European Language Grid – META-FORUM 2020
Session 4: News from the Language Communities

- Marko Turpeinen (1001 Lakes, Finland)
- Walter Daelemans (University of Antwerp, NCC Belgium)
- Svetla Koeva (Bulgarian Academy of Sciences, NCC Bulgaria)
- Maciej Ogrodniczuk (Polish Academy of Sciences, NCC Poland)
- Marta Villegas (Barcelona Supercomputing Center, NCC Spain)
- François Yvon (LIMSI/CNRS, NCC France)
Synergies between the Spanish Plan-TL & ELG

Marta Villegas (Barcelona Supercomputing Center – Centro Nacional de Supercomputación)
marta.villegas@bsc.es
01/02/03-12-2020 META-FORUM 2020 – Piloting the European Language Grid (virtual conference)
http://www.european-language-grid.eu
The Plan for the Advancement of Language Technology (Plan-TL)

Objective: to promote the development of NLP and MT for Spanish and co-official languages.

Calendar: the Plan-TL was approved in 2015.

Implementation: Organized into ‘subprojects’ (flagship projects) in strategic domains and collaborations with public administrations, universities and research centers and companies.

HEALTH & Platform

2016/19 Commitment CNIO
2019/22 Management assigned to BSC
Health + Platform HPC (6M €/4 years)

https://www.plantl.gob.es
Plan-TL / BSC Platform (January 2020)

- **VM** (virtual machines)
  - Openstack (physical structure)
- **Cloud:**
  - Openstack (physical structure)
  - Ranger 2.0 (cloud environment control)
  - Kubernetes (orchestration)
  - Docker (services deployment)
- **HPC:**
  - Nord3 (updated to SLURM & SO), up to 250 nodes (+ 4000 cores)

**REQUIREMENTS**

- Multi-user platform
- Able to be used in both cloud and HPC environment.
- Potential to accommodate up to thousands of cores with a good level of scalability.
- High availability of the solution.
- Support for scalable components.
- Support for Docker containers.
- Support for Docker component orchestration through tools such as Kubernetes.
- Monitoring, access control, availability of computing capacity, storage and security of the proposed solution.
Plan-TL / BSC Platform

- **VM** (virtual machines)
  - Openstack (physical structure)
- **Cloud**:
  - Openstack (physical structure)
  - Ranger 2.0 (cloud environment control)
  - Kubernetes (orchestration)
  - Docker (services deployment)
- **HPC**:
  - Nord3 (updated to SLURM & SO), up to 250 nodes (+ 4000 cores)

ELG contribution
Plan-TL / BSC Platform

- **VM** (virtual machines)
  - Openstack (physical structure)
- **Cloud**:
  - Openstack (physical structure)
  - Ranger 2.0 (cloud environment control)
  - Kubernetes (orchestration)
  - Docker (services deployment)
- **HPC**:
  - Nord3 (updated to SLURM & SO), up to 250 nodes (+ 4000 cores)
Plan-TL / BSC Data

Up to now, all our data is in Zenodo (44 resources)
Zenodo works well for us as,

- we get a DOI
- it allows large datasets (we are about to publish huge corpora and models)
- it has version control system
- includes Github, OpenAire linking
- ...

But we can also expose and share the resources in ELG (maybe via OAI-PMH)

Thank you!

Marta Villegas (Barcelona Supercomputing Center – Centro Nacional de Supercomputación)
marta.villegas@bsc.es
01/02/03-12-2020 META-FORUM 2020 – Piloting the European Language Grid (virtual conference)
http://www.european-language-grid.eu