

ROZHODNUTÍ

O POSKYTNUTÍ ÚČELOVÉ PODPORY PROJEKTU VELKÉ INFRASTRUKTURY PRO VÝZKUM, VÝVOJ A INOVACE V RÁMCI AKTIVITY PROJEKTY VELKÝCH INFRASTRUKTUR PRO VaVaI

(dále jen „rozhodnutí o poskytnutí účelové podpory“ nebo jen „rozhodnutí“)

Ministerstvo školství, mládeže a tělovýchovy

IČ 00022985

se sídlem Karmelitská 7, 118 12 Praha 1 – Malá Strana

(dále jen „**poskytovatel**“),

vydává podle ustanovení § 14 zákona č. 218/2000 Sb., o rozpočtových pravidlech a o změně některých souvisejících zákonů (rozpočtová pravidla)¹ a ustanovení § 3 odst. 2 písm. d) a § 4 odst. 1 písm. e) zákona č. 130/2002 Sb., o podpoře výzkumu, experimentálního vývoje a inovací z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu, experimentálního vývoje a inovací)²

ve prospěch

Univerzita Karlova v Praze

IČ: 00216208

právní forma: veřejná vysoká škola

se sídlem Ovocný trh 3/5, 116 36 Praha 1

číslo účtu: 3064490217/0100, Komerční banka, a.s., Na Příkopě 33, 114 07 Praha 1

statutární orgán příjemce: prof. RNDr. Václav Hampl, DrSc., rektor

(dále jen „**příjemce**“),

toto rozhodnutí o poskytnutí účelové podpory na projekt velké infrastruktury pro výzkum, vývoj a inovace (dále jen „velká infrastruktura“) s názvem **Projekt LINDAT/CLARIN – Vybudování a provoz českého uzlu pan-evropské infrastruktury pro výzkum** (dále jen „Projekt“)

Článek 1 Předmět rozhodnutí

Předmětem účelové podpory je poskytnutí dotace ze státního rozpočtu na podporu řešení projektu velké infrastruktury Projekt LINDAT/CLARIN – Vybudování a provoz českého uzlu pan-evropské infrastruktury pro výzkum s identifikačním kódem **LM2010013**. Návrh projektu schválený vládou ČR je Přílohou č. 1 tohoto rozhodnutí.

Článek 2 Řešení Projektu

1) Předmět řešení Projektu, jeho cíle, předpokládané výsledky Projektu, způsob jejich dosažení a ověření a vyplněné indikátory monitorování realizace Projektu jsou uvedeny v Příloze č. 1 tohoto rozhodnutí o poskytnutí účelové podpory.

- 2) Příjemce je povinen realizovat Projekt v souladu s rozhodnutím o poskytnutí účelové podpory, podmínkami v něm stanovenými, a plnit veškeré povinnosti z něj vyplývající.
- 3) V případě mezinárodního projektu je příjemce povinen řešením Projektu realizovat účast České republiky v mezinárodním projektu.
- 4) Fyzickou osobou, která je příjemci odpovědná za odbornou úroveň Projektu (dále jen „řešitel“), je **prof. RNDr. Jan Hajič, Dr.**
- 5) Příjemce je povinen zahájit řešení Projektu ve lhůtě do 60 kalendářních dnů ode dne vydání tohoto rozhodnutí. Řešení Projektu je příjemce povinen dokončit do 12/2015.
- 6) Rozhodnutí nabývá právní moci dnem doručení příjemci.

Článek 3

Výše účelové podpory z veřejných prostředků včetně termínů a způsobu jejího poskytnutí

- 1) Poskytovatel poskytne příjemci účelovou podporu formou dotace na podporu Projektu (dále jen „dotace“).
- 2) Celková výše poskytnuté dotace, sestávající z jednotlivých částí dotace pro jednotlivé kalendářní roky řešení Projektu, činí **114 324 tis. Kč** (slovy sto čtrnáct milionů tři sta dvacet čtyři tisíc korun českých).
- 3) Výše jednotlivých částí dotace pro jednotlivé kalendářní roky řešení Projektu, je stanovena v Příloze č. 2 rozhodnutí o poskytnutí účelové podpory. V případě, že dojde v rámci úsporných opatření státního rozpočtu ke krácení účelových výdajů, bude i výše této účelové podpory zkrácena o patřičný podíl.
- 4) Jednotlivé části dotace budou poskytovatelem příjemci poskytovány převodem na účet příjemce.
- 5) Příjemce je povinen užít poskytovatelem mu poskytnuté prostředky dotace výlučně k úhradě uznaných nákladů Projektu a v souladu s jejich časovým určením.
- 6) Nedojde-li v důsledku rozpočtového provizoria podle zvláštního právního předpisu k regulaci čerpání výdajů státního rozpočtu, první část dotace bude poskytnuta příjemci do 60 kalendářních dnů ode dne vydání tohoto rozhodnutí. V následujících letech řešení Projektu bude příslušná část dotace pro příslušný kalendářní rok dle rozhodnutí o poskytnutí účelové podpory poskytnuta příjemci vždy do 60 kalendářních dnů od začátku příslušného kalendářního roku, nedojde-li v důsledku rozpočtového provizoria podle zvláštního právního předpisu k regulaci čerpání výdajů státního rozpočtu.
- 7) Jednotlivé části dotace pro jednotlivé kalendářní roky řešení Projektu dle Přílohy č. 2 tohoto rozhodnutí poskytovatel příjemci poskytne pouze za podmínky, že příjemce řádně plnil všechny své povinnosti vyplývající z rozhodnutí o poskytnutí účelové podpory, mimo jiné, že předložil ve stanovených termínech příslušné zprávy a jiné dokumenty o postupu řešení Projektu a jiné poskytovatelem nebo tímto rozhodnutím stanovené podklady o Projektu a že jsou do informačního systému výzkumu, experimentálního vývoje a inovací zařazeny údaje o realizaci Projektu v souladu se zákonem o podpoře výzkumu, experimentálního vývoje a inovací.

Článek 4

Uznané náklady Projektu

- 1) Uznané náklady Projektu se stanoví ve výši **114 324 tis. Kč** (slovy sto čtrnáct milionů tři sta dvacet čtyři tisíc korun českých) a jsou specifikovány v Příloze č. 2 rozhodnutí o poskytnutí účelové podpory.
- 2) Uznanými náklady Projektu se rozumí takové způsobilé náklady, které poskytovatel schválí jako náklady nutné k realizaci Projektu, které budou vynaloženy během jeho řešení, budou zdůvodněné, prokazatelné a přiřazené k poskytovatelem schváleným činnostem.

- 3) Do uznaných nákladů Projektu lze zahrnout pouze položky uvedené v § 2 odst. 2 písm. 1) zákona o podpoře výzkumu, experimentálního vývoje a inovací. Do uznaných nákladů nelze zahrnout především zisk, daň z přidané hodnoty (pokud by byl příjemce plátcem daně z přidané hodnoty a uplatňoval odpočet této daně nebo jeho poměrnou část), odpisy (kromě nákladů na odpisy majetku odpovídající délce trvání Projektu, tyto odpisy musí být vypočteny pomocí správných účetních postupů), náklady na marketing, prodej a distribuci výrobků, úroky z dluhů a další povinnosti nesouvisející s řešením Projektu.
- 4) Výše uznaných nákladů a s tím související výše účelové podpory z veřejných prostředků poskytnuté na řešení Projektu na celou dobu řešení mohou být každoročně poskytovatelem na základě posouzení průběžné zprávy o realizaci projektu příjemce změněny až o 10 % a na základě kontroly realizace Projektu, která musí proběhnout nejméně jednou za dva roky, může být výše uznaných nákladů nebo výše účelové podpory z veřejných prostředků uvedených v rozhodnutí o poskytnutí účelové podpory změněna až o 50 %. Změny výše uznaných nákladů a s tím související výše účelové podpory musí být zdůvodněné, podložené průběžnými zprávami a kontrolou realizace Projektu. Jakákoliv změna výše poskytnuté dotace musí být podložena změnou rozhodnutí o poskytnutí účelové podpory a musí splňovat podmínky účelové podpory uvedené v zákoně o podpoře výzkumu, experimentálního vývoje a inovací. Na změnu výše uznaných nákladů a na změnu výše účelové podpory nemá příjemce právní nárok.
- 5) Změnu v dílčí skladbě uznaných nákladů uvedených v Příloze č. 2 lze v průběhu roku provést pouze se souhlasem poskytovatele na základě žádosti příjemce, a to do výše 10 % ročních uznaných nákladů, maximálně však do výše 5 mil. Kč během jednoho kalendářního roku.
- 6) Příjemce je povinen vést v souladu se zákonem č. 563/1991 Sb., o účetnictví³ oddělenou evidenci o všech vynaložených výdajích nebo nákladech Projektu a v jejím rámci sledovat výdaje nebo náklady hrazené z poskytnuté podpory na řešení Projektu. Na základě této evidence bude poskytovatelem podpory i místně příslušnými finančními úřady na poskytovateli nezávislými v průběhu řešení Projektu i po jeho ukončení prováděna kontrola. Dále je příjemce povinen uchovávat tuto účetní evidenci po dobu pěti let po ukončení řešení Projektu.

Článek 5

Výkazy uznaných nákladů Projektu

- 1) Výkazy uznaných nákladů Projektu je příjemce povinen předkládat poskytovateli ve formě stanovené nebo písemně schválené poskytovatelem. Pokud není v rozhodnutí o poskytnutí účelové podpory stanoveno jinak, bude výkaz uznaných nákladů přiložen ke každé průběžné zprávě o řešení Projektu a k závěrečné zprávě o řešení Projektu a bude pokrývat stejné období jako příslušná zpráva.
- 2) Příjemce poskytne poskytovateli všechny údaje požadované poskytovatelem za účelem kontroly plnění povinností příjemce vyplývajících z rozhodnutí o poskytnutí účelové podpory.
- 3) Podle ustanovení § 18 odst. 6 až 11 zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách)⁴, a podle ustanovení § 23 a § 26 zákona č. 341/2005 Sb., o veřejných výzkumných institucích⁵, mohou veřejné vysoké školy a veřejné výzkumné instituce převádět do fondu účelově určených prostředků prostředky účelové podpory z veřejných prostředků na výzkum a vývoj, které nemohly být efektivně využity v rozpočtovém roce, ve kterém byly poskytnuty, a to do výše 5 % objemu těchto prostředků poskytnutých na projekt výzkumu a vývoje v daném kalendářním roce. Převod do fondu musí příjemce podle uvedených právních předpisů poskytovateli oznámit listinnou formou a zdůvodnit. Převedené prostředky mohou být použity pouze k účelu, ke kterému byly poskytnuty.
- 4) Příjemce je povinen jednotlivé části dotace na příslušný rok finančně vypořádat nejpozději do 31. 1. následujícího roku, tzn. jednotlivé části dotace ročně zúčtovat a nepoužité prostředky dotace vrátit do státního rozpočtu v souladu s platnými předpisy, především s vyhláškou č. 52/2008 Sb., kterou se stanoví zásady a termíny finančního vypořádání vztahů se státním rozpočtem, státními finančními aktivy nebo Národním fondem⁶, a to předepsaným způsobem, zveřejněným každoročně na webových stránkách výzkumu a vývoje MŠMT.

- 5) Příjemce je povinen spolu se zúčtováním dotace předložit poskytovateli informace listinnou formou o tom, zda na účel, na který byla dotace poskytnuta, použil i finanční prostředky z jiných zdrojů, včetně ostatních veřejných prostředků, z jakých a v jaké výši, a to ve formě stanovené poskytovatelem.

Článek 6 Povinnosti příjemce

- 1) Příjemce je povinen splnit povinnosti stanovené rozpočtovými pravidly, zákonem o podpoře výzkumu, experimentálního vývoje a inovací a dalšími právními předpisy, jakož i povinnosti vyplývající z tohoto rozhodnutí.
- 2) Příjemce je povinen poskytnout poskytovateli veškeré jím požadované doklady a informace, týkající se plnění předmětu rozhodnutí o poskytnutí účelové podpory a využití poskytnuté dotace na řešení Projektu.
- 3) Příjemce je dále povinen
 - a) zajišťovat kontakt poskytovatele s řešitelem,
 - b) informovat poskytovatele o skutečném zahájení prací na řešení Projektu, o každé okolnosti, která by mohla podstatně ovlivnit řešení Projektu a o které se dozví, jako např. o změně dodavatelů, o každé změně forem kontroly (řízení) příjemce a o každé okolnosti, která by mohla ovlivnit podmínky jeho účasti na řešení Projektu,
 - c) přijímat opatření pro řádné provádění svých cílů stanovených v Příloze č. 1 rozhodnutí o poskytnutí účelové podpory,
 - d) respektovat pokyny poskytovatele týkající se obsahu a struktury zpráv a termínů a lhůt pro jejich odevzdání,
 - e) každoročně vypracovat a do 31. ledna následujícího roku předat zástupci poskytovatele průběžnou zprávu o realizaci Projektu včetně vyplněných indikátorů monitoringu realizace Projektu, a to v listinné a v digitální podobě; průběžná zpráva bude obsahovat všechny náležitosti dané Zásadami poskytování účelové podpory projektům velkých infrastruktur pro výzkum, vývoj a inovace (dále jen „Zásady“),
 - f) v roce ukončení řešení Projektu vypracovat a předat poskytovateli do 31. ledna následujícího roku závěrečnou zprávu o realizaci Projektu; tato závěrečná zpráva bude sloužit jako podklad pro závěrečné oponentní řízení organizované Radou pro velké infrastruktury pro výzkum, vývoj a inovace, která předloží výsledky závěrečného oponentního řízení poskytovateli, na jejichž základě bude poskytovatel informovat vládu o realizaci Projektu,
 - g) informovat poskytovatele o vyhlášených veřejných zakázkách dle zákona č. 137/2006 Sb., o veřejných zakázkách⁷, uzavřených smlouvách s dodavateli, jejich významných odchylkách v plnění a uchovávat řádně podepsané originály všech smluv týkajících se řešení Projektu,
 - h) účastnit se jednání, která byla svolána za účelem kontroly, sledování a hodnocení Projektu,

Článek 7 Kontrola průběhu řešení Projektu

- 1) Poskytovatel je oprávněn provádět kontrolu po ukončení řešení Projektu a kdykoli v průběhu jeho řešení na základě předložených průběžných zpráv o realizaci Projektu. Poskytovatel je oprávněn u příjemce provádět kontrolu plnění cílů Projektu, včetně kontroly čerpání a využívání účelové podpory, účelnosti vynaložených nákladů Projektu podle vydaného rozhodnutí o poskytnutí účelové podpory. Povinností příjemce je tuto kontrolu umožnit.
- 2) Kontrola podle odstavce 1 včetně zhodnocení dosažených výsledků a jejich právní ochrany se provádí také vždy po ukončení řešení Projektu, a to na základě předložené závěrečné zprávy o realizaci Projektu a oponentského posudku vypracovaného Radou pro velké infrastruktury pro výzkum, vývoj a inovace. V případě, že doba, po kterou se poskytuje účelová podpora, je delší než

dva roky, je poskytovatel povinen provést kontrolu podle odstavce 1 rovněž každý druhý rok řešení Projektu.

- 3) Finanční kontrola je prováděna zejména podle zákona č. 320/2001 Sb., o finanční kontrole ve veřejné správě a o změně některých zákonů (zákon o finanční kontrole)⁸, a podle vyhlášky č. 416/2004 Sb.⁹
- 4) Osobám provádějícím kontrolu je příjemce povinen poskytnout přístup na pracovištích příjemce k osobám podílejícím se na řešení Projektu, ke všem dokumentům, počítačovým záznamům a zařízením, která přísluší k Projektu. Povinnosti příjemce při finanční kontrole stanoví mimo jiné právní předpisy uvedené v odstavci 3.
- 5) Poskytovatel je oprávněn si pro účely kontroly, sledování a hodnocení řešení Projektu zajistit pomoc nezávislých odborných poradců, jmenovaných poskytovatelem. Poskytovatel odborné poradce písemně zaváže k zachování mlčenlivosti o informacích, které jim budou poskytnuty, a k závazku nezneužít tyto informace ve prospěch svůj nebo třetích osob. Poskytovatel seznámí příjemce s ustavením odborných poradců a umožní příjemci vznést připomínky vůči osobám odborných poradců. Poskytovatel posoudí námitky příjemce a shledá-li je oprávněnými, odvolá jmenovaného odborného poradce a navrhne jiného
- 6) Poskytovatel je oprávněn pozastavit poskytování prostředků dotace vždy až do příštího účtovacího období, pokud mu nebyly příjemcem předloženy podklady k prokázání uznaných nákladů Projektu, průběžná periodická zpráva o realizaci Projektu nebo ostatní podklady ve lhůtách stanovených rozhodnutím o poskytnutí účelové podpory.

Článek 8

Zprávy o realizaci Projektu a doklady o uznaných nákladech Projektu

- 1) Příjemce je povinen předložit poskytovateli ke schválení následující zprávy:
 - a) průběžné periodické zprávy o realizaci Projektu, tj. zprávy o postupu prací, vynaložených prostředcích, případných odchylkách od plánu práce a o dosažených výsledcích za uplynulé období; průběžné zprávy musí obsahovat všechny náležitosti dle Zásad,
 - b) průběžné neperiodické zprávy o dosažení dílčích cílů Projektu, tj. zprávy o jednotlivých výsledcích, u nichž byly zahájeny kroky k zajištění právní ochrany či jejich publikování, či které případně budou jako vlastnické informace předmětem komerčního využití, a to podle jejich povahy,
 - c) závěrečnou zprávu o všech pracích, cílech, výsledcích a závěrech se shrnutím všech těchto uvedených bodů,
 - d) redakčně upravené průběžné periodické zprávy a závěrečné zprávy ve formě vhodné pro poskytovatele a příjemce k publikování; pokud mohou být publikovány úplně periodické a závěrečné zprávy, nebudou vyžadovány redakčně upravené zprávy,
 - e) vyžádá-li si to poskytovatel, další zprávy s informacemi vyžadovanými poskytovatelem.
- 2) Příjemce je povinen respektovat pokyny poskytovatele týkající se obsahu a struktury zpráv a termínů a lhůt pro jejich odevzdání. Zprávy a podklady o uznaných nákladech Projektu, které podle tohoto rozhodnutí předkládá příjemce, se předkládají poskytovateli v 1 vyhotovení.
- 3) Průběžné periodické zprávy o realizaci Projektu a příslušné podklady o uznaných nákladech Projektu budou zahrnovat vždy období jednoho kalendářního roku řešení Projektu, počínaje kalendářním rokem, ve kterém má být dle tohoto rozhodnutí zahájeno řešení Projektu, nevyžádá-li si mimořádně poskytovatel předložení průběžné zprávy příjemcem mimo tuto periodicitu. Každá průběžná zpráva o realizaci Projektu bude předložena nejpozději do 30 kalendářních dnů ode dne ukončení etapy, o které se podává zpráva, nebo do 31. ledna následujícího roku.
- 4) V roce ukončení řešení Projektu na základě závěrečné zprávy o realizaci Projektu poskytovatel prostřednictvím Rady pro velké infrastruktury pro výzkum, vývoj a inovace zorganizuje závěrečné oponentní řízení za účasti tří nezávislých zahraničních expertů a předloží vládě zprávu o realizaci Projektu. Závěrečná zpráva bude mimo jiné obsahovat podrobný přehled finančních prostředků od

počátku realizace Projektu do data ukončení Projektu, a to včetně poskytnuté dotace a zdůvodnění způsobu čerpání všech finančních prostředků Projektu. Tato závěrečná zpráva se předává poskytovateli v listinné formě v 1 vyhotovení v českém a v anglickém znění a současně elektronicky nejpozději do 30 kalendářních dnů ode dne ukončení Projektu, (nejdéle však do 31. ledna následujícího roku).

- 5) Závěrečná zpráva o realizaci Projektu a příslušné podklady o uznaných nákladech Projektu budou zahrnovat celé období řešení Projektu. Závěrečná zpráva o realizaci Projektu bude příjemcem poskytovateli předložena vždy po ukončení řešení Projektu ve lhůtě podle odstavce 4 nebo po dokončení prací, jestliže dojde k předčasnému ukončení Projektu.
- 6) Bude-li řešení Projektu ukončeno před termínem stanoveným v článku 2 odst. 5 tohoto rozhodnutí, platí ustanovení o závěrečné zprávě o realizaci Projektu a o příslušných podkladech o uznaných nákladech Projektu pro období do termínu předčasného zastavení Projektu.

Článek 9 Účast třetích stran

- 1) Dodavatelé, jejichž plnění je potřebné k řešení Projektu, musí být příjemcem vybráni v souladu se zákonem o veřejných zakázkách. Pokud by příjemce pořizoval zboží nebo služby od podnikatelských subjektů, budou tyto pořizeny za tržní ceny obvyklé v místě a čase a nebude možné postoupit část podpory podnikatelským subjektům ani nepřímo.
- 2) Příjemce odpovídá za to, že příjemcem uzavřené smlouvy s dodavateli o účasti na řešení Projektu budou obsahovat ustanovení, dávající poskytovateli stejná práva kontroly dodavatelů ve vztahu k Projektu, jaká má poskytovatel vůči příjemci.
- 3) Návrh projektu je uspořádáním vzájemných vztahů, vč. rozdělení kompetencí a jednotlivých činností mezi příjemce a další účastníky Projektu. Pokud se na řešení Projektu podílí další účastníci, poskytne jim příjemce stanovenou část účelové podpory na základě smlouvy uzavřené mezi příjemcem a dalšími účastníky Projektu.
- 4) Výše účelové podpory dalších účastníků Projektu a její rozdělení v jednotlivých letech je uvedeno v Příloze č. 2 tohoto rozhodnutí.

Článek 10 Poskytování informací, způsob poskytnutí údajů o Projektu a jeho výsledcích pro informační systém výzkumu, experimentálního vývoje a inovací

- 1) Příjemce je povinen zveřejňovat úplné, pravdivé a včasné informace o Projektu a získaných poznacích a jiných výsledcích Projektu.
- 2) Příjemce plní povinnost poskytování informací podle odstavce 1 zejména prostřednictvím poskytovatele, kterému předává údaje o Projektu a údaje o získaných poznacích a jiných výsledcích Projektu ke zveřejnění prostřednictvím informačního systému výzkumu, experimentálního vývoje a inovací ve formě a v termínech stanovených poskytovatelem v souladu s požadavky zákona o podpoře výzkumu, experimentálního vývoje a inovací a nařízení vlády č. 397/2009 Sb., o informačním systému výzkumu, experimentálního vývoje a inovací¹⁰.
- 3) Údaje do centrální evidence projektů (databáze CEP) předkládá příjemce poskytovateli ve formě stanovené poskytovatelem. Poskytovatel předá údaje o projektech podporovaných z jeho rozpočtové kapitoly provozovateli informačního systému výzkumu, experimentálního vývoje a inovací do 50 kalendářních dnů
 - a) od počátku kalendářního roku u projektů zahájených v minulých letech a řešených v daném roce, nebo
 - b) ode dne vykonatelnosti rozhodnutí o poskytnutí účelové podpory nebo vykonatelnosti rozhodnutí o poskytnutí účelové podpory u nově zahajovaných projektů.

- 4) Pokud dojde v průběhu kalendářního roku ke změně údajů předaných podle odstavce 3, poskytovatel předá nové údaje o řešených projektech provozovateli nejpozději do 30 kalendářních dnů ode dne, kdy změna nastala nebo byla poskytovateli oznámena.
- 5) Údaje do rejstříku informací o výsledcích (databáze RIV) předává příjemce poskytovateli ve formě stanovené poskytovatelem. Poskytovatel předá provozovateli údaje o výsledcích projektů a o výsledcích výzkumných organizací podporovaných z jeho rozpočtové kapitoly včetně údajů o jejich zveřejnění a údajů o jejich ochraně podle zvláštních právních předpisů nebo realizování (dále jen „uplatnění“), nejpozději do 150 kalendářních dnů od počátku roku, obdobné údaje o výsledcích projektů nejpozději do 250 kalendářních dnů po ukončení poskytování účelové podpory.
- 6) Příjemce je povinen zveřejňovat výsledky Projektu v souladu s platnými právními předpisy a v informacích zveřejňovaných v souvislosti s tímto Projektem uvádět důsledně identifikační kód Projektu podle centrální evidence projektů a skutečnost, že výsledek byl získán za finančního příspěvní poskytovatele v rámci účelové podpory výzkumu a vývoje.
- 7) Při změně tohoto rozhodnutí o poskytnutí účelové podpory je příjemce povinen předat poskytovateli ve formě a v termínu stanoveném poskytovatelem informace o změně údajů zveřejňovaných v informačním systému výzkumu, experimentálního vývoje a inovací a platné údaje.
- 8) Pokud je předmět řešení Projektu předmětem obchodního tajemství, jiného tajemství nebo utajovanou skutečností podle zvláštního právního předpisu nebo skutečností, jejíž zveřejnění by mohlo ohrozit činnost zpravodajské služby, musí příjemce poskytnout konkrétní informace o Projektu a poznacích a jiných výsledcích Projektu tak, aby byly zveřejnitelné. Pokud je předmět řešení Projektu utajovanou informací, předá příjemce úplné údaje o Projektu a jeho výsledcích zároveň postupem stanoveným zákonem č. 412/2005 Sb., o ochraně utajovaných informací a o bezpečnostní způsobilosti¹¹.

Článek 11

Další povinnosti příjemce

- 1) Příjemce je povinen užít dotaci a její jednotlivé části výlučně k účelu, ke kterému byly prostředky dotace poskytovatelem určeny podle tohoto rozhodnutí, a v souladu s jejich časovým určením. Příjemce je povinen s prostředky dotace nakládat v souladu s právními předpisy a správně, hospodárně, efektivně a účelně; vymezení těchto pojmů obsahuje ustanovení § 2 zákona o finanční kontrole⁸.
- 2) Příjemce souhlasí se zveřejněním svého názvu, sídla, dotačního titulu, výše poskytnuté dotace a závěrečné zprávy o řešení Projektu.
- 3) Příjemce je povinen poskytovatele informovat listinnou formou o každé změně, rozhodné pro poskytování prostředků dotace, do 7 kalendářních dnů ode dne, kdy se o změně dozvěděl.
- 4) Příjemce je povinen zajistit, aby úhrada ostatních osobních nákladů, resp. výdajů, z účelové dotace byla uplatňována pouze příjemcem.
- 5) Příjemce je povinen dodržet maximální přípustný podíl účelové podpory Projektu z veřejných prostředků na uznaných nákladech Projektu, stanovený zákonem o podpoře výzkumu, experimentálního vývoje a inovací.
- 6) Příjemce je povinen uhradit z poskytnutých finančních prostředků příspěvek České republiky do mezinárodního projektu v případě, že je s tímto příspěvkem počítáno v rozpočtu návrhu Projektu schváleného vládou a je uveden v Příloze č. 2.
- 7) Příjemce je povinen pravidelně informovat o čerpání příspěvku do mezinárodního projektu a o výsledcích české účasti v mezinárodním projektu.
- 8) Příjemce nemá žádná přednostní práva na využití českého podílu v mezinárodním projektu.

Článek 12

Odpovědnosti za škodu

Poskytovatel neručí za jednání nebo naopak nečinnost příjemce. Poskytovatel žádným způsobem neodpovídá za nedostatky výrobků nebo služeb, které spočívají na poznacích dosažených v rámci Projektu.

Článek 13

Práva ke hmotnému majetku

Vlastníkem hmotného majetku, nutného k řešení Projektu a pořízeného z poskytnuté dotace, je příjemce, který si uvedený majetek pořídil nebo ho při řešení Projektu vytvořil. Po dobu realizace Projektu a po dobu tří let po skončení Projektu však příjemce není oprávněn bez souhlasu poskytovatele s tímto majetkem žádným způsobem disponovat ve prospěch třetí osoby, tj. není mimo jiné oprávněn bez souhlasu poskytovatele tento majetek zcizit, pronajmout, půjčit, zapůjčit či zastavit.

Článek 14

Práva k výsledkům a využití výsledků Projektu

- 1) Úprava užívacích a vlastnických práv k výsledkům a jejich využití a zpřístupnění mezi příjemcem a uživatelem výsledků, mezi příjemcem a třetími stranami účastnicími se projektu a má-li projekt více příjemců, pak mezi příjemci navzájem, je podmínkou pro poskytnutí účelové podpory. Pokud není tato úprava obsažena v Příloze č. 1, musí ji obsahovat samostatná smlouva o využití výsledků. V případě aplikovaného výzkumu musí mít tato smlouva všechny náležitosti podle ustanovení § 11 zákona o podpoře výzkumu, experimentálního vývoje a inovací.
- 2) Všechna práva k výsledkům patří příjemci. Práva autorů a původců výsledků a majitelů ochranných práv k nim jsou upravena zvláštními právními předpisy¹².
- 3) Pro využití výsledků platí ustanovení § 16 zákona o podpoře výzkumu, experimentálního vývoje a inovací.
- 4) Mohou-li si u příjemce činit nároky na práva k výsledkům třetí osoby, musí příjemce provést taková opatření nebo uzavřít takové smlouvy, aby tato práva byla vykonávána v souladu s jeho vlastními povinnostmi vyplývajícími z tohoto rozhodnutí.
- 5) Příjemce je povinen zajistit, aby výsledky, k nimž má majetková práva a které mohou být využity, byly přiměřené a účinně chráněny, zároveň je však povinen výsledky, ke kterým má majetková práva, využít nebo umožnit jejich využití, a to v souladu se zájmy poskytovatele při respektování nezbytné ochrany práv duševního vlastnictví a mlčenlivosti.
- 6) Postoupí-li příjemce majetková práva k výsledkům třetím osobám, zajistí odpovídajícími opatřeními nebo smlouvami, aby jeho povinnosti přešly na nového nositele práv tak, aby byly zajištěny zájmy poskytovatele vyplývající z tohoto rozhodnutí.
- 7) Využitím se pro účely tohoto rozhodnutí rozumí přímé nebo nepřímé použití výsledků k výzkumným nebo komerčním účelům. Za komerční se považuje přímé nebo nepřímé použití výsledků pro vývoj výrobku nebo technologie a jejich uplatnění na trhu nebo pro koncepci a poskytování služby.
- 8) Je-li výsledkem výzkumu, vývoje a inovací financovaného z veřejných prostředků patentovaný vynález, vztahuje se na něj ustanovení o zaměstnaneckých vynálezech a majitel patentu je povinen učinit nabídku licence podle zvláštních právních předpisů.
- 9) Je-li výsledek Projektu vytvořen v rámci české části mezinárodního projektu, patří výsledek přednostně mezinárodnímu projektu.

Článek 15

Sankce za porušení povinností obsažených v rozhodnutí o poskytnutí účelové podpory

- 1) Poskytovatel může zahájit řízení o odnětí účelové podpory po vydání rozhodnutí o poskytnutí účelové podpory v odůvodněných případech vyjmenovaných v § 15 rozpočtových pravidel.
- 2) Po vydání rozhodnutí o odnětí účelové podpory je příjemce povinen poskytovatelem mu poskytnuté prostředky dotace vrátit bezodkladně v plné výši na účet poskytovatele – tato výše vrácených prostředků může být snížena, stanoví-li tak poskytovatel, o prostředky ve výši úhrady uznaných nákladů Projektu vynaložené za poskytovatelem schválené výstupy (poznatky, podklady) z Projektu, kterých bylo dosaženo před vznikem důvodu pro zahájení řízení o odnětí účelové podpory; dále mohou být uhrazeny, stanoví-li tak poskytovatel, i uznané náklady Projektu, které byly uhrazeny z prostředků dotace v dobré víře a schváleny poskytovatelem před termínem zahájení řízení o odnětí účelové podpory.
- 3) Použije-li příjemce finanční prostředky dotace, poskytnuté mu podle rozhodnutí o poskytnutí účelové podpory poskytovatelem, na jiný účel, než stanoví toto rozhodnutí o poskytnutí účelové podpory, nebo v rozporu s jejich časovým určením, je příjemce povinen tyto neoprávněně použité prostředky dotace vrátit poskytovateli způsobem stanoveným poskytovatelem, a to nejpozději do 30 kalendářních dnů ode dne, kdy poskytovatel takové porušení sjednaného užití prostředků dotace příjemci vytkl a požádal jej o vrácení prostředků dotace.
- 4) Použití prostředků dotace příjemcem na jiný účel, než na jaký mu byla podle tohoto rozhodnutí dotace poskytnuta, nebo v rozporu s jejich časovým určením i jakékoliv jiné použití prostředků dotace v rozporu s podmínkami stanovenými právními předpisy nebo rozhodnutím o poskytnutí účelové podpory se považuje za neoprávněné použití peněžních prostředků poskytnutých ze státního rozpočtu a za porušení rozpočtové kázně. V takovém případě se bude následný postup řídit podle rozpočtových pravidel a podle zákona č. 337/1992 Sb., o správě daní a poplatků¹³.
- 5) Jestliže budou po ukončení řešení Projektu nebo po odnětí účelové podpory při finanční nebo jiné kontrole zjištěny závažné finanční nesrovnalosti nebo nedostatky v informacích nebo dokladech příjemce týkajících se užití dotace nebo budou příjemcem uvedeny nepravdivé údaje, může poskytovatel od příjemce požadovat bezodkladné vrácení všech příjemci již poskytnutých prostředků dotace a příjemce je povinen tyto prostředky vrátit bezodkladně poskytovateli.
- 6) Dojde-li po vydání tohoto rozhodnutí o poskytnutí účelové podpory k vázání prostředků státního rozpočtu, prostředky dotace, které již byly poskytnuté příjemci a užitě jím na úhradu uznaných nákladů Projektu, příjemce nevrací.
- 7) Příjemce je po obdržení rozhodnutí o odnětí účelové podpory povinen provést všechna nezbytná opatření k tomu, aby své povinnosti související s řešením Projektu řádně vypořádal.

Článek 16

Vymezení stupně důvěrnosti údajů

Je-li předmět řešení Projektu předmětem obchodního tajemství, postupuje se v souladu s obchodním zákoníkem¹⁴. Stupeň utajení a označení údajů, které podléhají ochraně podle zákona o ochraně utajovaných informací a o bezpečnostní způsobilosti, je součástí Přílohy č. 1 tohoto rozhodnutí.

Článek 17

Změny rozhodnutí

- 1) Toto rozhodnutí může být změněno pouze novým rozhodnutím poskytovatele.
- 2) Rozhodnutí podle odstavce 1 může být vydáno nejpozději dva měsíce před termínem ukončení řešení projektu.

Článek 18
Závěrečná ustanovení

- 1) Příloha č. 1 Návrh Projektu velké infrastruktury schválený vládou a Příloha č. 2 Detailní rozpočet Projektu a uznané náklady Projektu jsou nedílnými součástmi tohoto rozhodnutí o poskytnutí účelové podpory.

V Praze dne2010

Za poskytovatele: Ing. arch. Kryštof Hajn – I. náměstek ministra

Razítko:

Vysvětlivky:

1. **Zákon č. 218/2000 Sb.**, o rozpočtových pravidlech a o změně některých souvisejících zákonů (rozpočtová pravidla) - Změna: 493/2000 Sb., změna: 141/2001 Sb., změna: 187/2001 Sb., změna: 450/2001 Sb., změna: 320/2001 Sb., změna: 202/2002 Sb., změna: 320/2002 Sb., změna: 479/2003 Sb., změna: 479/2003 Sb. (část), 186/2004 Sb., 257/2004 Sb., změna: 482/2004 Sb., změna: 436/2004 Sb., změna: 1/2005 Sb., změna: 127/2005 Sb., změna: 377/2005 Sb., změna: 361/2005 Sb., změna: 546/2005 Sb., změna: 130/2006 Sb., změna: 230/2006 Sb., změna: 138/2006 Sb., 140/2006 Sb., změna: 112/2006 Sb., změna: 174/2007 Sb., změna: 270/2007 Sb., změna: 26/2008 Sb., změna: 267/2006 Sb., 218/2007 Sb., 306/2008 Sb., změna: 109/2009 Sb., změna: 154/2009 Sb., změna: 214/2009 Sb., změna: 417/2009 Sb., 421/2009 Sb., změna: 267/2006 Sb., 214/2007 Sb., 306/2008 Sb., změna: 139/2010 Sb., změna: 199/2010 Sb., změna: 227/2009 Sb.
2. **Zákon č. 130/2002 Sb.**, o podpoře výzkumu, experimentálního vývoje a inovací z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu, experimentálního vývoje a inovací) - Změna: 41/2004 Sb., 215/2004 Sb., změna: 342/2005 Sb., změna: 413/2005 Sb., změna: 81/2006 Sb., změna: 227/2006 Sb., změna: 171/2007 Sb., změna: 296/2007 Sb., změna: 124/2008 Sb., změna: 110/2009 Sb.
3. **Zákon č. 563/1991 Sb.**, o účetnictví - Změna: 117/1994 Sb., změna: 227/1997 Sb., změna: 492/2000 Sb., změna: 353/2001 Sb., změna: 575/2002 Sb., změna: 437/2003 Sb., změna: 437/2003 Sb. (část), 257/2004 Sb., změna: 669/2004 Sb., změna: 495/2005 Sb., změna: 179/2005 Sb., změna: 81/2006 Sb., změna: 57/2006 Sb., změna: 230/2006 Sb., změna: 264/2006 Sb., změna: 69/2007 Sb., změna: 261/2007 Sb., 296/2007 Sb., 348/2007 Sb., změna: 126/2008 Sb., změna: 304/2008 Sb. (část), změna: 230/2009 Sb., změna: 304/2008 Sb., změna: 227/2009 Sb.
4. **Zákon č. 111/1998 Sb.**, o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách) - Změna: 210/2000 Sb., změna: 147/2001 Sb., změna: 96/2004 Sb., změna: 121/2004 Sb., změna: 473/2004 Sb., změna: 436/2004 Sb., změna: 562/2004 Sb., změna: 342/2005 Sb., změna: 552/2005 Sb., změna: 161/2006 Sb., změna: 310/2006 Sb., změna: 165/2006 Sb., změna: 624/2006 Sb., změna: 362/2003 Sb., změna: 261/2007 Sb., 296/2007 Sb., změna: 189/2008 Sb., změna: 110/2009 Sb., změna: 110/2009 Sb. (část), 419/2009 Sb., změna: 159/2010 Sb.
5. **Zákon č. 341/2005 Sb.**, o veřejných výzkumných institucích - Změna: 533/2006 Sb., změna: 379/2007 Sb., změna: 296/2007 Sb., změna: 110/2009 Sb., změna: 227/2009 Sb.
6. **Vyhláška č. 52/2008 Sb.**, kterou se stanoví zásady a termíny finančního vypořádání vztahů se státním rozpočtem, státními finančními aktivy nebo Národním fondem.
7. **Zákon č. 137/2006 Sb.**, o veřejných zakázkách - Změna: 110/2007 Sb., změna: 296/2007 Sb., změna: 76/2008 Sb., změna: 124/2008 Sb., změna: 110/2009 Sb., změna: 41/2009 Sb., 417/2009 Sb., změna: 227/2009 Sb., změna: 179/2010 Sb. (část).
8. **Zákon č. 320/2001 Sb.**, o finanční kontrole ve veřejné správě a o změně některých zákonů (zákon o finanční kontrole) - Změna: 320/2002 Sb., změna: 123/2003 Sb., změna: 421/2004 Sb., změna: 482/2004 Sb., změna: 426/2003 Sb., 626/2004 Sb., změna: 377/2005 Sb., změna: 377/2005 Sb. (část), změna: 138/2006 Sb., změna: 342/2006 Sb., změna: 298/2007 Sb., změna: 227/2009 Sb.
9. **Vyhláška č. 416/2004 Sb.**, kterou se provádí zákon č. 320/2001 Sb., o finanční kontrole ve veřejné správě a o změně některých zákonů (zákon o finanční kontrole), ve znění zákona č. 309/2002 Sb., zákona č. 320/2002 Sb. a zákona č. 123/2003 Sb.
10. **Nařízení vlády č. 397/2009 Sb.**, o informačním systému výzkumu, experimentálního vývoje a inovací.
11. **Zákon č. 412/2005 Sb.**, o ochraně utajovaných informací a o bezpečnostní způsobilosti - Změna: 119/2007 Sb., změna: 177/2007 Sb., změna: 296/2007 Sb., změna: 32/2008 Sb., změna: 124/2008 Sb., 126/2008 Sb., změna: 250/2008 Sb., změna: 41/2009 Sb., změna: 227/2009 Sb.
12. Například zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), zákon č. 527/1990 Sb., o vynálezech a zlepšovacích návrzích, ve znění pozdějších předpisů, zákon č. 529/1991 Sb., o ochraně topografií polovodičových výrobků, ve znění pozdějších předpisů, zákon č. 478/1992 Sb., o užitných vzorech, ve znění pozdějších předpisů, zákon č. 206/2000 Sb., o ochraně biotechnologických vynálezů a o změně zákona č. 132/1989 Sb., o ochraně práv k novým odrůdám rostlin a plemenům zvířat, ve znění zákona č. 93/1996 Sb., zákon č. 408/2000 Sb., o ochraně práv k odrůdám rostlin a o změně zákona č. 92/1996 Sb., o odrůdách, osivu a sadbě pěstovaných rostlin, ve znění pozdějších předpisů, (zákon o ochraně práv k odrůdám), ve znění pozdějších předpisů.
13. **Zákon č. 337/1992 Sb.** (Zákon České národní rady), o správě daní a poplatků - Změna: 35/1993 Sb., změna: 157/1993 Sb., změna: 302/1993 Sb., 315/1993 Sb., 323/1993 Sb., změna: 85/1994 Sb., změna: 255/1994

Sb., změna: 59/1995 Sb., změna: 118/1995 Sb., změna: 323/1996 Sb., změna: 61/1997 Sb., změna: 242/1997 Sb., změna: 168/1998 Sb., změna: 91/1998 Sb., změna: 29/2000 Sb., změna: 159/2000 Sb., změna: 227/2000 Sb., změna: 218/2000 Sb., 367/2000 Sb., 492/2000 Sb., změna: 120/2001 Sb., změna: 271/2001 Sb., změna: 320/2001 Sb., změna: 226/2002 Sb., změna: 320/2002 Sb., změna: 322/2003 Sb., změna: 354/2003 Sb., 438/2003 Sb., 479/2003 Sb., změna: 440/2003 Sb., změna: 19/2004 Sb., 237/2004 Sb., změna: 254/2004 Sb., změna: 436/2004 Sb., změna: 554/2004 Sb., změna: 215/2005 Sb., změna: 444/2005 Sb., změna: 501/2004 Sb., 179/2005 Sb., 381/2005 Sb., 413/2005 Sb., změna: 56/2006 Sb., 62/2006 Sb., změna: 70/2006 Sb., 79/2006 Sb., změna: 230/2006 Sb., změna: 115/2006 Sb., 267/2006 Sb., změna: 214/2006 Sb., změna: 230/2006 Sb. (část), změna: 270/2007 Sb., změna: 261/2007 Sb., 296/2007 Sb., změna: 122/2008 Sb., 126/2008 Sb., změna: 254/2008 Sb., změna: 305/2008 Sb., změna: 301/2008 Sb., 7/2009 Sb., změna: 158/2009 Sb., změna: 285/2009 Sb., změna: 304/2008 Sb., 41/2009 Sb., 218/2009 Sb., 304/2009 Sb.

14. **Zákon č. 513/1991 Sb., obchodní zákoník** - Změna: 600/1992 Sb., změna: 264/1992 Sb., 591/1992 Sb. (s přihlédnutím k 4/1993 Sb.), změna: 286/1993 Sb., změna: 156/1994 Sb., změna: 84/1995 Sb., změna: 142/1996 Sb. (část), změna: 94/1996 Sb., změna: 142/1996 Sb., změna: 77/1997 Sb., změna: 15/1998 Sb., změna: 165/1998 Sb., změna: 356/1999 Sb., změna: 27/2000 Sb., 105/2000 Sb., změna: 29/2000 Sb., změna: 370/2000 Sb. (část), změna: 30/2000 Sb., 367/2000 Sb., 370/2000 Sb., změna: 370/2000 Sb. (část), změna: 120/2001 Sb., změna: 239/2001 Sb., změna: 501/2001 Sb., změna: 353/2001 Sb., změna: 15/2002 Sb., změna: 126/2002 Sb., změna: 308/2002 Sb., změna: 125/2002 Sb., 151/2002 Sb., 312/2002 Sb., změna: 476/2002 Sb., 87/2003 Sb., 88/2003 Sb., změna: 437/2003 Sb., změna: 370/2000 Sb. (část), 308/2002 Sb. (část), 85/2004 Sb., 257/2004 Sb., změna: 360/2004 Sb., změna: 484/2004 Sb., změna: 360/2004 Sb. (část), změna: 499/2004 Sb., 554/2004 Sb., změna: 179/2005 Sb., změna: 216/2005 Sb. (část), změna: 216/2005 Sb., změna: 377/2005 Sb., změna: 413/2005 Sb., změna: 56/2006 Sb., změna: 79/2006 Sb., 81/2006 Sb., změna: 57/2006 Sb., změna: 308/2006 Sb., změna: 269/2007 Sb., 296/2007 Sb., změna: 36/2008 Sb., změna: 344/2007 Sb., změna: 104/2008 Sb., změna: 126/2008 Sb., 130/2008 Sb., 230/2008 Sb., změna: 215/2009 Sb., 217/2009 Sb., změna: 230/2009 Sb., změna: 285/2009 Sb., změna: 420/2009 Sb., změna: 227/2009 Sb., 152/2010 Sb.

Příloha č. 1 – Návrh Projektu velké infrastruktury schválený vládou

Projekt LINDAT/CLARIN Vybudování a provoz českého uzlu pan-evropské infrastruktury pro výzkum

Uchazeč

Uchazečem o podporu je Univerzita Karlova v Praze; s prostředky bude hospodařit Ústav formální a aplikované lingvistiky MFF UK. Univerzita Karlova v Praze zajistí dvoustrannými smlouvami převod prostředků nezbytných pro provoz centrálního uzlu Clarin a pro zajištění národních úkolů na pracoviště, která se na projektu budou podílet podle schváleného rozpočtu. Univerzita Karlova v Praze rovněž zajistí (nebude-li do té doby rozhodnuto jinak) převod plánovaných prostředků do budoucí asociace Clarin-ERIC.

Osoba odpovědná za organizační, technické a personální zajištění projektu (řešitel-koordinátor): Prof. RNDr. Jan Hajič, Dr. (hajic@ufal.mff.cuni.cz), ředitel Ústavu formální a aplikované lingvistiky MFF UK, Malostranské nám. 25, 11800 Praha 1 (tel. +420 607 209 212).

Ústav formální a aplikované lingvistiky MFF UK (<http://ufal.mff.cuni.cz>) je předním národním pracovištěm v oblasti počítačového zpracování přirozeného jazyka, které bylo rozvíjeno v rámci programu výzkumných center i výzkumných záměrů a které má výrazné mezinárodní postavení v rámci již existujících celoevropských projektů. Plánovaný projekt výzkumné infrastruktury přímo navazuje na projekt CLARIN (FP7-RI-2122230), jehož je pracoviště uchazeče národním koordinátorem.

• Popis výzkumné infrastruktury

1.1. Koncepce

Projekt LINDAT-Clarin je koncipován jako český „uzel“ mezinárodní sítě Clarin (Common Language Resources and Technology Infrastructure, FP7-RI-2122230, zatím 2008-2010) a je rozšířen i na projekt “T4ME Net” (Technologies for the Multilingual European Information Society, NoE, 2011-2014, FP7-ICT-4-249119) pro volné sdílení jazykových dat a základních technologií mezi institucemi a jednotlivci ve vědě a výzkumu. Projekt Clarin je přitom cílen na humanitní vědy, projekt T4ME je širší s dosahem především na jazykové technologie a jejich aplikace.

Centrum LINDAT-Clarin bude tedy v českém jazykovém prostředí propojovat tyto dvě oblasti a bude zaměřeno na sběr jazykových dat a především jejich anotaci (tj. formální manuální, poloautomatickou a automatickou jazykovou analýzu). Sběr a anotace budou probíhat v takovém rozsahu, kvalitě a technologické přípravě (specifikace, schémata, formáty), která bude přímo využitelná jak v humanitní oblasti (jazykovědný a mezioborový výzkum, kde přirozený jazyk hraje podstatnou roli), tak i pro výzkum a vývoj v oblasti jazykových technologií za použití moderních statistických a hybridních metod.

Náplň práce centra a jeho výsledky se dotýkají řady oborů – v humanitních vědách to jsou jazykověda obecná a pro jazykověda zabývající se konkrétními jazyky, zejména češtinou, dále překladatelství, lexikografie, sociolingvistika, částečně i obory příbuzné (psychologie, sociologie, knihovnictví, neurovědy, cognitive science) s významným přesahem do informatiky (computer science, computational linguistics), matematiky (statistika a pravděpodobnost), a elektrotechniky (zpracování akustického signálu).

Z hlediska národních priorit aplikovaného výzkumu (dle dokumentu Národní politika výzkumu, vývoje a inovací České republiky na léta 2009 – 2015, Hlava VI) navrhované centrum spadá do priority 6 (Informační společnost) a 8 (Priority rozvoje české společnosti).

1.2. Současný stav

Jazykové zdroje a technologie pro jejich zpracování v jednotlivých evropských zemích (stejně jako v USA a Asii) již existují, ale současné centralizované distribuční agentury (Linguistic Data Consortium v USA a European Language Resources Association v Evropě) nevyhovují současným požadavkům na jednoduchý, nebyrokratický a zejména volný přístup k jazykovým datům pro většinu vědeckých, výzkumných a vývojových komunit.

Dochází tak k fragmentované, nekoordinované distribuci dat se všemi negativními důsledky (nekompatibilita formátů a tím obtížná použitelnost software pro jejich zpracování, velké množství různých licenčních podmínek, často i nemožnost přistupovat přímo k datům samotným a nutnost používat mnoho různých vyhledávacích systémů apod.). V ČR se data sbírají a anotují především na čtyřech pracovištích, která by se měla stát podílet na činnosti centra LINDAT-Clarin: UK v Praze, MU v Brně a ZČU v Plzni spolu s ÚJČ AV ČR v Praze. Tato pracoviště v současné době tvoří Centrum počítačnické lingvistiky (CKL – projekt MŠMT LC536), které je ovšem vědeckým pracovištěm, jazykové zdroje tvoří spíše okrajově a po roce 2010 už nebude existovat.

Projekt 7. RP (ESFRI) Clarin je zaměřen na odstranění těchto nedostatků v evropském rámci a ve prospěch humanitních a sociálních věd, zejména výzkumu ve všech oblastech lingvistiky. Projekt T4ME Net (zahajuje v březnu 2010) je zaměřen obdobně, ale jeho primárním cílem je sloužit vědecké komunitě mezioborově (lingvistika, informatika, statistika), a to zejména v jazykových technologiích, tzv. počítačové lingvistiky (computational linguistics), která už má následné dopady i v aplikační oblasti.

1.3. Věcná struktura projektu

V souladu s koncepcí projektu a provázaností s projekty Clarin a T4ME bude projekt strukturován takto:

1. Hlavní část: uzel distribuované evropské sítě Clarin
2. Národní část: české a multilingvální jazykové zdroje: sběr a tvorba jazykových korpusů a databází
3. Koordinační část: koordinace v oblasti právní, technologické, vzdělávací a vnějších vztahů

Na základě tohoto věcného členění bude vytvořena i organizační struktura projektu (viz kap. □) a členěn rozpočet projektu (kap. □).

1.3.1. Uzel distribuované evropské sítě Clarin

Jak už vyplývá z povahy projektu a jeho napojení především na ESFRI a projekt Clarin, tato část projektu je zásadní pro úschovu, sdílení a poskytování dat a služeb, což je jádrem projektu Clarin. Centrum LINDAT-Clarin bude ve své funkci distribuovaného uzlu zajišťovat technologické zázemí pro uzel typu „A“ (podle definice projektu Clarin v přípravné fázi), tj. nejvyšší model takového distribučního uzlu. Uzel typu „A“ zajišťuje úschovu jazykových dat (včetně přijímání nových dat a jejich začlenění do systému včetně přidělení unikátních persistentních identifikátorů), autorizovaný přístup s využitím celoevropské federace identit, přidělování identit pro uživatele systému, a webové služby lokální (plně) a distribuované (zprostředkovaně) pro zpracování jazykových dat, přístup k nim a jejich

poskytnutí dalším uzlům Clarin. Tato část projektu bude zajištěna navrhovatelem prostřednictvím MFF UK v Praze.

Rozhodování o směřování této části (na rozdíl od části národní, viz 1.3.2) bude probíhat ve struktuře projektu Clarin. Na tomto rozhodování se bude česká strana podílet podle pravidel projektu Clarin, a to rovněž prostřednictvím MFF UK, která byla v přípravné fázi pověřena koordinováním národních aktivit Clarin v ČR.

1.3.2. České jazykové zdroje: sběr a tvorba jazykových korpusů a databází

Zatímco jazyková data, která má distribuovaný uzel sítě Clarin uschovávat a sdílet, existují (a stále se tvoří) na řadě pracovišť v zahraničí, v českých podmínkách je nutno pro sběr a tvorbu jazykových korpusů stále aktivně podporovat. Z pochopitelných důvodů v zahraničí vznikají jazyková data v dané oblasti lokální (v Německu jde o němčinu, a Holandsku o holandštinu, v Belgii francouzštinu a vlámsčinu atd.), a česká jazyková data všech druhů je tedy nutno zajistit v ČR.

Zatímco sběrem dat se zabývají i jiná pracoviště v ČR a činnost LINDAT-Clarin zde bude pouze doplňovat chybějící oblasti, především mluvená data, paralelní data (s jinými jazyky) ve velkém rozsahu, a data kombinovaná (definice viz níže), při tvorbě anotovaných jazykových korpusů, které jsou základem pro další výzkum a vývoj jak v humanitních, tak v technologických a aplikačních oblastech klíčové, bude centrum LINDAT-Clarin v ČR jedinečné.

Sběrem jazykových dat se rozumí:

- získávání psaných (případně mluvených) jazykových dat z veřejně dostupných zdrojů (internet, otevřené zdroje); tato data mohou být i v kombinaci s dalšími modalitami, např. videem, znakovou řečí zaznamenanou symbolicky apod.
- získávání psaných nebo mluvených jazykových dat (s případnou vizuální složkou) od nakladatelů a majitelů těchto dat na základě smluv
- nahrávky řečových dat v předem určeném prostředí (laboratorní, aplikačně přizpůsobené, v terénu)
- elektronická data ze senzorů při expozici subjektů přirozenému jazyku v libovolné podobě (video, haptika, elektrické nebo magnetické snímání reakcí apod.)

Tvorbou jazykových korpusů a databází se rozumí takové zpracování sebraných dat, které je obohaceno pro použití v humanitních i technických vědách v těchto směrech:

- tzv. „čištění“, unifikace a normalizace dat do standardních datových formátů, které lze využít pro další zpracování nebo distribuci
- párování (alignment) pro paralelní jazyková data (psaná, mluvená, vizuální nebo jiná), s cílem zajistit explicitní vztahy mezi jazyky nebo různými modalitami dat navzájem, a to pomocí manuálních nebo automatických metod, jejichž výsledkem jsou synchronizační značky obsahového nebo časového rázu
- extrakce jazykových databází z jazykových dat; databázemi se rozumí především slovníky (fonetické, morfologické, syntaktické a sémantické, příp. s vazbou na ontologie)
- anotace jazykových korpusů; anotací se rozumí manuální nebo (polo)automatická analýza jazykových dat (s výstupem v elektronické formě), která je nutná pro další využití dat v humanitní i technologické oblasti, a to

podle světových standardů a nových postupů s ohledem na vlastnosti českého jazyka.

Cílem této části projektu je připravit jazyková data (jazykové korpusy a databáze, podle situace i s anotací) k publikaci v nejkratším možném termínu prostřednictvím uzlu Clarin (viz 1.3.1) a dalších distribučních kanálů tak, aby byla k dispozici co nejširší veřejnosti pomocí jednoduché licenční a autentizační politiky bez technologických nebo právních překážek, a to od institucionální úrovně až na úroveň jednotlivého výzkumníka nebo studenta.

Na této části projektu se budou podílet všechna čtyři česká partnerská pracoviště podle svého zaměření a specializace. V jednotlivých případech se očekává společný postup při tvorbě zvláště pracovně nebo technologicky náročných jazykových korpusů a databází.

1.3.3. Koordinace v oblasti právní, technologické, vzdělávací a vnějších vztahů

Sběr, tvorba, anotace a distribuce jazykových dat vyžaduje podle dosavadních zkušeností řadu navazujících činností, které je vhodné soustředit mimo vlastní jazykovědnou nebo technologickou práci.

Pro sběr dat na jedné a pro distribuci dat na druhé straně je třeba důsledně dbát na vhodnou aktivní i pasivní licenční politiku, s cílem co nejvíce zjednodušit přístup k vytvořeným jazykovým datům koncovým uživatelům a poskytnout jim rovněž nejvyšší reálně možnou právní jistotu, že data, která si pomocí distribuční sítě Clarin pořídí, nemají z hlediska vědeckého užití žádná omezení. Situace je nyní v EU jednodušší než před několika lety díky změnám v zákonech týkajících se autorských práv (v ČR zákon 121/200 Sb. v aktuálním znění), nicméně stále není pro vědecké využití jazykových dat ideální a je třeba ji dostatečně zajistit.

Technologickým cílem je sjednotit softwarové nástroje pro sběr, čištění, párování i (manuální) anotaci dat tak, aby byla zajištěna interoperabilita minimálně na úrovni datových formátů (XML a jeho varianty a restriktce pro jazykovou oblast). Stejněho cíle je nutno dosáhnout pro uchovávání a distribuci dat a webové služby v rámci sítě Clarin (na základě rozhodnutí koordinačního centra Clarin).

Nedílnou součástí centra LINDAT-Clarin je i výchova odborníků v oblasti jazykové, jazykově-technologické a organizační ve všech oblastech práce centra, a to na úrovni navazujících magisterských programů, na úrovni doktorandské a při dalším vzdělávání odborníků z vysokých škol, Akademie věd a aplikované praxe. Cílem je vzdělaná skupina pracovníků, kteří budou dále efektivně pracovat v centru LINDAT-Clarin, ve výzkumu humanitním i technologickém a praxi s jazykovými daty a technologiemi.

Centrum LINDAT-Clarin bude rovněž široce propagovat využití jazykových technologií a ukazovat jejich možnosti pro rozvoj znalostní společnosti pořádáním seminářů určených pro jednotlivé cílové skupiny (od vědeckých až po manažerské), podílem na organizaci celoevropských akcí ve výzkumně-jazykové oblasti (zejména ve spolupráci z projektem T4ME Net) a účasti pracovníků centra na obdobných akcích.

Tyto aktivity budou koordinovány jak s projektem Clarin (resp. jeho centrálou), a projektem T4ME Net na „evropské“ straně a rovněž v rámci národní účasti (ZČU, MU, ÚJČ AV ČR) prostřednictvím zástupců MFF UK v těchto projektech a vnitřní organizační struktury projektu LINDAT-Clarin (viz kap. □).

- ***Přínos pro výzkum a vývoj v Evropě a v České republice***

Centrum bude pokračovat v mezinárodní spolupráci při tvorbě a distribuci dat. V rámci projektu Clarin již nyní spolupracuje 171 institucí z 27 členských a 5 asociovaných zemí EU (na úrovni univerzit, akademických ústavů atd.). V projektu T4ME bude při jeho zahájení 17 partnerů, kteří jsou financováni EU, a předpokládá se vytvoření konsorcia uživatelů s cca 200-300 institucionálními členy z EU a asociovaných zemí. V obou projektech je partnerem z ČR Univerzita Karlova (prostřednictvím ÚFAL MFF UK).

Předpokládání partneri projektu byli a jsou rovněž zapojeni do dalších projektů v ČR i EU, ve kterých (dosud nekoordinovaně) vznikají jazyková data a různými způsoby se distribuují, například EuromatrixPlus a Faust (ÚFAL MFF UK - jazyková data pro automatický překlad), Companions (ÚFAL MFF UK, KKY FAV ZČU – řečová anotovaná data pro vývoj dialogových systémů), KYOTO (MU Brno, lexikální data). ÚFAL MFF UK je dále zapojen do existujícího menšího projektu NoE FlareNet, kde vznikají standardy a specifikace společného základu pro jazykové zdroje.

ÚFAL MFF UK spolupracuje dále s USA a některými asijskými pracovišti na projektech anotace jazykových dat: projekt NSF PIRE (anotace pro pokročilé porozumění mluvenému jazyku, strojový překlad: Johns Hopkins University, MD, USA, Brown University, RI, USA), projekty anotace syntaxe a valence (University of Colorado, CO, Brandeis University, MA, IIT Hyderabad, Indie), projekt anotace jazykových dat pro výzkum diskursu (University of Pennsylvania, PA, USA). ÚFAL MFF UK rovněž spravuje část dat a přístupový bod k 116 tisícům hodin nahrávek vzpomínek přeživších obětí holocaustu („Centrum Malach“, smlouva s University of Southern California, CA, USA), které jsou významným a velikostí dosud nepřekonaným zdrojem mluvených dat v mnoha jazycích.

ÚFAL MFF UK má spolu s Katedrou kybernetiky FAV ZČU v Plzni rovněž dlouholeté zkušenosti s vydáváním a „klasickou“ distribucí jazykových dat přes střediska LDC (USA) a ELRA/ELDA (Francie/EU).

Význam navrhované infrastruktury lze shrnout do těchto bodů:

1. vytvoří národní referenční zdroj jazykových dat pro veřejně dostupné, snadné a právně bezproblémové vědecké, pedagogické, výzkumné i aplikačně-vývojové použití
2. umožní široký přístup odborné komunitě i uživatelům k expertíze, která je ověřena
3. umožní široký přístup odborné komunitě i uživatelům k již vyvinutým počítačovým nástrojům a technologiím i službám
4. umožní široký přístup odborné komunitě i uživatelům nejen k jednojazyčným (českým) zdrojům, ale ke zdrojům multilingválním a k odpovídajícím technologiím
5. bude tvořit významnou „přidanou mezinárodní hodnotu“ k národním, v našem případě českým, iniciativám, umožní propojení center v celoevropském měřítku
6. poskytne významný potenciál k inovacím
7. posílí zájem o národní jazyk jako součást národní kultury a národního dědictví
8. v neposlední řadě významnou měrou přispěje k modernizaci pedagogického procesu (výuka jazyků, jazykové technologie, velká data a jejich zpracování)

Klíčová je přitom otázka, kdo a jak bude vytvořené centrum LINDAT-Clarin využívat; na základě dosavadních zkušeností s využitím jazykových dat předpokládáme, že uživatelé se budou rekrutovat z těchto kategorií:

1. v oblasti vědecké: všechna akademická pracoviště (na univerzitách i v AV ČR) zabývající se vědeckým zpracováním, výukou a počítačovým zpracováním);

- předpokládá se využití i ze zahraničí (pracoviště stejného nebo podobného typu) v rámci projektů Clarin a T4ME Net i mimo ně, studenti zapojení do vědecké práce (magisterské a doktorské studium, zahraniční studenti v ČR);
2. v oblasti aplikačního vývoje a inovací: všechna pracoviště a průmyslové organizace zabývající se informačními systémy (knihovnictví, dokumentační střediska, atd.), překladatelskými službami, vyhledáváním dokumentů a informací, terminologickými databázemi, podpůrnými nástroji pro tvorbu textů (korektory), dokumentací v oblasti historie, průzkumy sociologickými, psychologickými a aplikacemi podobného charakteru, v nichž se pracuje s texty a textovými záznamy; i zde je vysoká pravděpodobnost využití ze zahraničí (lokalizační projekty zahraničních firem, firem vyvíjejících nástroje pro podporu automatického překladu apod. – viz seznam výše);
 3. v oblasti výuky: využití ve školství všech stupňů v rámci jazykové výuky, výuky IVT a případně dalších předmětů.

• *Cíle projektu*

3.1. Úvod

Projekt EU/ESFRI Clarin, stejně jako budoucí T4ME Net a ve stejném duchu i navrhované centrum LINDAT-Clarin má překážky volného přístupu k jazykovým datům postupně odstranit a umožnit distribuované, nicméně jednotné poskytování jazykových dat a souvisejících technologií. Ambice Clarinu a do značné míry také T4ME (s výjimkou evaluačních dat pro konkurenční testování nástrojů pro zpracování přirozeného jazyka) jsou zejména ve sjednocení technologií a distribuce – tvorba je ponechána na kompetenci (a finančním zajištění) v jednotlivých státech, neboť se jedná často o národní jazyky. Sběr a anotace (tj. tvorba dat) je tedy nedílnou součástí navrhovaného centra LINDAT-Clarin.

Anotace dat v dostatečném rozsahu je nezbytná k tomu, aby výsledky tohoto výzkumu mohly být aplikovány v praxi (korektory textu, automatický překlad, extrakce informací z textu, porozumění textu, dialogové systémy apod.), neboť vývoj softwarových nástrojů je dnes založen na statistických metodách a ty vyžadují velké množství zejména jazykově interpretovaných (anotovaných) dat. Tato data jsou potřebná i pro to, aby čeština byla zařazena do programu lokalizace textových produktů velkých firem, které tyto metody vesměs využívají.

Existence anotovaných dat v daném jazyce rozhoduje často o zařazení do velkých výzkumných projektů EU (v minulosti UK získala několik takových projektů mj. díky vlastnictví anotovaných korpusů – např. projekty Euromatrix, EuromatrixPlus, Companions - se ZČU, Faust a samozřejmě i projekty Clarin a T4ME Net).

Anotace dat je časově i kapacitně velmi náročná - jedná se především o manuální práci vysoce odborně školených jazykových odborníků, a to studentů doktorského (výjimečně i navazujícího magisterského) studia jazykovědných oborů, zejména mezioborových v kombinaci s informatikou, s vysokým podílem technického zajištění odborníky z oblasti informatiky. Sběr dat je rovněž náročná práce, která vyžaduje odborníky několika profesí, od informatiky přes organizační zajištění, právní zabezpečení, až po výuku a školení uživatelů.

Jednotná distribuce pak zajistí širší publicitu a využití dat a tedy i vložených prostředků.

Nezanedbatelným cílem projektu je vychovat další vědeckou generaci, která bude umět s jazykovými daty pracovat, správně je tvořit a používat v národním i mezinárodním kontextu, a spolupracovat v rámci EU i mimo ni na budoucích projektech využívajících jazykové technologie.

3.2. Dílčí cíle projektu

Dílčí cíle projektu vycházejí z hlavních dvou hlavních cílů projektu popsaných v předchozím odstavci – pořídit jazyková data pro češtinu (a data paralelní) a zpřístupnit je. Mezi kontinuálně pojaté dílčí cíle projektu patří a zajištění výchovy mladé vědecké generace v oblasti jazykových dat a jazykových technologií v humanitní i technické oblasti.

3.2.1. Výstavba distribučního uzlu Clarin

Výstavba distribučního uzlu typu „A“ v metodologii Clarin je základním předpokladem fungování centra LINDAT-Clarin v rámci evropského výzkumného prostoru a v rámci projektu Clarin (a T4ME Net). Tento uzel bude budován postupně v letech 2010 (příprava, dokončení specifikací) a 2011-2013 (konstrukční fáze, soustředění technologií, postupné získávání dat a zahájení zkušebního provozu). V plném provozu se očekává v roce 2014.

Výstavba bude mít investiční část (ve smyslu pořízení výpočetní a síťové techniky na UK v Praze a na ostatních pracovištích) a část softwarovou při zavedení celosvětově jednotné identifikace dat, jednotné a alespoň celoevropsky platné autentikace uživatelů pomocí federace identit (v ČR: EduID), zavedení dohodnutého (v rámci Clarin) repozitáře pro úschovu vlastních i „cizích“ dat, zavedení jednotného popisu metadat, zavedení jednotného API pro webové služby v oblasti zpracování jazykových dat a zajištění právního rámce (vzorové smlouvy pro otevřené licence pro poskytovatele dat, licence pro uživatele).

3.2.2. Národní infrastruktura pro sběr a tvorbu jazykových dat

Jednotlivá pracoviště navrhovaného centra LINDAT-Clarin spolu již v oblasti sběru a anotace dat spolupracují na základě jednotlivých případů (projekt 6. RP „Companions“ a další). V rámci navrhovaného centra LINDAT-Clarin budou vzájemně předány nástroje na budování a anotaci jazykových dat, aby se dosáhlo co nejvyšší efektivity a co nejširší dostupnosti i mimo partnerské instituce.

Nástroje na sběr dat z otevřených zdrojů, na základní zpracování („čištění“ dat, tokenizace, segmentace) a na základní jazykové automatické předzpracování budou vzájemně volně dostupné a v případech, kdy to bude výhodné, budou sjednoceny (pokud se nyní používají podobné, nikoli však identické a všem vyhovující typy nástrojů).

Pro distribuci dat a pro ty služby v oblasti zpracování jazykových dat, kde i v rámci ČR to bude efektivnější, bude využíván uzel Clarin (viz 3.2.1).

3.2.3. Sběr dat

Sběr dat bude prováděn v závislosti na typu dat. Získávání textových českých a textových paralelních dat z otevřených zdrojů (tj. především internetu) je otázkou vhodné technologie, která bude mezi jednotlivými českými pracovišti sdílena.

Řečová data je nutno získat od subjektů-lidí, kteří budou předčítat vhodně vybrané pasáže pro daný konkrétní úkol. Podobně bude probíhat i záznam dialogů a senzorických reakcí lidí na jazykové podněty. Ve všech případech bude vyžádán souhlas subjektu na základě zákona o ochraně osobních dat; povaha získávání dat však nevyžaduje speciální opatření ohledně etických norem.

Součástí sběru dat je i jejich prvotní zpracování a ukládání s jednotnými metadaty a ve formátu vhodném pro další zpracování, zejména manuální transkripci a anotaci (viz 3.2.4).

3.2.4. Anotace dat

Lingvistická anotace jazykových dat, která je klíčová pro zhodnocení dat zejména pro jejich budoucí aplikační využití, je časově i finančně nejnákladnější a nejnáročnější operací při jejich přípravě. Základní metoda anotace se skládá ze čtyř fází:

- výběr dat (z dat sebraných a připravených)
- příprava anotačního nástroje pro přímou anotaci v elektronické formě
- vlastní anotace (včetně organizačního a administrativního zajištění)
- kontrola kvality anotace, příprava dat k uložení a veřejnému zpřístupnění

Rozdělení těchto čtyř oblastí z hlediska objemu práce je takové, že vlastní anotace je zastoupena cca polovinou. Kontrola kvality rovněž záleží na komplexnosti anotace – jedná se většinou o porovnání dvojité nebo trojité anotace a opravu neshody mezi anotátory v datech; lze aplikovat i různé automatické nebo poloautomatické metody kontroly formátu i obsahu anotace.

3.2.5. Příprava dat k publikaci, vlastní zpřístupnění a distribuce dat

Cílem projektu LINDAT-Clarin i samotných evropských projektů Clarin a T4ME Net je co nejvíce zpřístupnit (a to co nejjednodušeji) jazyková data a metody jejich zpracování výzkumné komunitě v České republice i Evropě. Data je tedy třeba publikovat – v novém konceptu Clarin to znamená, že data budou po jejich vytvoření a anotaci co nejdříve zpřístupněna v síti uzlů Clarin.

Pro tento účel je třeba vytvořená data popsat z hlediska budoucích uživatelů (anotační příručky, jejich překlad do relevantních jazyků), Data je pak nutno odpovídajícím způsobem uložit v repozitářích sítě Clarin.

3.2.6. Začlenění softwarových nástrojů do poskytovaných webových služeb

Pro přístup k jazykovým datům je obvykle třeba vytvořit software včetně vyhledávacích a statistických modulů (pokud se data nezpřístupňují prostým poskytnutím v kopii) a realizovat jej jako webovou službu v síti Clarin. Vzhledem k různorodosti jednotlivých datových souborů a databází je často nutno vytvářet specializovaný software, který umožní uživatelům např. i z humanitních oblastí plně a efektivně daná data využívat v jejich práci.

3.2.7. Výchova vědeckých pracovníků v oblasti jazykových dat

Nedílnou součástí centra LINDAT-Clarin je výchova vědeckých pracovníků pro práci s jazykovými daty, a to jak z hlediska lingvistického (teoretického i aplikovaného), tak z hlediska informatického. Všechna čtyři pracoviště disponují odborníky v těchto oblastech; budou využívány (v případě zájmu studentů, zapojených do prací v centru) i akreditované magisterské a doktorandské programy, neboť jazyková data i způsob jejich sběru a anotace jsou vhodným materiálem pro diplomové a doktorské práce.

V rámci možností bude centrum LINDAT-Clarin poskytovat toto vzdělání pro studenty EU i mimo EU podle platných předpisů a regulí, s využitím vzdělávacích programů EU i přímo.

3.2.8. Organizace vědeckých seminářů a konferencí v oblasti jazykových dat

Pro odborníky (ze všech relevantních oblastí) bude centrum LINDAT-Clarin pořádat semináře a případně se podílet i na organizaci konferencí v oboru řeči a zpracování přirozeného jazyka, v čemž mají spolupracující organizace již dlouholetou tradici (konference Text, Speech and Dialogue v ČR, konference a workshopy Association for Computational Linguistics).

3.2.9. Organizace seminářů a školení pro uživatele služeb

Uživatelé služeb uzlu sítě Clarin budou školeni v metodách používání jazykových dat zpřístupněných pomocí této sítě, a to jak v ČR, tak ve spolupráci s ústředím Clarin i v zahraničí.

3.3. Časový harmonogram pro roky 2010-2015

Obecně lze harmonogram rozdělit na tři části:

- 2010 příprava centra
- 2011-2013 konstrukční fáze, ve spolupráci s Clarin (Clarin-ERIC, bude-li založeno)
- 2014-2015(-2020 a dále) provoz (operační fáze) / uzel Clarin-ERIC.

V současné době jsou jazykové zdroje a příslušné programové nástroje uchovávány na pracovištích, kde vznikají, a také jimi zpřístupněny nebo distribuovány (web, fyzická média). Výjimečně jsou distribuovány dvěma existujícími mezinárodními středisky (LDC - USA, ELRA - Francie). Tvorba zdrojů se provádí zejména na pracovištích UK a ZČU, ÚFAL MFF UK má vybudované menší úložiště (několik TB diskového prostoru) a výpočetní kapacitu cca 200 CPU jednotek na základní zpracování jazykových dat.

1. Vybudování úložného a distribučního centra (Centrum „A“ v intencích projektu Clarin a podobné centrum v rámci projektu T4ME) se předpokládá ve dvou etapách: (i) vybudování technické infrastruktury (úložiště, síť, připojení) do r. 2013 (konstrukční fáze), a poté (ii) její provoz a případné kapacitní rozšiřování (operační fáze, kontinuálně). Zajištění toho distribučního centra je plánováno na ÚFAL MFF UK.
2. Sběr dat, archivace, zpřístupnění: tato aktivita v ČR probíhá na několika pracovištích průběžně (UK, MU, ZČU). Jedná se tedy o převedení těchto aktivit na společnou platformu, začlenění relevantních datových kolekcí pod distribuční schéma centra ad (1) do konce r. 2013. Operační fáze kontinuálně – částečně již paralelně s konstrukční, plně od r. 2014.
3. Specifikace a tvorba anotovaných dat, softwarových nástrojů pro zpracování dat: převedení těchto činností ze současného Centra počítačnické lingvistiky a částečně i několika „Výzkumných záměrů“ po jejich skončení, organizační zajištění, propojení s distribucí ad (1) do konce r. 2013. Operační fáze (včetně poskytnutí nástrojů a kapacit pro přípravu vlastních dat) od r. 2014 kontinuálně.

- **Dosavadní financování infrastruktury a začlenění České republiky**

Clarín: rozpočet pro ČR (spravováno UK): 90 tis. EUR, plus spolufinancování MŠMT na účast v rámcových programech - v rozsahu 33 % celkových nákladů na českou účast v projektu (26 tis. EUR), 2008-2010.

Flarenet: 6. RP, celková rozpočet pro UK: 9 tis. EUR, bez spolufinancování, 2007-2010.

Části projektů EU určené na tvorbu jazykových dat: celkem cca 660 tis. EUR, plus část příspěvku ČR na projekty 7. RP: cca 100 tis. EUR, různé projekty v letech 2006-2013 (plán).

Příprava dat, proj. v USA: po přepočtu cca 150 tis. EUR (zdroje: NSF, DARPA, 1998-2009)

Národní projekty (250 mil., LN, LC, MSM, GAČR, GAUK, jiné MŠMT) v jejichž rámci se sbírají nebo anotují jazyková data): po přepočtu cca 1800 tis. EUR celkem v letech 1996-2010 celkem (hrubý odhad) na vlastní sběr, anotaci a distribuci dat.

Strukturální fondy nelze pro sběr a pořizování dat použít.

Celkem dosud (1996-2009): 2835 tis. EUR, z toho ze zdrojů v ČR: 1926 tis. EUR
(v Kč: 70 mil. Kč, z toho ze zdrojů v ČR: 48 mil Kč)

Tyto zdroje byly a jsou určeny vesměs na pořizování a anotaci dat, a jsou vedené tedy nikoli jako *investiční*, ale jako *provozní* náklady (investiční náklady jsou řádově menší, pouze na datová úložiště a jejich obnovu - odpisy). Z hlediska celkového charakteru jazykové infrastruktury jsou však náklady na sběr a zejména anotaci a zpracování jazykových dat vždy dominující.

- **Trvání a jednotlivé fáze projektu**

Projekt je plánován na 10 let (s možností pokračování), tj. na roky 2010-2020. V roce 2010 budou v době od července 2010 (předpokládané zahájení poskytování podpory) probíhat přípravné a organizační práce (souběžně s doběhem přípravné fáze projektu Clarín v 7. RP), plné začlenění do sítě Clarín se očekává od roku 2011, stejně jako počátek prací na národním rozšíření a koordinaci dalších přidružených aktivit.

Fáze projektu jsou tři: přípravná (2010), konstrukční (2011-2013) a provozní (2014-2020 a dále).

- **Indikativní rámcový rozpočet projektu**

Odhad nákladů v tis. Kč:

Přípravná fáze v 2010 (iniciální investice, právní a organizační příprava)	15
000	
Konstrukční fáze, 2011-2013:	60 000
Operační fáze, 2014-2015:	80 000
Odhad na 2016-2020 (operační fáze, rozšiřování, nové technologie):	210 000

Pozn.: podle údajů ESFRI budou konstrukční fáze financovány z národních prostředků, přímý příspěvek v malé výši z EU lze očekávat až v operační fázi.

Z toho požadavek na financování z ČR:

Přípravná fáze (iniciální investice, právní a organizační příprava, mechanismus dofinancování projektů rámcových programů)

Konstrukční fáze, 2011-2013: 60 000

Operační fáze, 2014-2015: 60 000

Odhad na 2016-2020 (operační fáze, rozšiřování, nové technologie): 150 000

Podrobné zdůvodnění a rozpis na jednotlivé kategorie nákladů rozpočtu je uveden v části □ tohoto návrhu.

- **Indikativní detailní rozpočet na konstrukční fázi projektu (2011-2013) a na první dva roky provozní fáze (2014 – 2015)**

Rozpis podle struktury plánovaného centra LINDAT (2010-2015):

V tis. Kč		Rok						Celkem
		2010	2011	2012	2013	2014	2015	2011-15
1 Distribuce dat	Celkem	6777	4405	3447	3519	3520	3521	25 189
	<i>Inv.</i>	4800	1000	0	0	0	0	
	<i>Osobní vč. SZP</i>	642	2070	2085	2135	2136	2137	
	<i>Ost. NIV</i>	744	305	320	320	320	320	
	<i>Nepřímé</i>	591	1030	1042	1064	1064	1064	
2 Sběr, anotace	Celkem	6993	11724	12560	12575	12575	12575	69 002
	<i>Inv.</i>	5050	900	0	0	0	0	
	<i>Osobní vč. SZP</i>	340	6650	7800	7800	7 800	7 800	
	<i>Ost. NIV</i>	1030	950	1020	1030	1030	1030	
	<i>Nepřímé</i>	573	3224	3740	3745	3745	3745	
3 Koordinace	Celkem	1123	3802	3802	3802	3802	3802	20 133
	<i>Inv.</i>	0	0	0	0	0	0	
	<i>Osobní vč. SZP</i>	680	1745	1745	1745	1745	1745	
	<i>Ost. NIV</i>	100	180	180	180	180	180	
	<i>Nepřímé</i>	343	837	837	837	837	837	
	<i>ERIC přísp.</i>	0	1040	1040	1040	1040	1040	
Celkem		14893	19931	19809	19896	19897	19898	114324

Plán nákladů na jazykovou infrastrukturu – centrum LINDAT, v tis. Kč

Odůvodnění rozpočtu

Komentář k jednotlivým položkám nákladů (v závorce je odkaz na sekci návrhu popisující náplň práce a financované aktivity – viz kap. 1.3 návrhu):

Obecný a společný komentář

„Osobní náklady“ se vždy uvádějí včetně povinných odvodů na SZP a na sociální fond UK a zúčastněných institucí (nyní 36%).

Nepřímé náklady (režie) vychází z interních směrnic a propočtů jednotlivých institucí, a odpovídá v současnosti uplatňované režii vůči EU, nebo průměru interního předpisu a nepřímých nákladů uplatňovaných v rámci EU. Z částky odváděné pro ERIC-Clarín se nepřímé náklady nevyčísľují. Jednotlivé koeficienty jsou 0.44 (UK), 0.4 (ZČU), 0.4 (MU) a 0.4 (ÚJČ).

V oblasti osobních nákladů (viz níže pro jednotlivé oblasti práce v LINDAT-Clarín) disponují všechna čtyři pracoviště již dnes pracovníky s dostatečnou kvalifikací k pokrytí plánovaných 12-14 stálých pracovních míst jak z hlediska odborného, tak manažerského, a to i v dlouhodobém výhledu.

Distribuce (dat a služeb) (1.3.1)

Investiční náklady: v této části projektu (technické zajištění uzlu Clarín, které bude provedeno na UK v Praze) je nutno vybudovat i záložní síťovou strukturu pro zajištění nepřetržitého provozu a významně posílit výpočetní a úložnou kapacitu pro poskytování webových služeb spojených s provozem uzlu Clarín typu „A“. Hlavní investice bude provedena již v roce 2010 a dokončena v roce 2011; v dalších dvou letech se investice neplánují. K dalšímu rozšíření dojde podle skutečného využití Centra v roce 2014 nebo později.

Osobní náklady se plánují rovněž převážně na UK v Praze, která bude zajišťovat provoz uzlu sítě Clarín. Rozpočet vychází z potřeby přípravy a zajištění investic v roce 2010 (1/2 pracovní úvazek tech. pracovníka) a 2 pracovních míst (technické zajištění (1), software – vývoj, údržba a podpora (2 místa)) v době konstrukce (a zkušebního provozu). V osobních nákladech je 1/2 pracovního úvazku zahrnut i vědecký koordinátor centra LINDAT (na UK v Praze). Osobní náklady na ostatních třech institucích se v této části týkají zajištění instalace a údržby data a software pro webové služby dodané těmito institucemi jako součást uzlu Centra Clarín.

Další neinvestiční náklady se týkají pořízení drobného majetku a softwarových licencí, které budou jednotně předepsány z ústředí Centra Clarín a které neexistují v prostředí open source, zejména pak repozitářového software a software k autentikaci a zajištění identifikátorů pro datové soubory (placená služba: bude zajištěno buď prostřednictvím ústředí Clarín, nebo samostatně jednotlivými uzly sítě Clarín).

Nepřímé náklady: viz obecná část komentáře výše.

Sběr a anotace dat (1.3.2)

Investiční náklady: na všech pracovištích centra LINDAT-Clarín je třeba posílit výpočetní sílu (v rámci stávajících výpočetních clusterů), zejména kapacitu a zálohování úložišť dat. Posílená výpočetní kapacita bude použita na zpracování dat před, při i po anotaci a na přípravu dat k vydání. Investice budou pořizovány v prvních dvou letech.

Osobní náklady tvoří nejpodstatnější složku při pořizování a nahrávání dat. Obecně lze osobní náklady rozdělit do tří kategorií:

- náklady na stálé zaměstnance (experty v lingvistice a informatice pro lingvistickou, organizační a softwarovou podporu anotací dat)
- náklady na anotační práce, převážně lingvistického charakteru (studenti a v menší míře experti z oboru jazykovědy; většinou formou dohod o pracích konaných mimo pracovní poměr na dobu trvání toho kterého anotačního projektu)