



PANACEA WP3

The platform

UPF – ILC- ILSP –
LG – DCU - ELDA



Overview



1. WP Details & Deliverables
2. Major strengths of the WP
3. Main Challenges
4. Addressing the Challenges
5. Fall-back strategy
6. Work plan for the First 6 months
7. Questions0



WP3 Details & Deliverables



- The PANACEA LR factory will be build as a web service-based platform
- Web services will form a number of user configured production lines that will vary according to languages, resources to be produced and tools used.



WP3 Details & Deliverables



- WP3 will develop the PANACEA platform: a space of interoperability (standardized formats, protocols, registry and common interfaces, CI) for the easy integration of latest technological components for Automatic Acquisition & Production of LR's.
- WP3 will supply PANACEA production line user tools (workflow editor and engine)



WP3 Details & Deliverables



- Tasks:
- WP3.1 – Architecture and design
- WP3.2 – Workflow editor and engine
- WP3.3 – CI, middleware and temporal files
- WP3.4 – The registry
- WP3.5 – Deployment of web-services



WP3 Details & Deliverables



- D3.1 (t6) Requirement analysis of the platform. Architecture and design of the platform, guidelines for the specification of metadata and common interfaces for the most used web services.
- D3.2 (t14) First version (v1) of the integrated platform and documentation (First Web service integration with WP4 CAA prototype, and WP5 aligners, and metadata and common interfaces for them).
- D3.3 (t22) Second version (v2) of the integrated platform and documentation (v1 + WP4 CAA + WP5 aligners and large data management solutions. Metadata and common interfaces for all the modules integrated).
- D3.4 (t30) Third version (v3) of the integrated platform and documentation (v2 + WP4 PoS modules + WP5 Bilingual Dictionary Extractor + WP5 Transfer Grammar Extractor + WP6 Lexical Acquisition components + WF editor + Registry).



Major strengths of the WP3



- Participant's experience in NLP components and pipelines
- Participant's experience in LT standards and production of LR
- Minimization of integration problems using latest technical proposals
- Highly adaptive for including new proposals (standards or operative)



Main Challenges



1. Definition of CI that cover the majority of NLP tools considered in PANACEA to contribute convincingly to show the usefulness of standardization of formats and interfaces
2. To integrate legacy systems
3. Working with massive data. Web service-based architecture cannot scale as expected



Addressing the Challenges



1. CI & Standards. PANACEA will actively contribute to standardization and interoperativity initiatives to disseminate information about PANACEA's proposals and to gather information about problems as soon as possible.
2. Creation of a specific group to handle legacy problems.
3. Massive data handling included in the requirement analysis



Fall-back strategy



1. UIMA platform (familiar for most of the partners) will be considered a model to follow in case web service-based platform cannot be fully implemented
2. Re-implementation of parts of legacy systems
3. Follow current developments in grid computing such as grid-ftp, creating special temporal files, ...



Work plan for the first 6 months



- Analysis of requirements according to currently defined goals and components.
- Also requirements for the modules to be wrapped as web services
- Work out the architecture and design of the platform accordingly
- First use-case (WP4 CAA and WP5 Aligners): definition of metadata and CI proposals. Evaluation of the methodology and guidelines for other similar exercises
- Detailed work plan for the rest of the project
- First WP3 meeting will be in February



- Questions?
- Thanks!!!

Resource distribution

WP Task	3.1	3.2	3.3	3.4	3.5	total
UPF	2	4	4	4	4	18
ILC	2	4	4	4	4	18
ELDA	2	3	3	3	3	14
LG	2		5		5	12
ILSP	1		2		3	6
DCU	1		2		3	6
	10	11	20	11	22	74



Task Participation in WP3



- UPF (18 pm) – Marc Poch
- ILC (18 pm) – Ricardo Gratta
- ELDA (14 pm) – Olivier Hamon
- LG (12 pm) – Boris Vaisman, Matthias Kistler
- ILSP (6 pm) – Prokopis Prokopidis
- DCU (6 pm) – Antonio Toral, Pavel Pecina