

Jak je možno se zapojit?

Jan Hajič (MFF UK)

- Současný stav
 - CEF.AT v začátcích
 - Potřebujeme sbírat data (texty, překlady)
 - Jazyková data „nestárnou“ (příliš) – je třeba a možno začít hned
- Příspěvek veřejných organizací a státní správy
 - Možnosti:
 - Identifikovat webové stránky, kde data jsou ke stažení
 - jsou-li chráněna, domluvit podmínky a způsob předání
 - Identifikovat stávající archívy, domluvit předání, právní aspekt
 - Ideální stav (technologicky)
 - publikované API pro přístup k otevřeným datům
- Jednou to bude třeba udělat stejně...
 - EU Open Data Portal (textové i datové zdroje)

- Uložení do repozitáře (prozatím do Clarin/META-SHARE na <http://lindat.cz>)
 - Rozliší se plně otevřená data (texty) a texty jen pro výzkum/vývoj
 - Metadata
 - Systém postupného doplňování novými texty a překlady
 - API, periodické předávky, ...
- Systémy překladu budou vyvinuty v rámci CEF.AT
 - Podle domény (zdraví, právo, průmysl, zemědělství, ...)
 - V závislosti na (nad)kritickém množství dat
 - Podle párů jazyků
- CEF.AT bude výsledný systém provozovat
 - Po registraci dostupný a využitelný pro všechny úrovně státní správy

- ELRC Technical and Legal Helpdesk
 - <http://www.lr-coordination.eu>
- Právní aspekty
 - Autorské právo, ochrana osobních údajů, (zákon o archivnictví)
- Technické aspekty
 - Uchovávání dokumentů v instituci
 - Přístup, omezení
- [Organizační aspekty]
 - Oběh dokumentů, spisová služba, ...
- Používání systému překladu (ÚFAL MFF UK)
 - Co lze a nelze překládat automaticky
 - Právně, technicky, časově
- Použití jiných technologií analýzy jazyka (ÚFAL MFF UK)
 - Posouzení vhodnosti použití, možnost spolupráce na specifikaci
 - Možnost dlouhodobější spolupráce