

Pisanje predloga projekta v H2020

Marko Grobelnik

Institut Jozef Stefan



Vsebina

- Izhodišča pri pisanju predloga projekta
- Kako se lotimo pisanja predloga projekta?
- Praksa pri pisanju projektne predloga

Nekaj osnovnih izhodišč pri
pisanju predloga projekta

Kdo so “igralci” v prijavljanju projektov?

- **Mi**, ki prijavljamo
 - ...običajno se trudimo pridobiti projekt in smo za to pripravljeni investirati nekaj energije in sredstev
- **Naši partnerji**, konzorcij s katerimi prijavljamo
 - ...običajno imajo podobne cilje kot mi (ni pa nujno in tega se je dobro zavedati vnaprej)
- **Projektni uradniki** iz EC
 - ...njihova motivacija je, da imajo dobre projekte s katerimi se lahko promovirajo in da nimajo težav
- **Evalvatorji**, ki ocenijo naš predlog projekta
 - ...to so strokovnjaki, ki se trudijo opraviti dobro delo v kratkem času za solidno plačilo

Za koga pišemo projektni predlog?

- V prvi vrsti **za nas!**
 - ...če slučajno dobimo projekt, potem smo mi tisti, ki trpimo naloge in težave, ki smo jih obljubili
- Projektni predlog mora biti sprva izrazito obarvan **za oči evalvatorjev**
 - ...to zahteva nekaj izkušenj
- Če je projektni predlog sprejet, potem ga v drugi fazi skupaj s **projektnimi uradniki** spremenimo v pogodbo
 - ...tu lahko v dokument vnesemo veliko sprememb

Kakšna je perspektiva evalvatorja? (1/3)

- Povprečni evalvator našega projekta je strokovnjak, ki **ne pozna** nujno našega področja **v podrobnosti**
 - ...v predlogu projekta moramo evalvatorja izobraziti in ne smemo pričakovati, da pozna vsebino v podrobnosti
- Evalvator ima vedno omejen čas (od ene do treh ur), da prebere kar smo mu poslali
 - ...predlog mora biti napisati pregledno in prijazno
 - ...ne smemo pričakovati, da bo evalvator posebej iskal nečesa česar ne najde takoj tam, kjer jo pričakuje
 - ...koristno je, če damo predlog pred pošiljanjem v pregled nekomu, ki ima izkušnjo z evalviranjem

Kakšna je perspektiva evalvatorja? (2/3)

- Evalvator ne ocenjuje **nikoli sam** in je zanj rizično, če skuša biti **nekompetentno zloben**
 - ...v predlogu se moramo izogniti čim več razlogom, ki bi jih lahko nekdo uporabil proti nam
 - ...če je predlog predhodno usklajen s projektnim uradnikom iz EC, potem lahko pričakujemo, da bo projektni uradnik, ki sedi na panelu, na naši strani
- Evalvator se skuša obnašati racionalno in skuša vedno **zmanjšati tveganje, da ga bodo drugi spoznali za nekompetentnega v debati**
 - ...naš predlog projekta mu mora dati dovolj argumentov, da se lahko tak evalvator bori za nas

Kakšna je perspektiva evalvatorja? (3/3)

- Evalvatorji so običajno izkušeni in jim le težko uide kakšna od pomembnejših zadev
 - ...ne računati, da evalvatorji morda nečesa ne bodo opazili
 - ...vsaj eden od evalvatorjev bo nekaj opazil in povedal naprej (dobro ali slabo)
- Evalvator ima pred sabo evalvacijski formular, ki mu daje vodilo kako ocenjuje naš predlog
 - ...na naslednjem slajdu si oglejmo kako so predlogi ocenjevani

FP7 Evaluation criteria

- Scale of 0-3 (and 0) with min threshold
- **1. Scientific and/or technological excellence (relevant to the topics addressed by the call)**
 - ...kako dobro idejo imamo?
- **2. Quality and efficiency of the implementation and the management**
 - ...kako dober je plan, da izvedemo našo idejo?
- **3. Potential impact through the development, dissemination and use of project results**
 - ...ali bo naša ideja imela kakšen učinek na okolje?

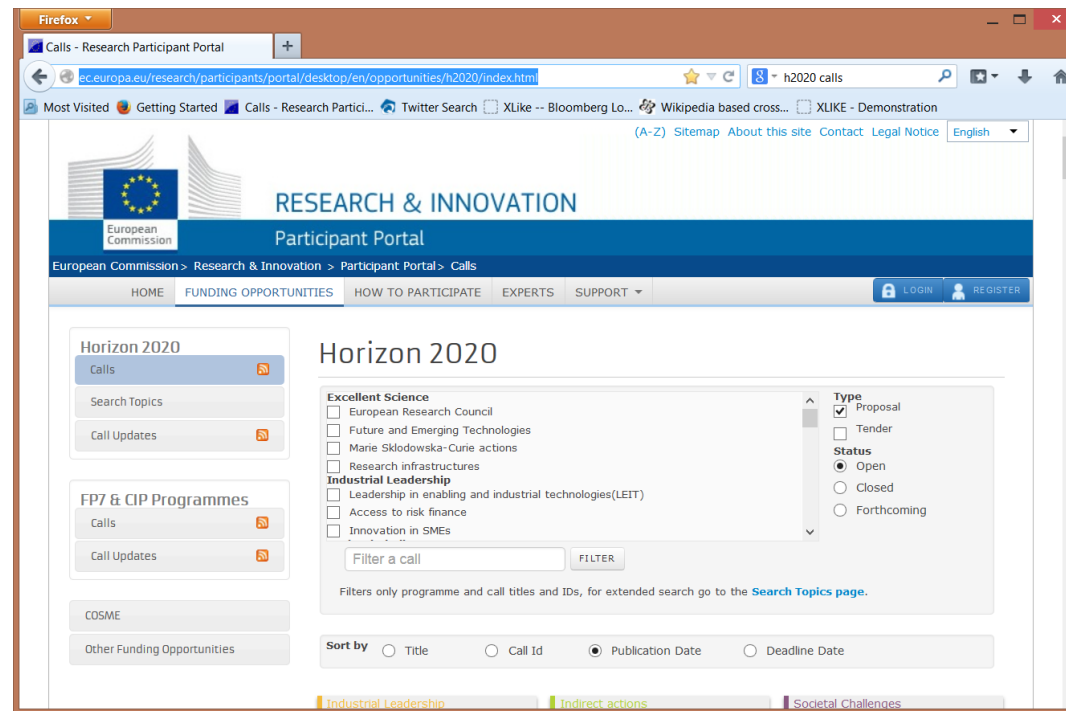
Kako se lotimo pisanja predloga projekta?

Kaj je predlog projekta?

- Predlog projekta je dokument, ki ga pošljemo na Evropsko komisijo in ki opisuje vse vidike projekta kot smo si ga zamislili
- Dokument je dolg od 10 strani do preko 100 strani (odvisno od klica)
- Struktura dokumenta je podana s predlogo, ki jo definira projektni klic
 - običajno je del informacijskega paketa projektnega klica

Kje najdemo informacije o klicih?

- Informacije o klicih so zbrane na strani:
<http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/index.html>
- Stran vsebuje klice po vseh kategorijah in vsebinskih sklopih projektov



Kakšna je informacija o posameznem klicu?

- Vsak klic ima splošno informacijo in informacijski paket
- Vse podrobnosti o klicu (administrativne in vsebinske) vsebuje informacijski paket

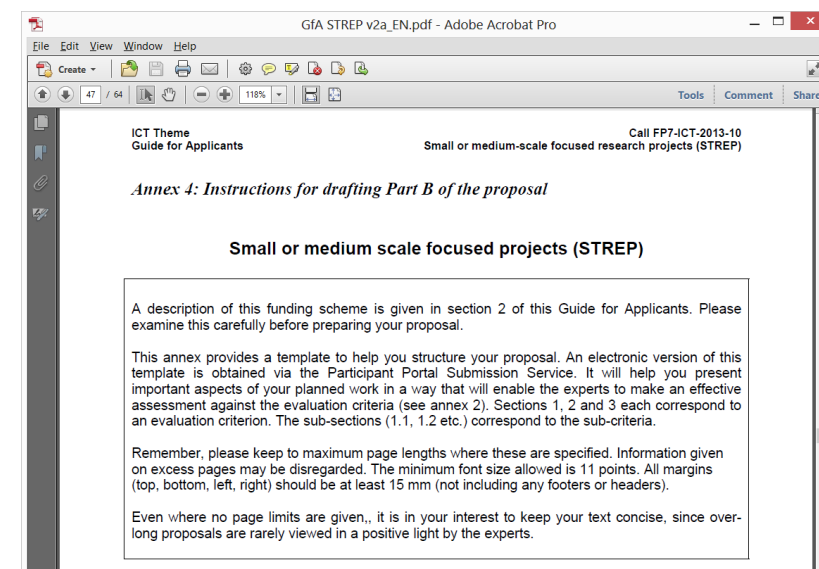
The screenshot displays the European Commission's Research & Innovation Participant Portal. The main heading is "RESEARCH & INNOVATION Participant Portal". The navigation menu includes "HOME", "FUNDING OPPORTUNITIES", "HOW TO PARTICIPATE", "EXPERTS", and "SUPPORT". The page is titled "ICT 2014 - Information and Communications Technologies" (H2020-ICT-2014-1) and is a sub-call of H2020-ICT-2014. Key details include:

Publication date	2013-12-11	Deadline Date	2014-04-23 17:00:00 (Brussels local time)
Total Call Budget	€658,500,000	Main Pillar	Industrial Leadership
Status	Open	OJ reference	OJ C361 of 11.12.2013

The topic is "Big data and Open Data Innovation and take-up" under the call "ICT-15-2014". The page also features sections for "Horizon 2020 calls", "FP7 & CIP Programmes", and "COSME". A "Specific Challenge" description is visible at the bottom: "The activities supported under this topic address the general technological and systemic data challenges that concern entire value chains and/or bridge across borders, languages, industries and sectors. The aim is to..."

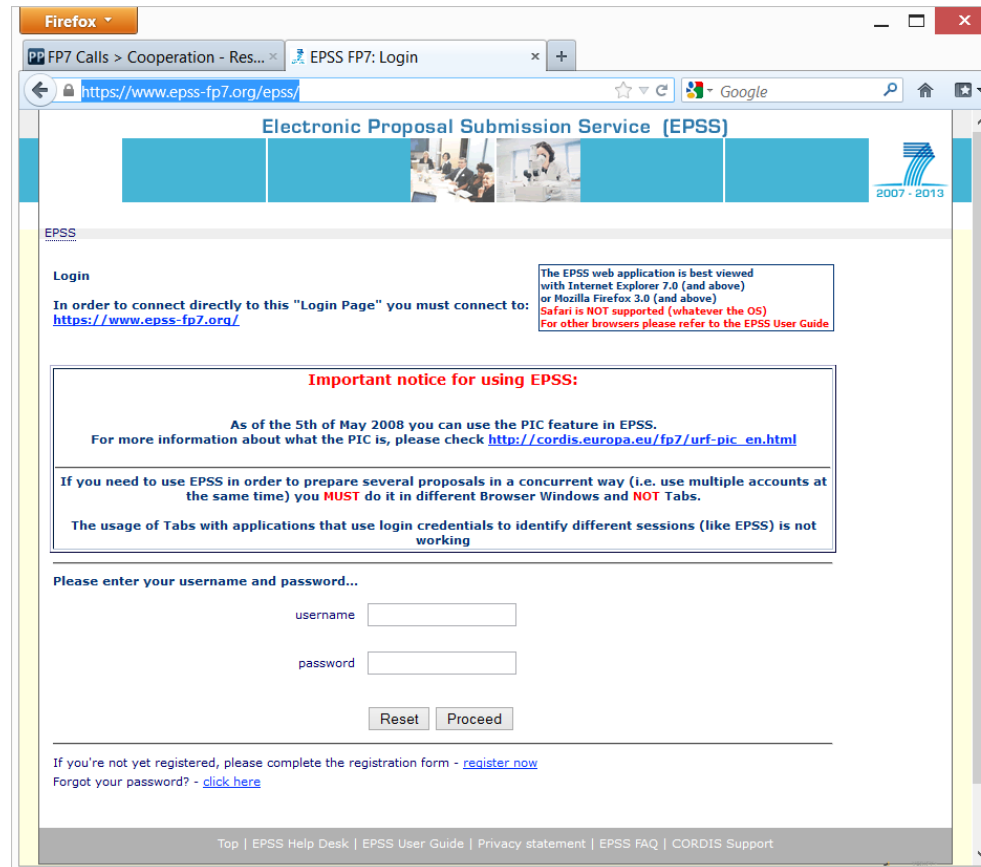
Kakšna je struktura predloga projekta?

- Prelog projekta je sestavljen iz dveh delov:
 - Del A – izpolnimo ga preko spletnega portala EPSS
 - Del A vsebuje administrativne informacije o konzorciju, kratek povzetek projekta, in generalne finančne informacije
 - Del B – vsebuje vsebino projekta in je dokument definiran s predlogo, ki je del informacijskega paketa



Del „A“ projektnega predloga

- DEMONSTRACIJA na strani <https://www.epss-fp7.org/epss/>



The screenshot shows a Firefox browser window displaying the EPSS login page. The browser's address bar shows the URL <https://www.epss-fp7.org/epss/>. The page header includes the title "Electronic Proposal Submission Service (EPSS)" and a logo with the years "2007 - 2013".

The main content area features a "Login" section with the following text:

Login
In order to connect directly to this "Login Page" you must connect to:
<https://www.epss-fp7.org/>

A warning box states: "The EPSS web application is best viewed with Internet Explorer 7.0 (and above) or Mozilla Firefox 3.0 (and above). Safari is NOT supported (whatever the OS). For other browsers please refer to the EPSS User Guide."

An "Important notice for using EPSS:" box contains the following information:

As of the 5th of May 2008 you can use the PIC feature in EPSS.
For more information about what the PIC is, please check http://cordis.europa.eu/fp7/urf-pic_en.html

If you need to use EPSS in order to prepare several proposals in a concurrent way (i.e. use multiple accounts at the same time) you **MUST** do it in different Browser Windows and **NOT** Tabs.
The usage of Tabs with applications that use login credentials to identify different sessions (like EPSS) is not working

Below the notice is a login form with the following fields and buttons:

Please enter your username and password...

username

password

Reset Proceed

At the bottom of the page, there are links for registration and password recovery:

If you're not yet registered, please complete the registration form - [register now](#)
Forgot your password? - [click here](#)

The footer contains navigation links: Top | EPSS Help Desk | EPSS User Guide | Privacy statement | EPSS FAQ | CORDIS Support

Del „B“ projektne predloga

- Del „B“ vsebuje vso vsebinsko informacijo projekta in je ključni del predloga
- Kazalo tipičnega dela „B“:
 - **Title Page & Abstract**
 - **Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call**
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - 1.3 S/T methodology and associated work plan
 - **Section 2. Implementation**
 - 2.1 Management structure and procedures
 - 2.2 Individual participants
 - 2.3 Consortium as a whole
 - 2.4 Resources to be committed
 - **Section 3. Impact**
 - 3.1 Expected impacts listed in the work programme
 - 3.2 Dissemination and/or exploitation of project results, and management of intellectual property
 - **Section 4. Ethical Issues**

Komentar k strukturi dela „B“

- Sekcija 1 dela „B“ je najpomembnejša in najzahtevnejša pri pisanju predloga projekta
 - Vsebuje vizijo projekta (1.1) in podrobno strukturo projekta (1.3)
 - Evalvatorji tipično preletijo sekciji 1.1 in 1.3 s pripadajočimi tabelami, kar da hiter in dokaj natančen vpogled v projekt.
 - Koristno je, da Sekcijo 1 napiše majhen krog ljudi in da je vsebinsko in organizacijsko zelo homogena
- Sekcija 2 predstavlja upravljanje projekta in konzorcij
 - Sekcija je pomembna v smislu predstavitve kdo sodeluje na projektu
- Sekcija 3 predstavlja kakšen vpliv bo imel projekt na zunanji svet (znanstveni in poslovni)
- Sekcija 4 je običajno kratka in se dotika etičnih vidikov projekta

Title Page & Abstract

Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call

1.1 Concept and objectives

1.2 Progress beyond the state-of-the-art

1.3 S/T methodology and associated work plan

Section 2. Implementation

2.1 Management structure and procedures

2.2 Individual participants

2.3 Consortium as a whole

2.4 Resources to be committed

Section 3. Impact

3.1 Expected impacts listed in the work programme

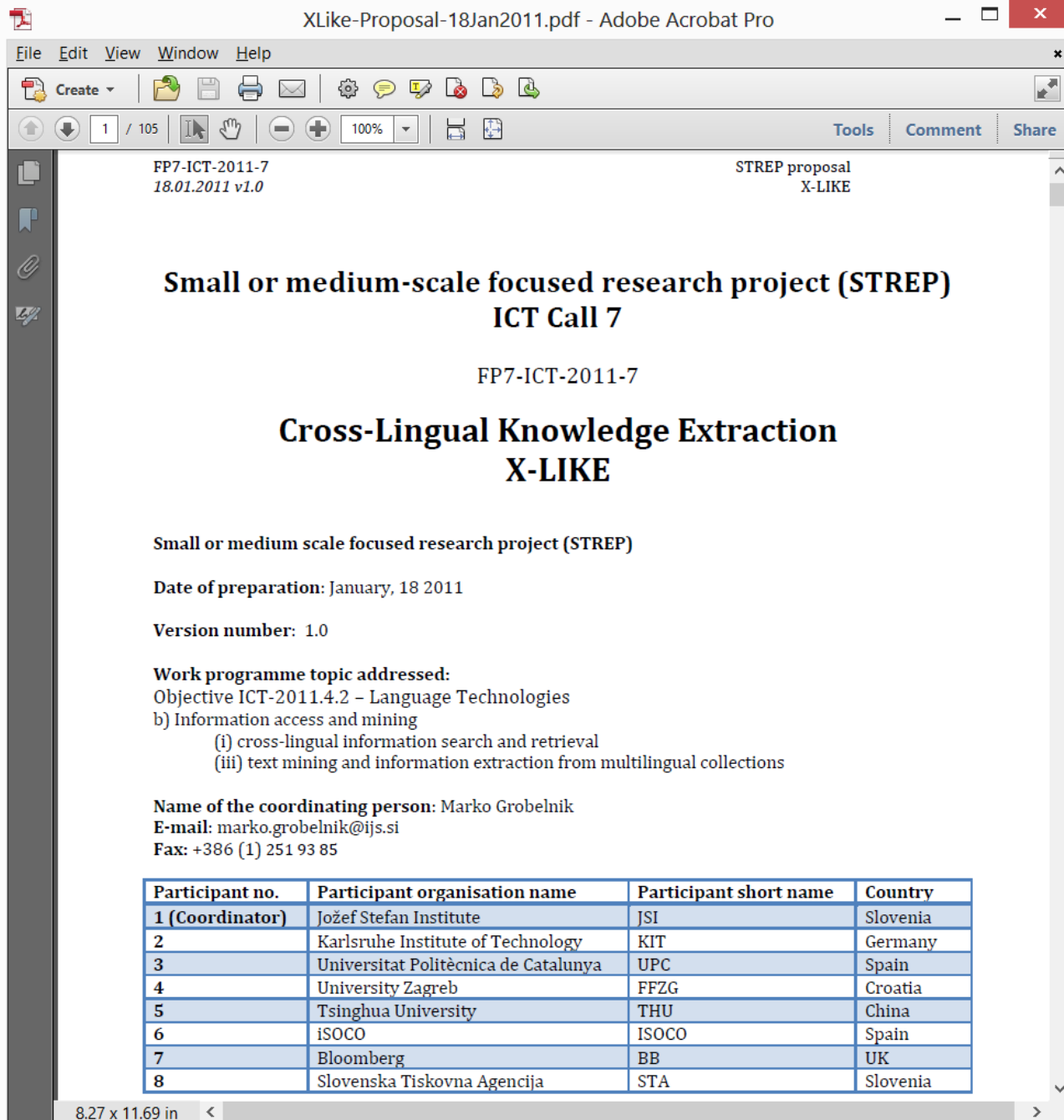
3.2 Dissemination and/or exploitation of project results, and management of intellectual property

Section 4. Ethical Issues

Naslovna stran in povzetek

Naslovna stran

- Naslovna stran (prva stran) vsebuje splošne informacije o projektu
 - Naslov in kratica projekta
 - Tip projekta
 - Oznako klica
 - Informacijo o koordinatorju
 - Institucije iz konzorcija



The screenshot shows the Adobe Acrobat Pro interface with a PDF document titled "XLike-Proposal-18Jan2011.pdf". The document content includes the following text:

FP7-ICT-2011-7
18.01.2011 v1.0

STREP proposal
X-LIKE

**Small or medium-scale focused research project (STREP)
ICT Call 7**

FP7-ICT-2011-7

**Cross-Lingual Knowledge Extraction
X-LIKE**

Small or medium scale focused research project (STREP)

Date of preparation: January, 18 2011

Version number: 1.0

Work programme topic addressed:
Objective ICT-2011.4.2 – Language Technologies
b) Information access and mining
(i) cross-lingual information search and retrieval
(iii) text mining and information extraction from multilingual collections

Name of the coordinating person: Marko Grobelnik
E-mail: marko.grobelnik@ijs.si
Fax: +386 (1) 251 93 85

Participant no.	Participant organisation name	Participant short name	Country
1 (Coordinator)	Jožef Stefan Institute	JSI	Slovenia
2	Karlsruhe Institute of Technology	KIT	Germany
3	Universitat Politècnica de Catalunya	UPC	Spain
4	University Zagreb	FFZG	Croatia
5	Tsinghua University	THU	China
6	iSOCO	ISOCO	Spain
7	Bloomberg	BB	UK
8	Slovenska Tiskovna Agencija	STA	Slovenia

8.27 x 11.69 in

Povzetek

- Povzetek (druga stran) je posebej pomemben del predloga projekta, ker ga vsi, ki pridejo v stik z dokumentom zagotovo preberejo
 - Napisan mora biti v brezhibni angleščini
 - Napisan mora biti jasno in vsebovati ključne prispevke projekta po raziskovalni in komercialni plati
 - Koristno je, da povzetek napiše nekdo, ki dobro razume cel projekt in razume mentaliteto evalvatorjev

FP7-ICT-2011-7
18.01.2011 v1.0

STREP proposal
X-LIKE

Proposal Abstract

The goal of the X-LIKE project is to develop technology to monitor and aggregate knowledge that is currently spread across global mainstream and social media, and to enable cross-lingual services for publishers, media monitoring and business intelligence.

In terms of research contributions, the aim is to combine scientific insights from several scientific areas to contribute in the area of cross-lingual text understanding. By combining modern computational linguistics, machine learning, text mining and semantic technologies we plan to deal with the following two key open research problems:

- **to extract and integrate formal knowledge from multilingual texts with cross-lingual knowledge bases, and**
- **to adapt linguistic techniques and crowdsourcing to deal with irregularities in informal language used primarily in social media.**

As an interlingua, knowledge resources from Linked Open Data cloud (<http://linkeddata.org/>) will be used with special focus on general common sense knowledge base CycKB (<http://www.cyc.com/>). For the languages where no required linguistic resources will be available, we will use a probabilistic interlingua representation trained from a comparable corpus drawn from the Wikipedia.

The solution will be applied on two case studies, both from the area of news. For the Bloomberg case study the domain will be financial news, while for the Slovenian Press Agency we will deal with general news. The technology developed in the project will be used to introduce cross-lingual and information from social media in services for publishers and end-users in the area of **summarization, contextualization, personalization, and plagiarism detection**. Special attention will be paid to analysing **news reporting bias** from multilingual sources.

The developed technology will be language-agnostic, while within the project we will specifically address English, German, Spanish, and Chinese as major world languages and Catalan and Slovenian as minority languages.

Title Page & Abstract

Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call

1.1 Concept and objectives

1.2 Progress beyond the state-of-the-art

1.3 S/T methodology and associated work plan

Section 2. Implementation

2.1 Management structure and procedures

2.2 Individual participants

2.3 Consortium as a whole

2.4 Resources to be committed

Section 3. Impact

3.1 Expected impacts listed in the work programme

3.2 Dissemination and/or exploitation of project results, and management of intellectual property

Section 4. Ethical Issues

Sekcija 1 dela „B“

Section 1. Scientific and/or technical quality, relevant to the topics addressed by the call

Sekcija 1 dela „B“

- Sekcija 1 je po predlogi sestavljena iz treh delov, dobra praksa pa je, da vsakega od treh delov še razgradimo naprej:
- **Section 1. Scientific and/or technical quality, relevant to the topics addressed by the call**
 - **1.1 Concept and objectives**
 - **1.1.1 Introduction**
 - **1.1.2 Vision and Objectives**
 - **1.1.3 Indicators and Metrics**
 - **1.1.4 Relevance to the call**
 - **1.2 Progress beyond the state-of-the-art**
 - **1.2.X Progress beyond the state-of-the-art in ...**
 - **1.3 S/T methodology and associated work plan**
 - *Describe the overall strategy of the work plan*
 - *Show the timing of the different WPs and their components (Gantt chart or similar).*
 - *Provide a detailed work description broken down into work packages:*
 - *Work package list (please use table 1.3a);*
 - *Deliverables list (please use table 1.3b)*
 - *List of milestones (please use table 1.3c)*
 - *Description of each work package (please use table 1.3d)*
 - *Summary effort table (1.3e)*
 - *Provide a graphical presentation of the components showing their interdependencies (Pert diagram or similar)*
 - *Describe any significant risks, and associated contingency plans*

1.1.1 Introduction

- V uvodu opišemo na učinkovit in jasen način celo idejo projekta, ki ga predlagamo
- Cilj je, da evalvatorja na kratek in opisen način informiramo (a) kaj bomo pravzaprav delali, (b) zakaj je to pomembno, in (c) kako bomo to izvedli
- Ta sekcija je še posebej pomembna, ker jo zagotovo prebere večina evalvatorjev

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - **1.1 Concept and objectives**
 - 1.2 Progress beyond the state-of-the-art
 - 1.3 S/T methodology and associated work plan

FP7-ICT-2011-7
18.01.2011 v1.0

STREP proposal
X-LIKE

1. SCIENTIFIC AND/OR TECHNICAL QUALITY, RELEVANT TO THE TOPICS ADDRESSED BY THE CALL

1.1. CONCEPT AND OBJECTIVES

1.1.1. INTRODUCTION

Automatic text understanding has been a research goal for many years and it seems that it will continue to be explored for many years to come. The problem is hard because of two main issues: (a) the ungraspable nature of the world we are living in, and (b) the dynamic nature of human languages, trying to capture abstractions about the world for the purpose of communication between humans.

On this long path of exploring language and simultaneously modeling the world, we have transitioned through many stages. From the early years, when linguists were exploring the structure of words and sentences, and philosophers were discussing ontology, up till today, when a large part of written information is available to almost everybody and modern communication allows researchers to join their efforts and exploit opportunities offered by technology. Still, the apparent opportunities seem to be much bigger than the actual depth of the solutions developed in recent years in the field of text understanding.

One of the key properties of natural languages is redundancy in the encoded information and the structure used. As a consequence, different techniques can extract different aspects of information from a text. They range from simple techniques such as character counting, to more sophisticated ones such as linear algebra, to the advanced techniques which exploit the structural aspects of text. Many of these techniques deliver something useful and solve somebody's problem. Some examples of such problems are: language identification (solved with character counting), document categorization (solved with linear algebra methods), question-answering (solved typically with shallow linguistic methods), and reasoning (solved typically using logic).

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - **1.1 Concept and objectives**
 - 1.2 Progress beyond the state-of-the-art
 - 1.3 S/T methodology and associated work plan

1.1.2 Vision and Objectives

- V sekciji razgradimo projekt na posamezne komponente v smislu vizije in ciljev, ki jih želimo doseči
- Predstavimo arhitekturo projekta (po možnosti z grafičnim prikazom)
- Vsako od komponent na kratko opišemo
- Cilj je prepričati evalvatorja, da razume našo vizijo in ji tudi verjame

1.1.2. VISION AND OBJECTIVES

The vision of the X-LIKE project is to develop technologies to monitor and aggregate knowledge spreading across global mainstream and social media and to enable cross-lingual services for publishers, media monitoring and business intelligence.

- The key tangible result of the project will be an "X-LIKE Software Toolkit" consisting from:
- A set of components partly developed within the project and partly taken and integrated from existing linguistic repositories
 - An API offering X-LIKE functionality as a web-service
 - A web front-end on the top of the API demonstrating capabilities of the Toolkit.

The X-LIKE Toolkit will serve as a basis for use case applications described in more detail in the section 1.1.2.2. Use cases applications are supposed to be implemented on the top of X-LIKE Toolkit API.

Functionally, X-LIKE Toolkit consists from a pipeline of six separated stages as shown on Figure 1.1.d. Each stage corresponds to one of the workpackages as designated on the top of the figure.

An intuitive way of understanding the information flow through the pipeline is that each next stage uses the data and information from one or several previous stages and adds something on the top of it.

Proposal Part B: page [8] of [105]

FP7-ICT-2011-7
18.01.2011 v1.0

STREF proposal
X-LIKE

While stages 1 and 6 are more engineering oriented, stages 2 to 5 deal with research topics.

In terms of types of transformations on language, stages 1 and 2 are operating with documents in separate languages. In stages 3 and 4 documents get transformed into semantic and keyword interlingua representations. In stage 5 the interlingua representation is used to produce higher level results. In stage 6 information created in the previous stages is being delivered into applications.

Next we briefly describe each of the stages through its key contributions, technical challenges and scientific insights and solutions. Detailed description of each of the stages is found in section 1.3.2.

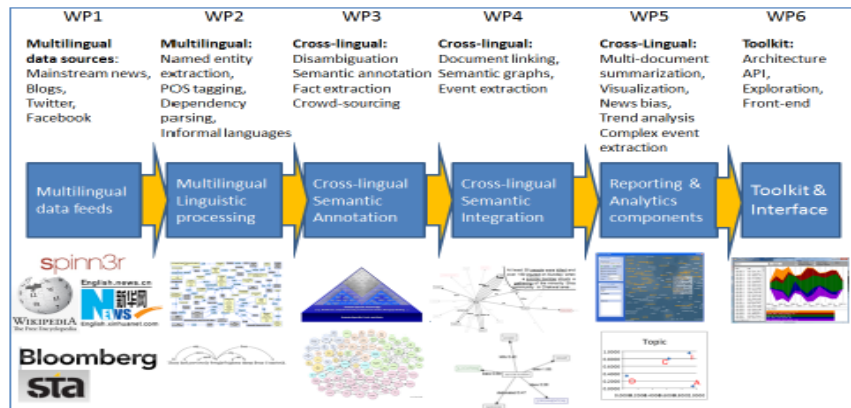


Figure 1.1.d: Schematic diagram of functional stages in X-LIKE pipeline.

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - **1.1 Concept and objectives**
 - 1.2 Progress beyond the state-of-the-art
 - 1.3 S/T methodology and associated work plan

1.1.3 Indicators and Metrics

- V sekciji navedemo zelo konkretne merljive cilje, ki jih želimo doseči v projektu
 - Merljive cilje strukturiramo po vsebinskih področjih
 - Za vsak cilj navedemo merljivo oceno (če je mogoče, številčno)
 - Tak način opisovanja ciljev zelo konkretizira načrte in naredi celotno besedilo bolj pregledno

1.1.3. INDICATORS AND METRICS

With the aim of defining the expectations of the project in a measurable way, X- LIKE has filled-in the table below with target outcomes for different categories.

Target outcome	Indicator	Target
OBJECTIVE 1		
Quality of extracted knowledge from multilingual texts		
General aspects	International and National scientific papers published	>=5
Linguistic part of pipeline	Precision and recall of properly resolved anaphoras	Near state-of-the-art performance on all covered languages
	Precision and recall of triples extracted from text	Near state-of-the-art performance on all covered languages
Semantic part of pipeline	Precision and recall of named entities properly resolved into the relevant knowledge bases	Methods improving X% over state-of-the-art as of M1 of the project: <ul style="list-style-type: none"> • X < 3: unsatisfactory • 3 ≤ X < 15: satisfactory • X ≥ 15: very satisfactory
	Precision and recall of disambiguated concepts and relationships into the relevant knowledge bases	Methods improving X% over state-of-the-art as of M1 of the project: <ul style="list-style-type: none"> • X < 3: unsatisfactory • 3 ≤ X < 15: satisfactory • X ≥ 15: very satisfactory
Cross lingual part of pipeline	Accuracy and quality of translation of lexical groundings in knowledge based	Methods improving X% over state-of-the-art as of M1 of the project: <ul style="list-style-type: none"> • X < 3: unsatisfactory • 3 ≤ X < 15: satisfactory • X ≥ 15: very satisfactory
	Accuracy in linking of cross-lingual documents	Methods improving X% over state-of-the-art as of M1 of the project:

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - **1.1 Concept and objectives**
 - 1.2 Progress beyond the state-of-the-art
 - 1.3 S/T methodology and associated work plan

1.1.4 Relevance to the call

- V sekciji pojasnimo zakaj naš predlog projekta ustreza klicu na katerega pošiljamo dokument.
- Načinov, da to opišemo je več
 - pogost način je, da skopiramo besedilo iz klica, kjer označimo ključne besede in jih v nadaljevanju razložimo v relaciji do našega predloga projekta – tak način pomaga evalvatorju razumeti relacijo med projektom in klicom

1.1.4. RELEVANCE TO THE CALL

The project is aligned and emphasize most of the objectives in general part of the "Objective ICT-2011.4.2 - Language Technologies" call and specifically focuses on the item "(b) Information access and mining".

Specifically, from the general part of the call ("Objective ICT-2011.4.2 - Language Technologies") we specifically address the following highlighted parts:

*"Projects shall address **multiple languages** and cater for **written** and/or spoken language as appropriate. Technologies must be **adaptive**, they must handle language **in its various uses**, cope **efficiently with massive volumes**, and be **embedded within information flows**. **Contextualisation** is a common requirement and so is **personalisation**."*

- **Multiple languages** - within the project we deal with 6 primary languages (English, German, Spanish, Slovene, Catalan, Chinese) while the technology will be further tested on several more languages depending on needs of the use cases.
- **Written language** - within the project we will concentrate only on written language.
- **Adaptive technologies** - the project technology will be highly adaptive since it will use machine learning techniques on most of the steps; the aim is to minimize the cost and time of an introduction of new languages into the project pipeline.
- **Language in its various uses** - the project will specifically address informal languages (like used in Twitter, Facebook and blogs) as well as traditional formal language (as used in mainstream media).
- **Efficiently with massive volumes** - the whole project infrastructure is specifically designed to deal with real-life social media feeds; initial experiments show ability to deal with over 100Gb texts within 30-40 million documents per day.
- **Embedded within information flows** - X-LIKE technology addresses information flows by analysing streams of documents from publishers like Bloomberg, Slovenian Press Agency, and Xinhua Press Agency as well as streams of data from Twitter and public part of Facebook. Information flows will be analysed and automatically structured into events and story-lines.

1.2 Progress beyond the state-of-the-art

- Sekcija vsebuje množico podsekcij, ki razlagajo vsa relevantna (raziskovalna) podpodročja, ki so potrebna za razumevanje predloga projekta
- S sekcijo:
 - Pokažemo da razumemo tematike
 - Izobrazimo evalvatorje o področjih, s katerimi se ukvarjamo v predlogu projekta
 - Nakažemo, kako bomo vsako od teh področij razširili z inovacijami

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - **1.2 Progress beyond the state-of-the-art**
 - 1.3 S/T methodology and associated work plan

1.2.	Progress beyond the State-of-the-Art.....	18
1.2.1.	Progressing State-of-the-Art in Multilingual Data Sources and Streams	18
1.2.2.	Progressing State-of-the-Art in Multilingual Dependency Parsing	18
1.2.3.	Progressing State-of-the-Art in Text Processing of Informal Documents	19
1.2.4.	Progressing State-of-the-Art in Cross-lingual Text Processing.....	21
1.2.5.	Progressing State-of-the-Art in Cross-lingual Knowledge Bases	22
1.2.6.	Progressing State-of-the-Art in Linking to Cross-lingual Knowledge Bases	23
1.2.7.	Progressing State-of-the-Art in Cross-lingual Disambiguation.....	24
1.2.8.	Progressing State-of-the-Art in Cross-lingual Document Linking	25
1.2.9.	Progressing State-of-the-Art in Semantic Graph and Event Extraction.....	26
1.2.10.	Progressing State-of-the-Art in News Reporting Bias Detection.....	27
1.2.11.	Progressing State-of-the-Art in Document Organisation and Visualization.....	28

1.2.X Progress beyond the state-of-the-art in ...

- Vsaka od podsekcij opisuje eno področje raziskav oz. aktivnosti na projektu
- Tipično vsebuje dva dela:
 - Opis področja, kjer pokažemo razumevanje področja in izobrazimo evalvatorje
 - Opis inovacij, ki jih nameravamo razviti v okviru projekta

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - **1.2 Progress beyond the state-of-the-art**
 - 1.3 S/T methodology and associated work plan

1.2.7. PROGRESSING STATE-OF-THE-ART IN CROSS-LINGUAL DISAMBIGUATION

In this section we focus on research that is concerned with a major challenge when grounding semantic facts in a knowledge base, namely disambiguation. In case of ontological information extraction, the extracted facts have to be mapped to entities and relations in the ontology. In many cases, this mapping is ambiguous. E.g. since many location names, companies, or product names are ambiguous, finding the intended meaning of a word is often a difficult task. This likewise applies to relations or other semantic facts.

Word Sense Disambiguation (WSD) has been employed in the Semantic Web community to improve specific tasks like ontology mapping or data annotation with respect to a reference ontology, but also Semantic Web resources from the LOD cloud have been used to aid WSD, the most popular being WordNet. WSD techniques have been previously used to validate ontology mappings, by analysing the semantics of the ontological terms; they exploit ontological context as well as information provided by WordNet [1]. Aside from WordNet, another semantic resource, namely Wikipedia has been used for building sense tagged corpora [2]. A Wikipedia-based sense tagged corpus is generated for a subset of the Senseval-2 and Senseval-3 datasets, which is further used to train a classifier; the results achieved by this classifier exceed the accuracy of a baseline (which assumes the availability of a sense tagged corpus) that selects the most frequent word sense.

Supervised approaches to WSD have recorded better results in the past semantic evaluation workshops, compared to their unsupervised and knowledge-based counterpart. NUS-PT [3], the best

....

Progressing beyond the state-of-the-art

X-LIKE will work towards advancing the-state-of-the-art disambiguation from two directions and, most importantly, emphasising cross-lingual scenarios for disambiguation. First, approaches for disambiguation which rely on lexical groundings and descriptions will be generalized to cover cross-lingual scenarios. Second, this work will be combined with ontological structure and constraints to (a) increase the precision of disambiguation and (b) to compensate for the case where little or no lexical descriptions are available.

1.3 S/T methodology and associated work plan

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

- Sekcija 1.3 je ključna za podrobno razumevanje projekta
 - Gre za najbolj tehnični opisa projekta z opisom poteka projekta iz katerega evalvatorji vidijo kaj se bo v projektu zares počelo
 - V tej sekciji opišemo projekt po tehnični in znanstveni plati iz več vidikov
 - Srce projektne predloga je sekcija „**Description of each work package**“, ki opisuje podrobno vsak delovni paket („**Work package**“)

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

Komponente sekcije 1.3

- Navodilo o strukturi sekcije 1.3:
 - *Please present your plans as follows:*
 - *i) Describe the overall strategy of the work plan (Maximum length – one page)*
 - *ii) Show the timing of the different WPs and their components (Gantt chart or similar).*
 - *iii) Provide a detailed work description broken down into work packages:*
 - *Work package list (please use table 1.3a);*
 - *Deliverables list (please use table 1.3b);*
 - *List of milestones (please use table 1.3c)*
 - *Description of each work package (please use table 1.3d)*
 - *Summary effort table (1.3e)*
 - *iv) Provide a graphical presentation of the components showing their interdependencies (Pert diagram or similar)*
 - *v) Describe any significant risks, and associated contingency plans*

Arhitektura projekta

- V sekciji predstavimo arhitekturo projekta
 - Z grafičnim prikazom na katerega se lahko večkrat sklicujemo v dokumentu
 - Z podrobnim in jasnim opisom komponent arhitekture
- Posebej tehničnim evalvatorjem lahko opis arhitekture dobro pojasni predlagani projekt

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

1.3.1.1. PIPELINE ARCHITECTURE

The components developed in the project fit into pipeline architecture. Each of the components processes and enriches the content. Well-defined, public interface documentations enable the usage of the components outside of the pipeline, so that they can be reused and recombined for alternative requirements. The components are modular and can be recombined. The use cases have their respective combination of modules, especially using components from WP5 and WP6 directly (and those of WP1-4 directly).

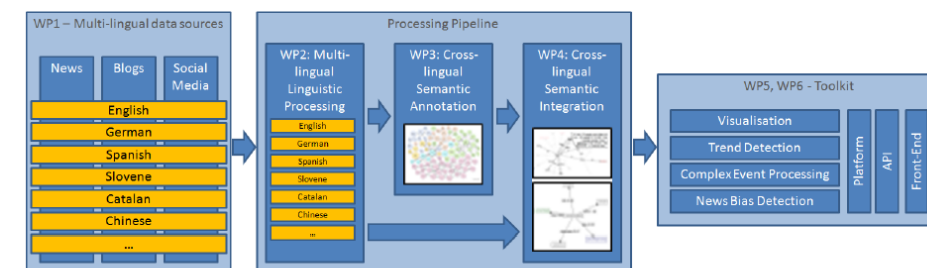


Figure 1.3a: Overview of the project pipeline architecture

The components in the project pipeline architecture are covered by WP1-5 as follows.

WP1 Definition and Data Provision collects and makes available for the rest of the pipeline the **multilingual data feeds**. WP1 gathers and systemizes existing resources and requirements, and develops common-agreed models of all data sources.

WP2 Multilingual Linguistic Processing prepares a linguistic pipeline for formal languages, and in parallel develops linguistic tools to deal with informal languages (informal language, within X-LIKE, refers to short items, often consisting of just fragments, without necessary adhering to standard

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

OVERVIEW OF WORK PACKAGES

- V sekciji na kratko opišemo vsak delovni paket (workpackage) preko nekaj vidikov:
 - Cilji
 - Vloga v projektu
 - Glavni izzivi
 - Inovacije
 - Kriteriji za uspešnost

WP2: Multilingual Linguistic Processing (Lead UPC)

Goal: The main goal of WP2 is to establish linguistic infrastructure for each of the supported languages, ranging from language identification, via lemmatisation, part-of-speech tagging and named entity extraction to dependency parsing. Separate infrastructure will be built for dealing with formal languages used in sources such as articles from mainstream news sources and for informal languages used more social media sources such as micro-blogs.

Thus, we pursue two main goals in WP2:

- Preparing a linguistic pipeline for formal languages for all required languages.
- Developing linguistic tools to deal with informal languages for all required languages.

To achieve these goals, X-LIKE will build on top of existing linguistic infrastructure when dealing with formal language and develop new techniques to deal with informal language. In particular, we will analyse the styles of language as appearing in social media and develop methodologies and tools, based on machine learning and crowd sourcing, which can deal with informal language while still extracting enough linguistic structure required by the rest of the project.

Role in the X-LIKE system: WP2 belongs to the Research Activity and will provide a linguistic processing toolbox for:

- WP3 to support annotation and disambiguation activities;
- WP4 to support extracting predicate-argument relations for constructing semantic graphs.

Challenges: The main challenges are:

- Scaling existing linguistic tools across all required languages;
- Scaling triple extraction on top of linguistic tools across all required languages;
- Extracting sufficient structure from informal languages to support triple extraction.

Innovation: The key innovation in WP2 will be developed methodologies, techniques and tools for dealing with informal language. Currently available solutions either rely on constraints usually available in formal language or do not deal with the structure of informal language on a level sufficient for subsequent work packages.

Success criteria: Success criteria of WP2 are providing sufficient linguistic support to the dependent work packages and to move beyond current state-of-the-art in dealing with informal language.

WORK PACKAGE LIST

- V sekciji navedemo vse delovne pakete

- Kdo jih vodi
- Čas trajanja
- Čas začetka
- Čas konca

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

1.3.4. WORK PACKAGE LIST

Work package no.	Work package title	Type of activity	Lead part. no.	Lead part. short name	Person-months	Start month	End month
1	Definition and Data Provision	RTD	1	JSI	57	M1	M27
2	Multilingual Linguistic Processing	RTD	3	UPC	90	M1	M30
3	Cross-lingual Semantic Annotation	RTD	2	KIT	81	M7	M30
4	Cross-lingual Semantic Integration	RTD	1	JSI	57	M7	M30
5	Reporting and Analytics	RTD	5	THU	54	M1	M30
6	Integration and Toolkit	RTD	6	ISOCO	67	M1	M36
7	Use cases and Evaluation	RTD	7	BB	46	M1	M36
8	Dissemination, Exploitation and Community Building	RTD	4	FFZG	43	M1	M36
9	Management	MGT	1	JSI	26	M1	M36
	TOTAL				521		

DELIVERABLE LIST

- V spisku navedemo za vsak deliverable v projektu
 - Oznako
 - Naslov
 - Delovni paket
 - Dostopnost
 - Mesec zaključka

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

1.3.5. DELIVERABLE LIST

Del. No	Deliverable name	WP no	Nature	Diss. Level	Delivery
D1.1.1	Report on library on the existing technology and data	WP1	R	PU	M3
D1.2.1	Requirements for early prototype	WP1	R	PU	M3
D1.2.2	Requirements for demonstrator	WP1	R	PU	M15
D1.2.3	Requirements for fully functional prototype	WP1	R	PU	M27
D1.3.1	Early prototype of data infrastructure	WP1	P	PU	M3
D1.3.2	Final prototype of data infrastructure	WP1	P	PU	M15
D2.1.1	Shallow linguistic processing prototype	WP2	P	PU	M6
D2.2.1	Early deep linguistic processing prototype	WP2	P	PU	M12
D2.2.2	Final deep linguistic processing prototype	WP2	P	PU	M21
D2.3.1	Informal language analysis report and prototype	WP2	R+P	PU	M12
D2.4.1	Early informal language structure extraction prototype	WP2	P	PU	M21
D2.4.2	Final informal language structure extraction prototype	WP2	P	PU	M30
D3.1.1	Early text annotation prototype	WP3	P	PU	M12
D3.1.2	Final text annotation prototype	WP3	P	PU	M21
D3.2.1	Early ontological word-sense-disambiguation prototype	WP3	P	PU	M15

DESCRIPTION OF WORK PACKAGES (1/2)

- V sekciji opišemo vsak delovni paket na najbolj podroben način
 - Delovni paket opišemo s partnerji in s koliko človek meseci sodelujejo
 - Delovni paket na kratko opišemo v sekciji „Objectives“
 - Vsak paket je razdeljen na naloge

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

Work package number	2	Start date or starting event:				
Work package title	Multilingual Linguistic Processing					
Activity type	RTD					
Participant number	2	3	4	5		
Participant short name	KIT	UPC*	FFZG	THU		
Person-months per participant	6	48	12	24		

Objectives
For each of the languages identified in WP1 as required for the case studies:
<ul style="list-style-type: none">• To prepare and develop tools for shallow and deep linguistic processing of formal language corpora.• To analyse informal languages and develop tools for extracting structured information from informal language corpora.

Description of work
T2.1: Shallow linguistic processing of formal language (Lead: FFZG, Contributors: UPC, THU)
This task will prepare tools for shallow processing across all required languages. The tools will include language detection, tokenization, stemming and lemmatization, part-of-speech tagging, bag-of-words generation and named entity extraction. Named entity extraction will be done with the help of external multilingual resources such as multilingual Wikipedia corpus and Linked-Open-Data providing large coverage of relevant entities across many languages. The task will mostly rely on existing monolingual tools and language resources.
T2.2: Deep linguistic processing of formal language (Lead: UPC, Contributors: FFZG, THU)
This task will prepare and develop tools for deep processing of all required languages. This includes deep parsing, co-reference resolution (de-duplication of entity references), pronoun resolution and extraction of grammatical relations from parse trees. The task will rely on existing monolingual tools for deep parsing and co-reference resolution. New tools will be developed for extraction of predicative grammatical relations, like agent-predicate-object triples and adjunct relations such as temporal, locative, causal, etc. We will extract representations of such relations at different levels, from purely syntactic but language dependent, to semantic abstractions based in semantic roles that are invariant across different languages.
T2.3: Analysis of informal languages (Lead: UPC, Contributors: THU, KIT)
The main goal of this is to analyze informal languages from social media feeds (blogs, Twitter, Facebook) and to develop initial set of tools and resources for working with them. First, analysis will be done to identifying specifics of informal languages. Based on this, language resources required by T2.4 (e.g. lexicons) will be created using statistical methods and crowd sourcing. A special emphasis will be given to creating parallel collections of informal textual expressions (words, phrases, sentences) paired with their formal counterpart.

DESCRIPTION OF WORK PACKAGES (2/2)

- Vsaka naloga ima oznako, naslov, spisek sodelujočih partnerjev, opis in spisek deliverablov

- Vsak “deliverable” ima oznako naslov, rok oddaje, tip, dosegljivost, odgovorno institucijo in kratek opis

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

Description of work

T2.1: Shallow linguistic processing of formal language (Lead: FFZG, Contributors: UPC, THU)

This task will prepare tools for shallow processing across all required languages. The tools will include language detection, tokenization, stemming and lemmatization, part-of-speech tagging, bag-of-words generation and named entity extraction. Named entity extraction will be done with the help of external multilingual resources such as multilingual Wikipedia corpus and Linked-Open-Data providing large coverage of relevant entities across many languages. The task will mostly rely on existing monolingual tools and language resources.

T2.2: Deep linguistic processing of formal language (Lead: UPC, Contributors: FFZG, THU)

This task will prepare and develop tools for deep processing of all required languages. This includes deep parsing, co-reference resolution (de-duplication of entity references), pronoun resolution and extraction of grammatical relations from parse trees. The task will rely on existing monolingual tools for deep parsing and co-reference resolution. New tools will be developed for extraction of predicative grammatical relations, like agent-predicate-object triples and adjunct relations such as temporal, locative, causal, etc. We will extract representations of such relations at different levels, from purely syntactic but language dependent, to semantic abstractions based in semantic roles that are invariant across different languages.

T2.3: Analysis of informal languages (Lead: UPC, Contributors: THU, KIT)

The main goal of this is to analyze informal languages from social media feeds (blogs, Twitter, Facebook) and to develop initial set of tools and resources for working with them. First, analysis will be done to identify specifics of informal languages. Based on this, language resources required by T2.4 (e.g. lexicons) will be created using statistical methods and crowd sourcing. A special emphasis will be given to creating parallel collections of informal textual expressions (words, phrases, sentences) paired with their formal counterpart.

D2.2.1 Early deep linguistic processing prototype (M12, Prototype, PU, Responsible: UPC)

This deliverable will provide an initial version tools for deep linguistic processing of required languages with sufficient language infrastructure at the project start.

D2.2.2 Final deep linguistic processing prototype (M21, Prototype, PU, Responsible: UPC)

This deliverable will provide a final version of tools for deep linguistic processing of all required languages.

D2.3.1 Informal language analysis report and prototype (M12, Report + Prototype, PU, Responsible: UPC)

This deliverable will provide an analysis on dealing with informal languages and specific language resources required by T2.4.

SUMMARY OF STAFF EFFORT

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

- V tabeli navedemo koliko človek mesecev bo sodeloval kateri od partnerjev na katerem delovnem paketu
 - Pomembno je, da so številke uravnotežene in da so vsa izstopanja strokovno opravičljiva

1.3.7. SUMMARY OF STAFF EFFORT

The leading partner for each work package is marked with '*'.

	JSI	KIT	UPC	FFZG	THU	ISOCO	BB	STA	Total
WP1	10*	10	3	3	3	12	8	8	57
WP2	0	6	48*	12	24	0	0	0	90
WP3	9	30*	12	24	6	0	0	0	81
WP4	33*	12	12	0	0	0	0	0	57
WP5	12	12	0	0	18*	12	0	0	54
WP6	9	3	3	3	11	38*	0	0	67
WP7	0	0	0	10	0	0	18*	18	46
WP8	5	8	2	4*	2	10	6	6	43
WP9	26*	0	0	0	0	0	0	0	26
Total	104	81	80	56	64	72	32	32	521

LIST OF MILESTONES

- V tabeli naštejemo ključne mejnike v projektu
- Vsak gradnik ima ime, spisek delovnih paketov na katere se nanaše, čas in kratek opis
- Mejniki so pomembni za razumevanje strukture projekta in morajo biti usklajeni s preostankom projekta
 - Delovni paketi in deliverabli se tipično ravnajo po mejnikih

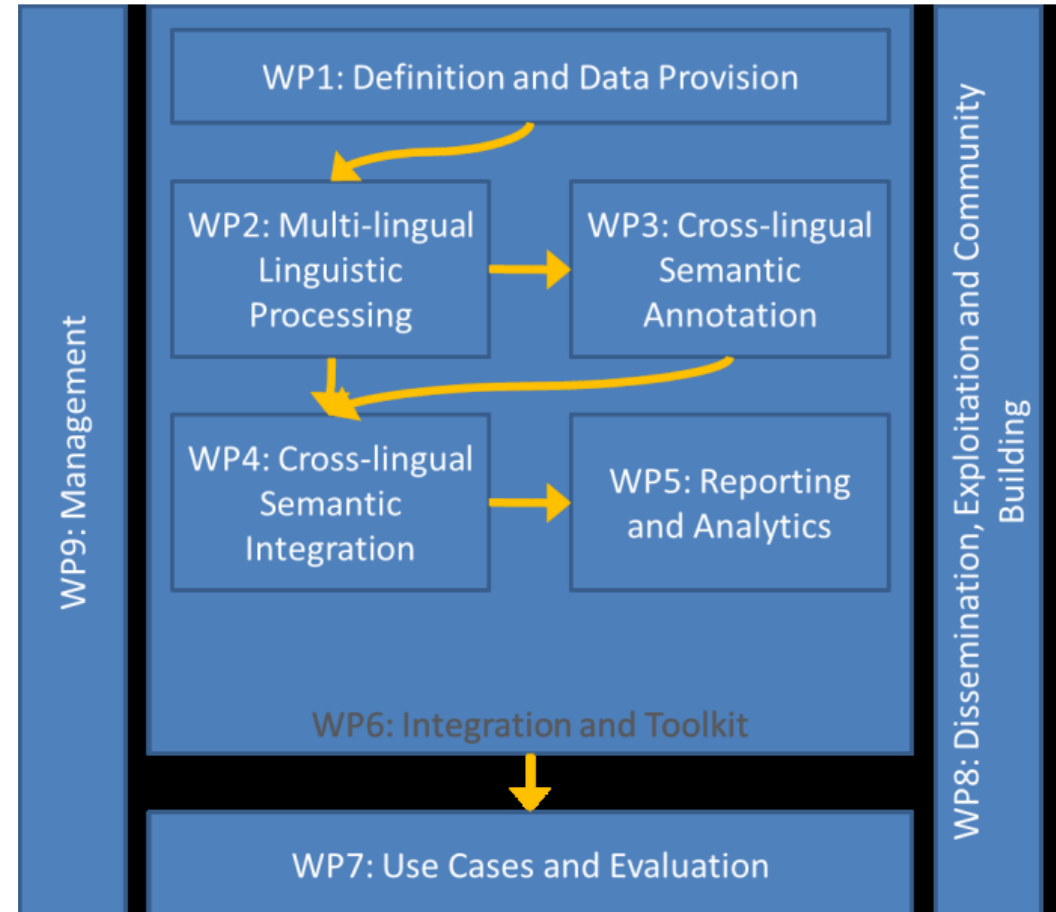
- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**

1.3.8. LIST OF MILESTONES				
Milestone no.	Milestone name	Work packages involved	Expected date	Means of verification
1	Inception check	WP1, WP2, WP6, WP8, WP9	M3	Project Management: Project satisfactorily commenced, all partners fully resourcing project Dissemination and Exploitation: Website established Requirements: Initial Requirements captured
2	Project establishment	All	M12	RTD: Initial components available, Pipeline functional Dissemination and Exploitation: Initial presentation material Use cases: Initialized, First data evaluations finished Project Management: Deliver initial progress report
3	Project refinement	All	M24	RTD: Demonstrator components available, published and documented, Pipeline refined and open for external development Use cases: Experience from initial feed back round fully incorporated, demonstrator version evaluated Project Management: Carry out risk assessment
4	Project validation	All but WP1	M30	RTD: Final version of pipeline components completed Use cases: Use case experiences fed back into component and use case development
5	Project impact	WP6, WP7, WP8, WP9	M36	Dissemination and Exploitation: Exploitation plans from all industrial partners in place Use cases: Evaluation finished and reported Project Management: Completion of all project objectives, technical, commercial, and dissemination

INTERDEPENDENCIES AMONG COMPONENTS (PERT CHART)

- „Pertov“ diagram opisuje soodvisnosti med delovnimi paketi (včasih celo med nalogami)

- Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call
 - 1.1 Concept and objectives
 - 1.2 Progress beyond the state-of-the-art
 - **1.3 S/T methodology and associated work plan**



Sekcija 2 dela „B“

Section 2. Implementation

Title Page & Abstract

Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call

1.1 Concept and objectives

1.2 Progress beyond the state-of-the-art

1.3 S/T methodology and associated work plan

Section 2. Implementation

2.1 Management structure and procedures

2.2 Individual participants

2.3 Consortium as a whole

2.4 Resources to be committed

Section 3. Impact

3.1 Expected impacts listed in the work programme

3.2 Dissemination and/or exploitation of project results, and management of intellectual property

Section 4. Ethical Issues

Sekcija 2 dela „B“

- Sekcija 2 vsebuje informacije, ki so pomembne za izvedbo projekta iz organizacijske strani
- Vsebuje 4 podsekcije:
 - ***2.1 Management structure and procedures***
 - ***2.2 Individual participants***
 - ***2.3 Consortium as a whole***
 - ***2.4 Resources to be committed***

2.1 Management structure and procedures

- Sekcija vsebuje opis vodenja projekta, kar vključuje
 - Definicijo vlog posameznikov in projektnih odborov
 - Komunikacijo znotraj projekta
 - Postopke odločanja in razreševanje konfliktnih situacij
 - Preverjanje kontrole kakovosti
 - Glavne rizike v izvajanju projekta in načine zavarovanja pred njimi
- Praksa je, da pisci predloga projekta vzamejo že izdelane opise vodenja projekta in ga prilagodijo glede na specifične trenutnega projekta



2.2 Individual participants

- V sekciji opišemo vse sodelujoče institucije
- Vsako institucijo opišemo z dvema glavnima rubrikama:
 - Kratkim opisom institucije, ki je prilagojen vlogi v projektu
 - Kratke življenjepise ljudi (dveh do treh), ki bodo sodelovali na projektu
- Opcijsko lahko dodamo tudi rubrike:
 - Vlogo institucije na projektu
 - Motiv za sodelovanje na projektu
 - Izraba rezultatov po končanem projektu

2.2.1. JOZEF STEFAN INSTITUTE, SLOVENIA (JSI)



Jožef Stefan Institute (JSI) is the leading research institution for natural sciences in Slovenia having over 900 researchers within 25 departments working in the areas of computer science, physics, and chemistry and biology. **Artificial Intelligence Laboratory**, having approx. 30 researchers, is

one of the largest European research groups working in the areas of machine learning, data mining, language technologies, semantic technologies and recently sensor networks. One of the key items on the research agenda is combining modern statistical data analytic techniques with more traditional logic based knowledge representations and reasoning techniques with the purpose to progress in solving complex problems such as text understanding, large scale probabilistic reasoning, building broad coverage knowledge bases, and dealing with scale. The members have developed several software tools for multimodal data analysis, among others: the *Text-Garden* suite of text mining tools, the *OntoGen* system for ontology learning,¹³ the *Document-Atlas* for complex visualization, the *AnswerArt* system for semantic search over large textual databases,¹⁴ the *Enrycher* system for semantic enrichment of textual data,¹⁵ the *SearchPoint* system for visual and contextualized Web browsing.¹⁶ The department members are also active in the area of women in science participating in several EU projects; the activities are presented at the Web portal ScienceWithArt.¹⁷ The **Artificial Intelligence Laboratory** closely collaborates with **Centre for Knowledge Transfer in Information Technologies** which has approximately ten researchers and technical staff working in the areas of research results dissemination and eLearning. In particular, the centre is well known by the portal VideoLectures.NET¹⁸ with multimedia materials of numerous scientific events, on-line training materials, and collection of tutorials on different scientific fields. The centre is covering management, training and dissemination activities of several EU projects.

Key personnel in X-LIKE

Marko Grobelnik (X-LIKE coordinator) is an expert in the areas of analysis and knowledge discovery in large complex databases. In particular, the areas of expertise comprise: Data Mining, Text Mining, Semantic Technologies, Network Analysis, and Complex Data Visualization. Marko collaborates with major European and US academic institutions and consults industries such as British Telecom, Microsoft Research, Nature, New York Times, Bloomberg, and Accenture. Marko is author of several books in the area of machine learning, data mining, text mining and semantic technologies and authors of many scientific papers. He is also W3C AC representative for IJS, CEO of the company Quintelligence and co-founder of the company Cycorp Europe. In the past years he organized series of workshops on text-mining TextKDD and network analysis LinkKDD at ACM KDD conferences. Marko served also as a program chair for European Machine Learning conference (ECMLPKDD 2009) and for Extended Semantic Web Conference (ESWC 2011). In terms of the past project experience, Marko has been technical coordinator for projects FP6 IST-World and FP7 VIDI; he was a member of project management board in several FP6 and FP7 Integrated Projects (SEKT, NEON, ACTIV, COIN) and STREP projects.

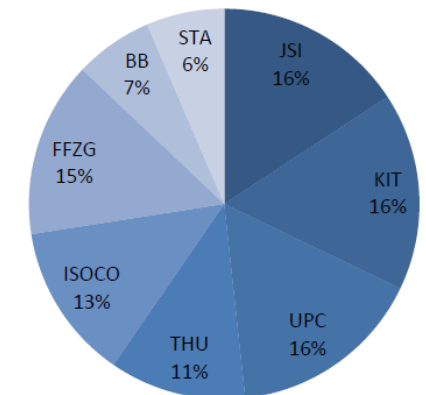
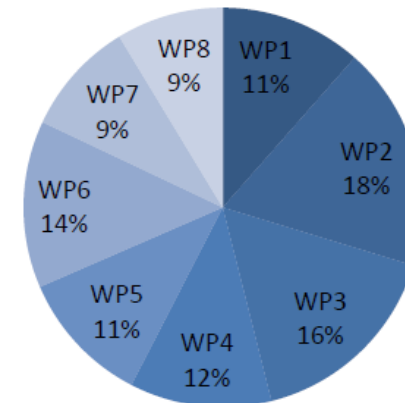
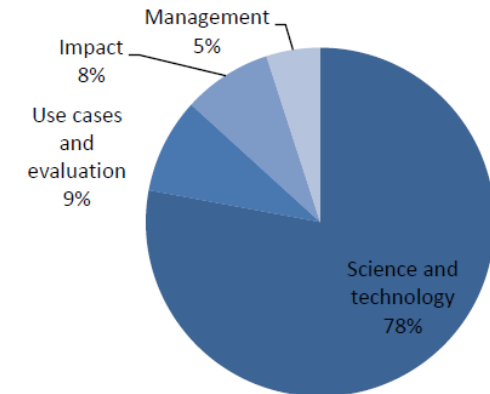
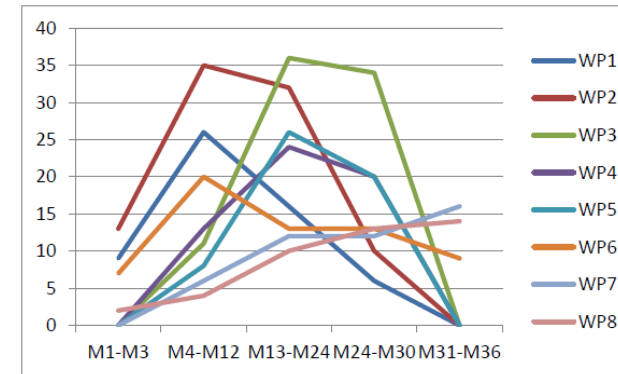
2.3 Consortium as a whole

- Sekcija opisuje na razne načine zakaj je konzorcij sestavljen ravno iz izbranih institucij
- Pomembno je, da poudarimo komplementarnost in dodeljene vloge posameznih institucij
- Cilj je, da evalvatorje prepričamo, da je ta sestava konzorcija dovolj dobra, da lahko uspešno izvede projekt

No.	Partner (short name)	Country	Expertise	Main contribution to the project
1.	JSI	SI	Information extraction	Cutting-edge mining and machine learning expertise
2.	KIT	DE	Semantic technologies	Cross-lingual annotation and semantic metadata management
3.	UPC	ES	Multilingual parsers	Deep parsing of multiple languages
4.	FFZG	HR	Machine translation	Shallow parsing and machine translation techniques towards learning annotations
5.	THU	CN	User interface	Experiences from news interpretation and presentation and application to Chinese data sources
6.	ISOCO	ES	Software engineering	Software services integration and commercial exploitation
7.	BB	UK	News and data services	Case study
8.	STA	SI	News agency	Case study

2.4 Resources to be committed

- Razdelek na razne načine opisuje kako so porazdeljena sredstva v projektu
- Tipični prikazi vključujejo:
 - Distribution of the budget and funding across activities
 - Distribution of the work across activity types
 - Distribution of the work across work packages
 - PM allocation over time
 - Budget distribution across partners
 - Effort distribution across partners
 - RTD effort across partners
 - Resources brought in by individual partners



Sekcija 3 dela „B“

Section 3. Impact

Title Page & Abstract

Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call

1.1 Concept and objectives

1.2 Progress beyond the state-of-the-art

1.3 S/T methodology and associated work plan

Section 2. Implementation

2.1 Management structure and procedures

2.2 Individual participants

2.3 Consortium as a whole

2.4 Resources to be committed

Section 3. Impact

3.1 Expected impacts listed in the work programme

3.2 Dissemination and/or exploitation of project results, and management of intellectual property

Section 4. Ethical Issues

Sekcija 3 dela „B“

- Sekcija opisuje:
 - Kakšen učinek bo imel projekt na okolje
 - Zakaj potrebujemo evropska sredstva (in ne nacionalna)
 - Kako bomo razširjali rezultate projekta
 - Kako bomo uredili intelektualno lastnino
- Sekcija vsebuje dve podsekciji:
 - **3.1 Expected impacts listed in the work programme**
 - **3.2 Dissemination and/or exploitation of project results, and management of intellectual property**

3.1 Expected impacts listed in the work programme

- Sekcija opisuje učinke, ki bi jih imel uspešen projekt na okolje.
- Nekateri vidiki, ki jih lahko uporabimo za opis učinkov:
 - EXPECTED IMPACT LISTED IN THE WORK PROGRAMME
 - SOCIETAL IMPACT
 - ECONOMIC IMPACT
 - SCIENTIFIC AND TECHNICAL IMPACT
 - RELATION TO OTHER INITIATIVES
 - NEED FOR EUROPEAN APPROACH

3.2 Dissemination and/or exploitation of project results, and management of intellectual property

- Sekcija opisuje naslednje rubrike:
 - Kanale po katerih bomo informirali javnost in zainteresirane stranke o rezultatih projekta
 - Kako bodo partnerji vgradili rezultate v svojo raziskovalno in/ali poslovno strategijo
 - Kako bo v projektu regulirana intelektualna lastnina

Target audience	Core dissemination channel	Description
Researchers and academia	Scientific events and publications Education and training	Publications at scientific events and in scientific journals, organization of and contribution to tutorials, workshops, summer schools
Industry and customers	Industry outreach Education and training	Community Advisory Group, customer workshops, support of and participation in industry conferences
Interested public	Web and public promotion material	Project Web site

Sekcija 4 dela „B“

Section 4. Ethical Issues

Title Page & Abstract

Section 1: Scientific and/or technical quality, relevant to the topics addressed by the call

1.1 Concept and objectives

1.2 Progress beyond the state-of-the-art

1.3 S/T methodology and associated work plan

Section 2. Implementation

2.1 Management structure and procedures

2.2 Individual participants

2.3 Consortium as a whole

2.4 Resources to be committed

Section 3. Impact

3.1 Expected impacts listed in the work programme

3.2 Dissemination and/or exploitation of project results, and management of intellectual property

Section 4. Ethical Issues

Sekcija 4 dela „B“

- Sekcija je pomembna le v primeru, da predlog projekta vsebuje etično problematične elemente
 -v tem primeru je potrebno natančno opredeliti problematične elemente
- Opis etično spornih elementov:
 - *Describe any ethical issues that may arise in their proposal. In particular, you should explain the benefit and burden of their experiments and the effects it may have on the research subject. Identify the countries where research will be undertaken and which ethical committees and regulatory organisations will need to be approached during the life of the project.*
 - *The following special issues should be taken into account:*
 - **Informed consent:** *When describing issues relating to informed consent, it will be necessary to illustrate an appropriate level of ethical sensitivity, and consider issues of insurance, incidental findings and the consequences of leaving the study.*
 - **Data protection issues:** *Avoid the unnecessary collection and use of personal data.*
 - **Use of animals**
 - **Human embryonic stem cells**

ETHICAL ISSUES TABLE

- Tabela z etičnimi vprašanji je potrebno nujno izpolniti in dodati v predlog projekta
 - ...tudi v primeru, ko projekt nima etično problematičnih elementov

	YES	Page Number
Informed Consent		
• Does the proposal involve children?		
• Does the proposal involve patients?		
• Does the proposal involve persons not able to give consent?		
• Does the proposal involve adult healthy volunteers?		
Biological research		
• Does the proposal involve human genetic material?		
• Does the proposal involve human biological samples?		
• Does the proposal involve human biological data collection?		
• Does the proposal involve human embryos?		
• Does the proposal involve human foetal tissue or cells?		
• Does the proposal involve human embryonic stem cells?		
Privacy		
• Does the proposal involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?		
• Does the proposal involve tracking the location or observation of people without their knowledge?		
Research on Animals		
• Does the proposal involve research on animals?		
• Are those animals transgenic small laboratory animals?		
• Are those animals transgenic farm animals?		
• Are those animals cloned farm animals?		
• Are those animals non-human primates?		
Research Involving Third Countries		
• Is any part of the research carried out in countries outside of the European Union and FP7 Associated states?		
Dual Use		
• Does the research have direct military application		
• Does the research have the potential for terrorist abuse		
ICT Implants		
• Does the proposal involve clinical trials of ICT implants?		
(IF NONE) I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL		

Praksa pri pisanju projektne
predloga

Uravnoveženost predloga projekta

- Pri pisanju moramo paziti na naslednje uravnoveženosti predloga projekta:
 - **Finančno**
 - ...tipično imamo dva nivoja partnerjev, nihče ne sme izstopati navzdol in navzgor
 - **Nacionalna zastopanost**
 - ...če se le da, naj bo zastopana cela Evropa, vsaj dve od velikih držav, vsaj en partner iz NMS (“new member states”)
 - **Akademsko/komercialno**
 - ...projekt naj prikaže simbiozo med akademsko in komercialno platjo v smislu vlog, denarja, sodelovanja
 - **Vloge na projektu**
 - Projekti partnerji naj imajo jasno opredeljene vloge (jedro projekta, ki povezuje in podporni partnerji s točno opredeljenimi vlogami)
 - **Velika in mala podjetja**
 - Skoraj vsi sprejeti projekti vsebujejo vsaj eno veliko podjetje in vsaj en SME
 - **Standardi in inovacije**
 - Projekt naj ima kontakt s standardizacijskimi institucijami in naj nosi nekaj (vsaj potencialnih) inovacij (z akademskim in komercialnim potencialom)

Glavni razlogi za neuspeh predlogov projektov

- sestava konzorcija 76%
- nerelevantnost 59%
(EU, eksploatacija, razširjanje)
- neprepričljivost pristopa 32%
- neprepričljivost inovativnosti 29%
- premalo informacije 21%
- neprimerno vodenje 20%
- izven področja klica 10%
- previsoki stroški 10%

Pogosti problemi (1/3)

- Pomanjkanje vizije
 - vizija je samo “hočemo projekt” ali “denar iz EU”
- Ideja projekta nima potenciala
 - samo prijavitelju se zdi zanimiva
- Akademiki bi se šli samo bazično znanost in smatrajo komercialne partnerje kot nebodigatreba
 - ...ali obratno
- Partnerji na projektu so prijatelji namesto partnerji
 - ...če ga ne vzameš je užaljen, če ga vzameš je projekt slab

Pogosti problemi (2/3)

- Koordinacija priprave projekta je preveč anarhična
 - vsak lahko izsili svoje, vodja nima avtoritete ali znanja
- Pozabljanje na malenkosti:
 - zastopanost obeh spolov, zastopanost SMEjev, prispevek k EU, ...
- Ignoriranje kriterijev za evaluacijo predlogov
- Čakanje s pripravo predloga do zadnjega trenutka
 - ...priprava projekta postane panično lepljenje kosov besedila

Pogosti problemi (3/3)

zamujanje pri pošiljanju

Dear partners,
after busy weeks working on the XXX proposal and with some of you in parallel on the YYY proposal I have to admit that I have underestimated the work and organisational efforts.

At the end we missed the deadline only by some hours after working also the last night very hard without stop.

I take the responsibility for the bad situation.

Many thanks to you all for your engagement especially ...

We have become a good team and I hope this will enable us to use the proposal for the next call ...

Povzetek

- Predlog projekta je dokument, ki ima vnaprej podano strukturo
 - Rubrike iz strukture izpolnimo na najbolj informativen način, da prepričamo evalvatorje, da predlagana ideja zasluži financiranje
 - Pomembno si je zapomniti, da predlog projekta pišemo za evalvatorje
 - ...če je predlog projekta sprejet v financiranje, potem dokument v naslednji fazi (skupaj z EC uradniki) pretvorimo v pogodbo, ki šele postane obvezujoča