  - Define criteria and methodology for evaluating/validating components integration
    - validating formats
    - Validating/evaluating metadata
    - Validating workflows
- ....



# Workplan

## 1<sup>st</sup> semester (2/2)



...

- Identify most relevant evaluation tasks for resource to be produced by PANACEA and collect evaluation data for all the languages involved
- Identify gaps and start constructing missing datasets
- Define baselines for all the different evaluation tasks within PANACEA
- Fix criteria, for first evaluation cycle, in D7.1

WP7 first meeting will be organized after meetings of WPs 3,4,5 and 6.