

# META=NET

## Core Language Resources and Technologies

**Stelios Piperidis**

**ATH/ILSP, Greece**  
spip@ilsp.athena-innovation.gr

January 25, 2013 – Berlin, Germany



Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

# Outline

- ❑ Shared Infrastructures, Resources and Technologies as a basis for the successful development and implementation of the priority themes
- ❑ Aggregates common issues appearing in all three priority themes
- ❑ Based on important initiatives, like FLaReNet Blueprint of Actions and Infrastructures, the CLARIN Research Infrastructure, etc.
- ❑ Based also on the domain independent needs (and visions) articulated during the SRA building process (from Vision Group meetings until the SRA )
- ❑ Organised in 6 Research Priorities
  - Infrastructure
  - Coverage, Quality, Adequacy
  - LR Acquisition
  - Openness
  - Interoperability
  - Promotion

# Roadmap

| Priority        | Phase 1: 2013-2014  | Phase 2: 2015-2017  | Phase 3: 2018-2020  |
|-----------------|---|---|---|
| Infra-structure | <ul style="list-style-type: none"> <li>• <b>Maintain and extend facility for LR sharing</b></li> <li>• Promote reliable <b>documentation</b> of LRs through agreed <b>formal metadata models.</b></li> <li>• <b>Integrate web services</b> in LR infrastructures</li> <li>• Adopt <b>persistent LRs identification – ISLRN</b></li> <li>• Ensure <b>cooperation</b> between infrastructures</li> <li>• Identify missing functionalities - prioritise their development</li> <li>• Assess Infrastructure(s)</li> </ul> | <ul style="list-style-type: none"> <li>• Develop mechanisms for <b>automatically extracting and updating MD</b> of as well as actual LRs</li> <li>• Easy conversion into <b>uniform formats</b></li> <li>• Solutions for integrating language processing services to help <b>organic growth of LR infrastructures</b></li> <li>• Identify missing functionalities - prioritise their development</li> <li>• Assess Infrastructure(s)</li> </ul> | <ul style="list-style-type: none"> <li>• Identify missing functionalities and prioritise their development</li> <li>• Assess Infrastructure(s)</li> </ul> |

# Roadmap

| Priority                               | Phase 1: 2013-2014   | Phase 2: 2015-2017   | Phase 3: 2018-2020   |
|--|--|--|--|
| Coverage<br>Quality<br>and<br>Adequacy | <ul style="list-style-type: none"> <li>• Increase <b>quantity of LRs</b> available to address language technology and application needs (cf. priority themes)</li> <li>• <b>Methods &amp; tools to cover new forms of language use/new types of content</b> (social media, user generated content)</li> <li>• Address <b>formal and content quality of LRs</b> by promoting <b>evaluation and validation</b></li> <li>• Promote initiatives for <b>LRs collaborative creation</b></li> <li>• <b>Identify gaps per technology &amp; language</b> – BLARKS for less-resourced languages</li> </ul> | <ul style="list-style-type: none"> <li>• Further increase quantity of LRs available to address language and application needs</li> <li>• Provide <b>high quality LRs for all European languages</b></li> <li>• <b>Coordinated effort at the international level by providing under-resourced languages with examples and best practices</b></li> </ul> | <ul style="list-style-type: none"> <li>• Further increase quantity of LRs available to address language and application needs</li> <li>• Provide <b>high quality LRs for all European languages</b></li> </ul> |
|  | <ul style="list-style-type: none"> <li>• <b>Establish common and standard LRs and Technology validation and evaluation procedures</b></li> </ul>   |  |  |

# Roadmap

| Priority       | Phase 1: 2013-2014  | Phase 2: 2015-2017 | Phase 3: 2018-2020 |
|----------------|---|--------------------|--------------------|
| LR Acquisition | <ul style="list-style-type: none"> <li>• Define and disseminate <b>resource production best practices</b></li> <li>• Enforce <b>recycling, reusing and repurposing</b> (also through appropriate metadata)</li> </ul>   |                    |                    |
|                | <ul style="list-style-type: none"> <li>• Research work towards <b>full automation of LR data production</b></li> <li>• Invest in Web 2.0/3.0 methods for <b>crowdsourcing-based, collaborative creation</b> and extension of high-quality resources, also as a means to achieve better coverage</li> <li>• Implement <b>workflows of language processing services for acquisition of LRs</b> required for the implementation of the priority themes</li> <li>• <b>Bridge LR acquisition methods with linked open data (&amp; big data)</b></li> <li>• <b>Build/adapt basic tools to new data and content types</b></li> <li>• Ensure that <b>Public Sector Information/Data/Content</b> is available for <b>repackaging and repurposing</b>.</li> <li>• Share the effort for production of LRs between international bodies and individual countries</li> </ul> |                    |                    |

# Roadmap

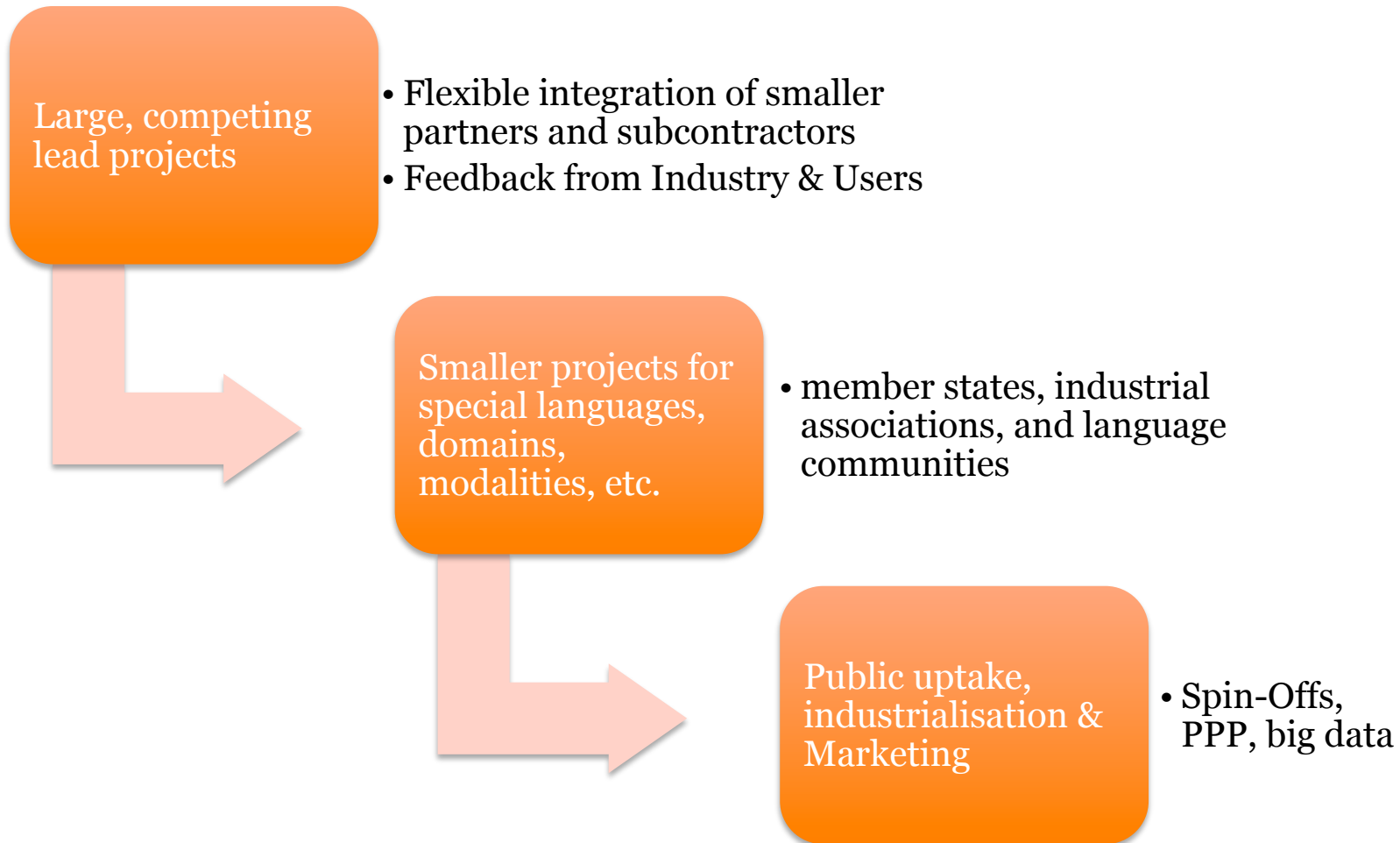
| Priority | Phase 1: 2013-2014  | Phase 2: 2015-2017 | Phase 3: 2018-2020 |
|----------|---|--------------------|--------------------|
| Openness | <p><b>Educate</b> key players with basic <b>legal know how</b></p> <p>Establish <b>agreements for National Corpora</b> for all EU languages</p> <p>Establish <b>agreements with funding bodies</b> (EC/National) for LRT developed in <b>publicly funded projects</b></p>   |                    |                    |
|          | <p>Elaborate specific, simple &amp; harmonised licensing solutions for data resources - Ensure <b>legal interoperability</b></p> <p>Promote <b>copyright exception for research purposes</b></p> <p>Develop <b>legal and technical solutions for privacy protection</b></p> |                    |                    |
|          | <p>Ensure that publicly funded resources are publicly available either free of charge or at a small distribution cost</p> <p>Establish mechanisms for <b>clearing IPR at early stages of production</b>; ensure that <b>re-use is permitted</b></p>                         |                    |                    |
|          |   |                    |                    |

# Roadmap



| Priority                | Phase 1: 2013-2014  | Phase 2: 2015-2017 | Phase 3: 2018-2020  |
|-------------------------|---|--------------------|---|
| <b>Interoperability</b> | Invest in <b>standardisation activities</b> , make <b>standards operational</b> and put them in use   |                    |   |
|                         | <p>Create a permanent <b>Standards Observatory/ Standards Watch</b></p> <p>Set up <b>training initiatives</b> to promote and disseminate standards (tertiary education/ postgrads/ young researchers)</p> <p>Encourage/enforce <b>use of best practices or standards</b> in LR production projects</p> <p>Set up strong <b>liaisons with ISO</b> (in particular TC 37 &amp; SC 35), W3C, ETSI, etc. to promote existing best practices &amp; turn some of them into standards</p> |                    | <p>Identify new mature areas for standardisation and promote joint efforts between R&amp;D and industry</p> |

# Actors & Research Organisation





# Discussion

- ❑ Are there any major missing issues?
- ❑ Do you consider the plan feasible and credible?
- ❑ Are there any obstacles (e.g. some necessary technologies to implement the particular roadmap tasks are still immature) to their implementation?
- ❑ Are there any other prerequisites (e.g. technical, infrastructural, social, legal or ethical factors) that could impede the process?
- ❑ Is the breakdown into the three phases reasonable?
- ❑ Should the 4 roadmaps be linked?

**Q/A**

**META=NET**

**Thank you!**

**[office@meta-net.eu](mailto:office@meta-net.eu)**

**<http://www.meta-net.eu>**

**<http://www.facebook.com/META.Alliance>**