QT LAUNCH PAD

MultilingualWeb Workshop 2013

META-NET

Quality Translation:

Addressing the Next Barrier to Multilingual Communication on the Internet

Hans Uszkoreit
DFKI

A Concern: Language boundaries

☆ commerce

☆ technology

☆ education

☆ health

☆ social change

A Necessity: The Need for MT

META NET

☆ for the single digital market

☆ for the information and knowledge society

☆ for horizontal plus vertical mobility

☆ for e-participation and e-democracy

An Observation:
Concentration of Research on Gist Translation

META-NET

- ☆ Most of MT research has concentrated on in-bound content overview translation
- ☆ Reasons: Funding sources, new applications and opportunities for fast success
- ☆ Nearly all existing translation markets are for out-bound translation
- ☆ There is a lack of systematic research on quality obstacles and on shared quality metrics for MT and HT

A New Analytical Approach

good enough

almost good enough
requires just a few
editing steps

not usable for
out-bound purposes

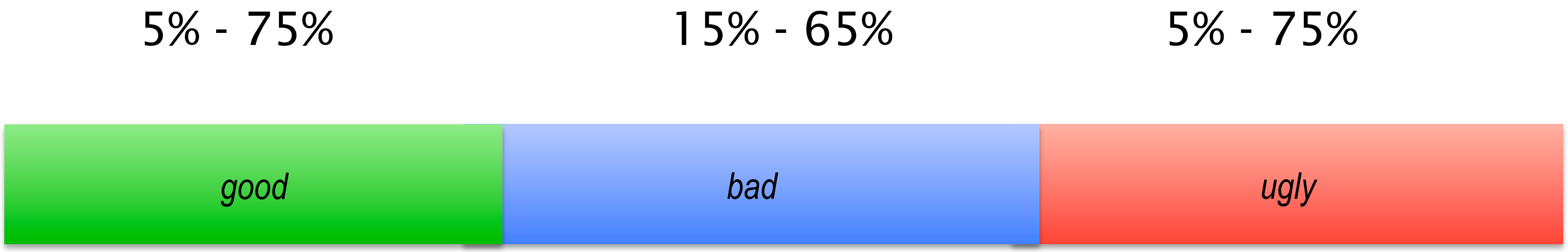
A New Analytical Approach

good

bad

ugly

A New Analytical Approach



5% - 75%

15% - 65%

5% - 75%

good

bad

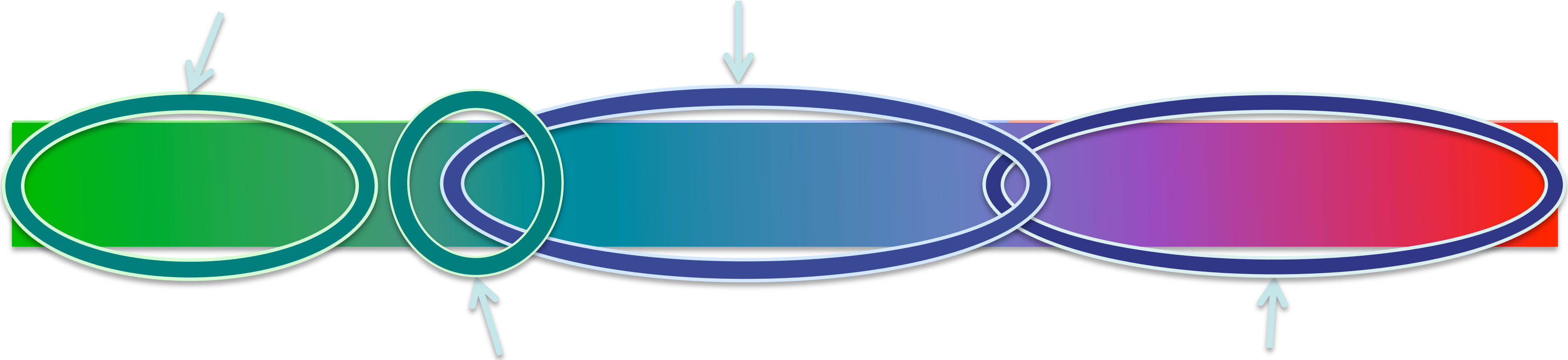
ugly

A New Analytical Approach

A New Analytical Approach

Recognize the truly *good*
High Quality Estimation

help the translator to
improve the *bad*:
Computer Assisted Translation



Make the nearly *good*
truly *good*:
High Quality MT

help the end user to make
the *ugly* at least understandable:
Improve Gist Translation

META-NET

☆ META-NET - Network of Excellence 60 research centers in 34 countries

☆ META, an alliance with 638 members (organizations) in 47 countries

☆ Vision Process with three vision groups

☆ 31 Language White Papers on 31 individual languages

☆ A first version of META-SHARE, the infrastructure for sharing resources

☆ A Strategic Research Agenda for Multilingual Europe

☆ Inclusion of language communities - language policy bodies

☆ Inclusion of industrial and professional associations

The Strategic Research Agenda

- a vision with a plan
- with more than 200 contributors
- drafts were discussed at 83 conferences, workshops and other meetings
- the prefinal draft was distributed, reviewed and revised in the Fall of 2012
- In the last discussion round of the prefinal draft, we again received und used more than 50 proposed pieces of textual input
- the SRA was finalised in December 2012
- it will be publicly presented next week

META-NET

STRATEGIC
RESEARCH
AGENDA FOR
MULTILINGUAL
EUROPE 2020

presented by the
META Technology Council

Springer

SRA: Contents – Brief Glimpse

- Set the stage and describe the European situation, the needs and the LT research and industry.
- Discuss the state of IT, predictions and mega-trends.
- Our technology vision for 2020.
- Select and specify priority themes.
- Suggest a model for speeding up innovation.
- Outline proposals for the organisation of research and innovation.

META-NET

STRATEGIC
RESEARCH
AGENDA FOR
MULTILINGUAL
EUROPE 2020

edited by the
META Technology Council

Strategic Considerations

We decided on priority themes that

- …support technology progress
- …lead to solutions that European society needs
- …to solutions from which European industry will benefit as users or as providers

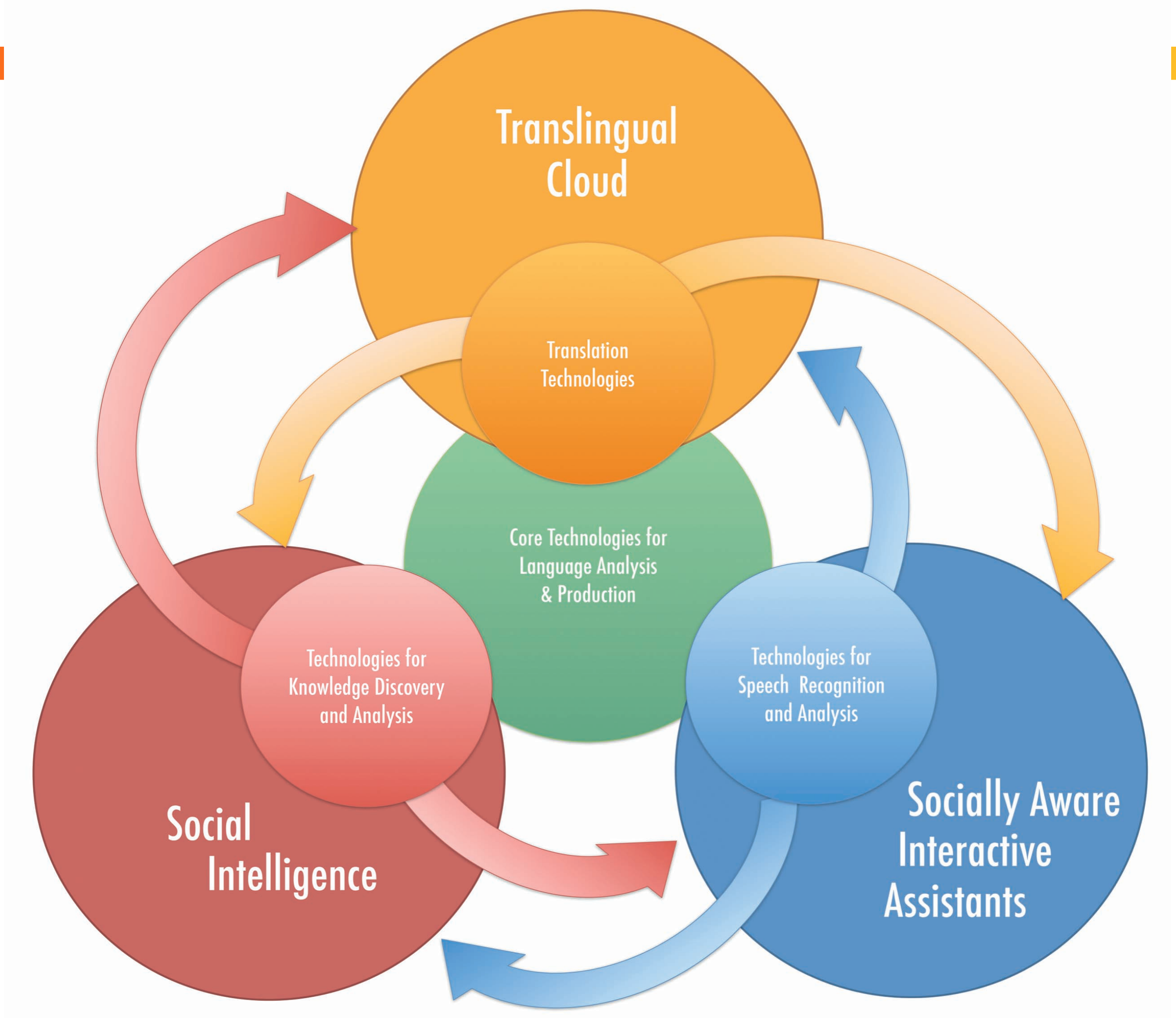
SRA-Priority Themes

- Translingual Cloud –
  Understanding everything,
  everywhere, everytime

- Social Intelligence –
  Technologies for e-participation
  and better decisions

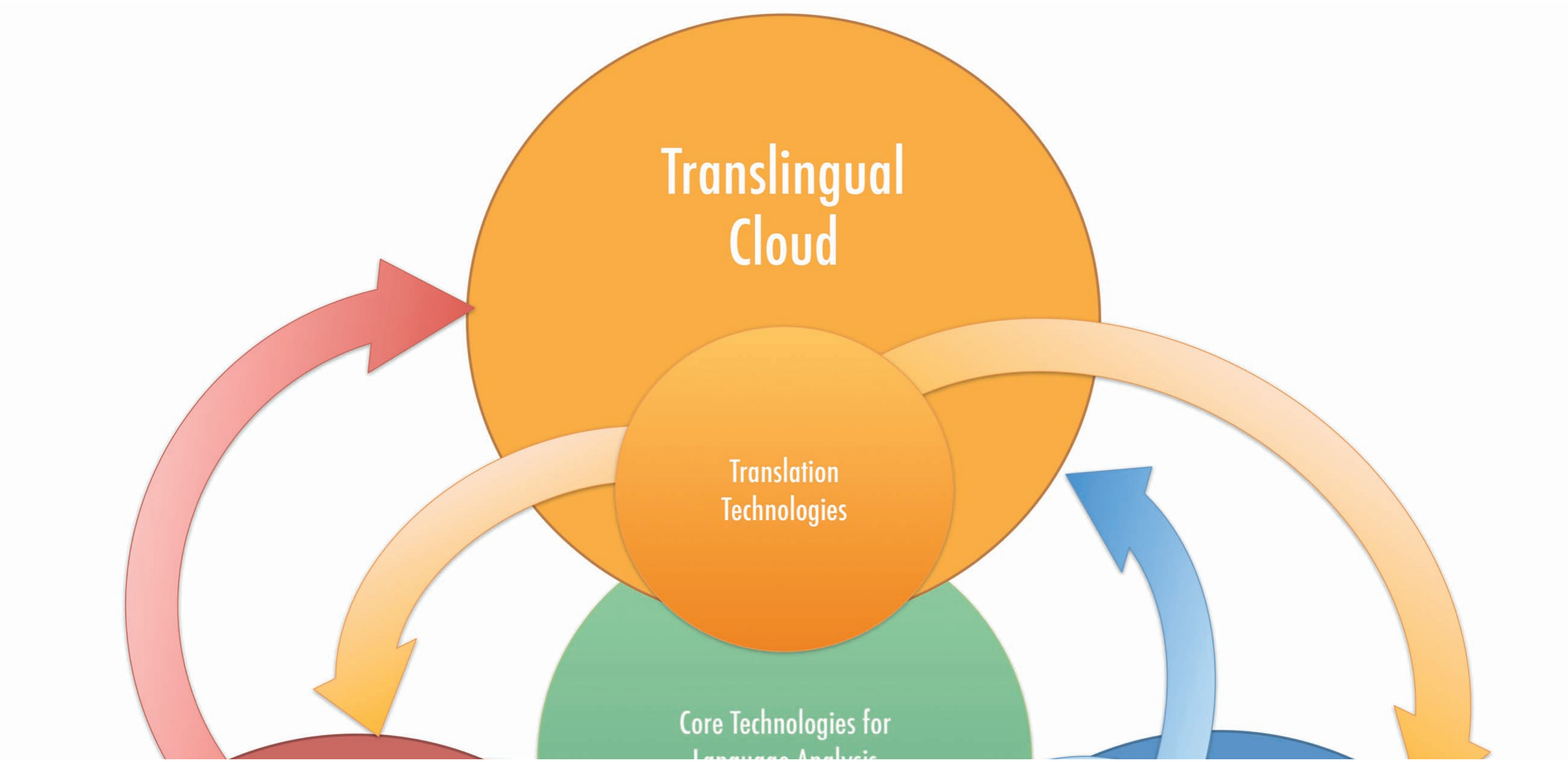
- Socially aware interactive assistants

SRA-Priority Themes



Translingual Cloud
Translation Technologies
Core Technologies for Language Analysis & Production
Technologies for Knowledge Discovery and Analysis
Technologies for Speech Recognition and Analysis
Social Intelligence
Socially Aware Interactive Assistants

Translingual Cloud



What is it?

- A network of generic and special-purpose services combining automatic translation, language checking, post-editing, as well as human creativity and quality assurance where needed.
  - Free for small volume use and for high-volume base-line quality
  - Business opportunities for a wide range of service and technology providers

Ingredients

- Systematic concentration on quality barriers
  - A unified dynamic-depth weighted-multidimensional quality assessment model with task profiling
  - Significantly improved automatic quality estimation
- Inclusion of translation professionals and enterprises in the entire research and innovation process
  - Ergonomic work environments for computer-supported creative top-quality human translation and multilingual text authoring

More Ingredients

- A Semantic translation paradigm by extending statistical translation by semantic data such as linked open data, ontologies including semantic models of processes and textual inference models
- Exploitation of strong monolingual NL analysis and generation methods and resources
- Modular combinations of specialized analysis, generation and transfer models, permitting accommodation of registers and styles and also enabling translation within a language (e.g. between specialists and laypersons).

European Service Platform

- creation of an ambitious large-scale sky-computing platform

  European Multilingual Service Platform for Language Understanding, Communication to be extended to Knowledge, Inference, Emotion…

- central motor for research and innovation in the next phase of IT evolution

- ubiquitous resource for the multilingual European society (an idea suggested by several experts from industry in META-NET Vision Group meetings).

- The platform will be used for testing, show casing, proof-of-concept demonstration, avant-garde adoption, experimental and operational service composition, and fast and economical service delivery to enterprises and end-users.

The Proposed Platform is …

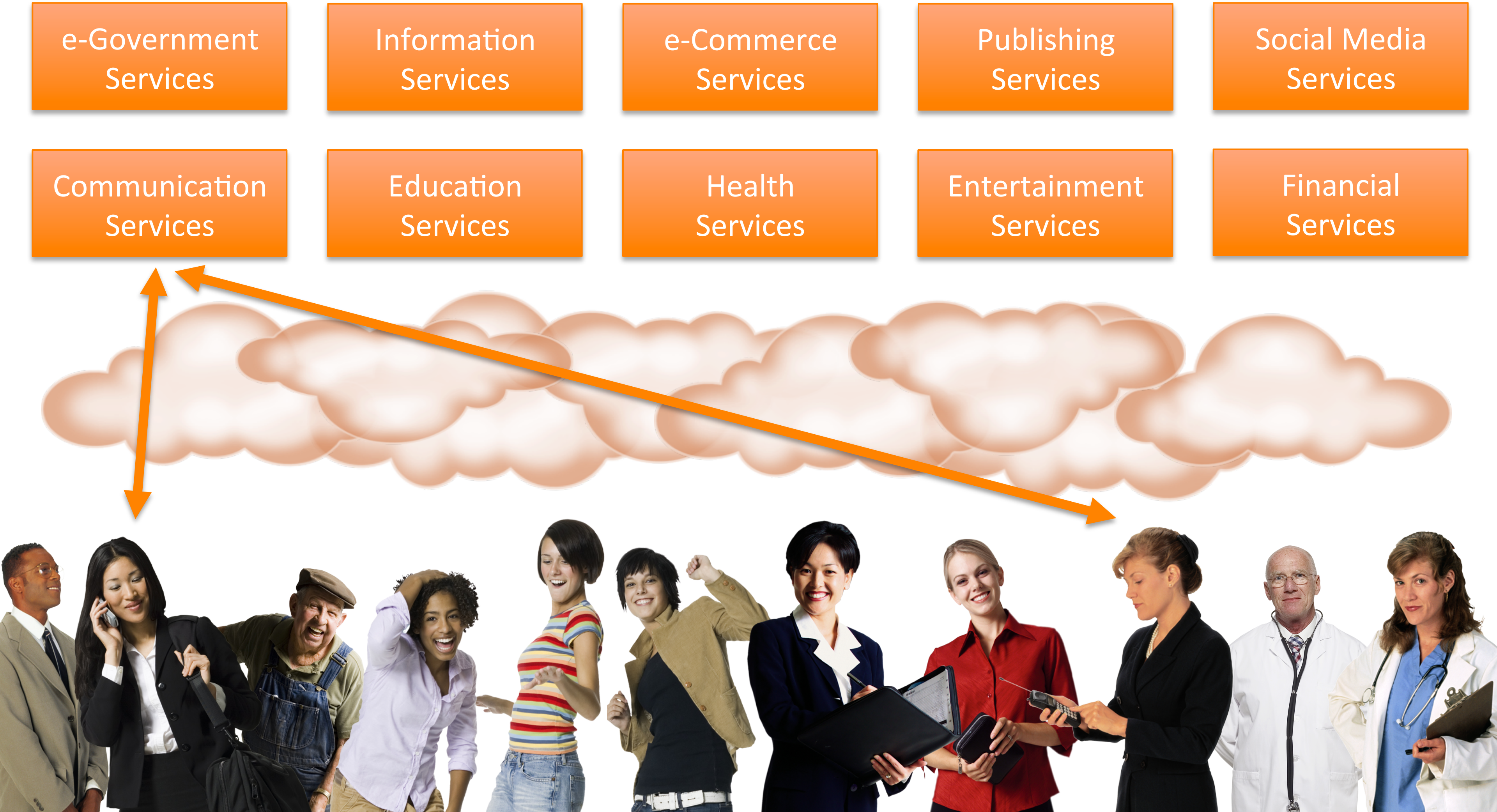
META-NET

- intended for a mix of commercial and non-commercial services.

- It would be cost-free for all providers of non-commercial services (cost-free and advertisement-free) including research systems, experimental services and freely shared resources but it would raise revenues by charging a proportional commission on all commercially provided services.

- In order to reduce dependence on individual companies and software products, the base technology should be supplied by open toolkits and standards such as OpenNebula and OCCI.

Translation Integrated in Many Services



e-Government Services

Information Services

e-Commerce Services

Publishing Services

Social Media Services

Communication Services

Education Services

Health Services

Entertainment Services

Financial Services

Translation Integrated in Many Services

META-NET

e-Government Services

Information Services

e-Commerce Services

Publishing Services

Social Media Services

Communication Services

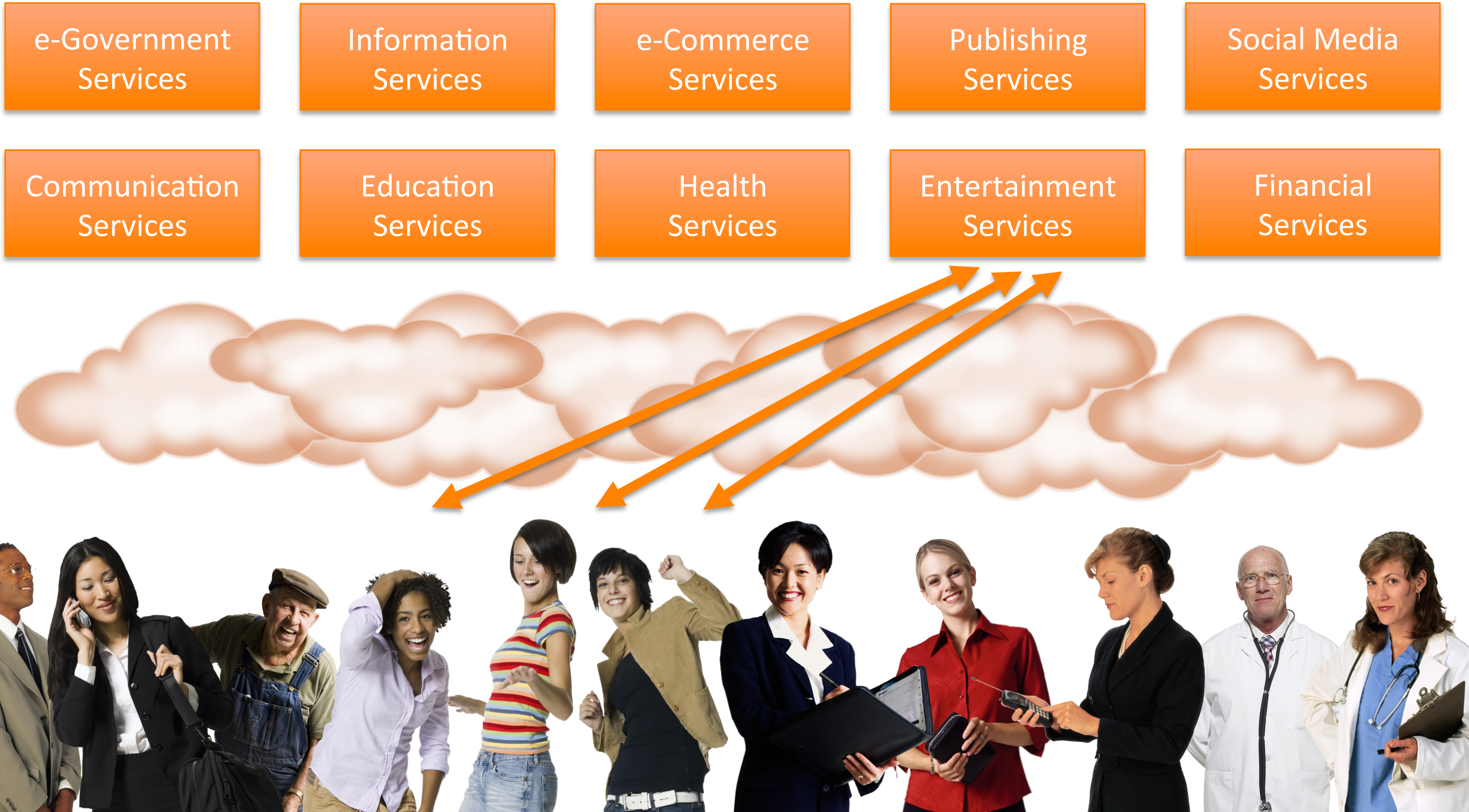
Education Services

Health Services

Entertainment Services

Financial Services

Translation Integrated in Many Services



e-Government Services

Information Services

e-Commerce Services

Publishing Services

Social Media Services

Communication Services

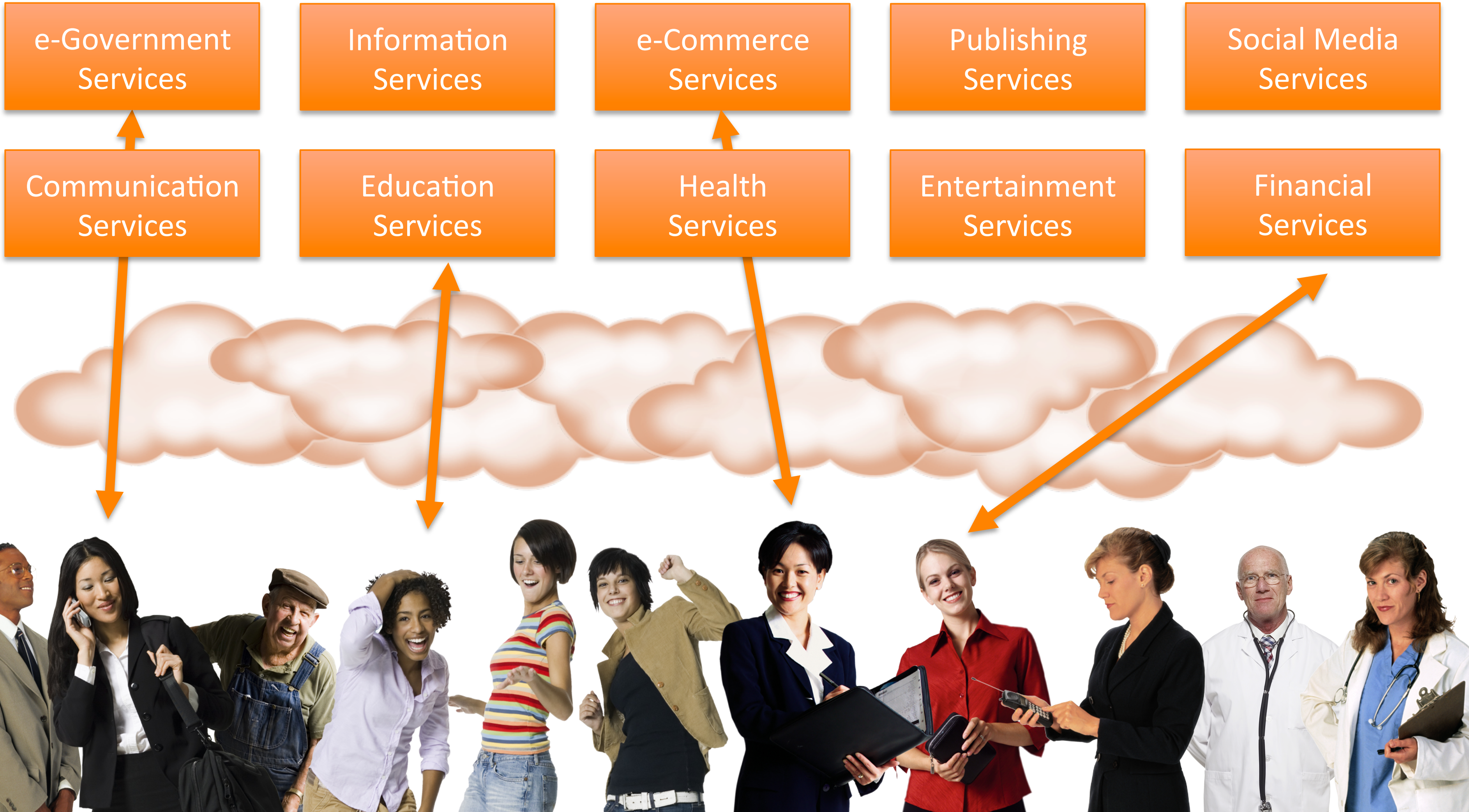
Education Services

Health Services

Entertainment Services

Financial Services

Translation Integrated in Many Services



e-Government Services

Information Services

e-Commerce Services

Publishing Services

Social Media Services

Communication Services

Education Services

Health Services

Entertainment Services

Financial Services

QTLaunchPad – Objectives

- ☆ The support action will prepare the grounds for a new type of collaborative MT research dedicated to overcoming existing quality barriers.
- ☆ To this end, QTLaunchPad will
  - – assemble and provide needed data and tools including specialised translation corpora, test suites and tools for quality assessment,
  - – create a shared quality metrics for human and machine translation, improve automatic translation quality estimation,
  - – extend an existing platform for resource-sharing to the needs of quality-MT research,
  - – define strategies and challenges and then plan and launch a large-scale research and innovation action ("QT21") for a breakthrough in quality translation technology.

QTLaunchPad – Objectives

- ☆ The support action will prepare the grounds for a new type of collaborative MT research dedicated to overcoming existing quality barriers.
- ☆ To this end, QTLaunchPad will
  - – assemble and provide needed data and tools including specialised translation corpora, test suites and tools for quality assessment,
  - – create a shared quality metrics for human and machine translation, improve automatic translation quality estimation,
  - – extend an existing platform for resource-sharing to the needs of quality-MT research,
  - – define strategies and challenges and then plan and launch a large-scale research and innovation action (“QT21”) for a breakthrough in quality translation technology.

QTLP Consortium

☆ DFKI

☆ CNGL DCU

☆ ILSP Athena

☆ U. Sheffield

☆ Subcontractor GALA

QTLP Planning Panel

QT LAUNCH PAD

- Jan Hajic of U. Prague,
- Stephan Oepen of U. Oslo,
- Philipp Koehn of U. Edinburgh,
- Alex Waibel of Karlsruhe KIT,
- Marcello Federico of FBK Trento,
- Mikel Forcada of U. Alicante,
- Hermann Ney and Volker Steinbiss of RWTH Aachen,
- Nuria Bel of U. Barcelona,
- Joseph Mariani of LIMSI Paris,
- Johann Roturier of Symantec Ireland,
- Spyridon Pilos of EC DGT Luxembourg,
- Serge Gladkoff, Kim Harris and Hans Fenstermacher of GALA.
- Andrejs Vasiljevs, Tilde

Inportant Trend

☆ On the way to quality translation, MT will increasingly employ:

- morphological processing
- syntactic processing
- semantic processing

Semantics-Based Translation

☆ Warren Weaver (1949): “Thus it may be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication – the real but as yet undiscovered universal language – and then re-emerge by whatever particular route is convenient.”

☆ Kevin Knight (2012): “As long as we get the **“who did what to who”** wrong, we are optimizing with respect to the wrong metric (BLEU)!”

Where are the Semantic Resources?

They are growing fast:

☆ linked open data, FreeBase, DBPedia, YAGO

☆ WordNet, and the VerbNet, FrameNet, BabelNet, UWN, …

☆ PropBank, OntoNotes

Prediction

- ☆ A lengthy massive bootstrapping process involving at least the following communities:
  - – MT: structure-based approaches to MT
  - – Lexical semantics: WSD, VerbNet, FrameNet, WordNet, UWN/MENTA
  - – Knowledge: SW, linked (open) data, KDBs (Yago, Freebase, Dbpedia)
  - – IE, especially Relation/Event Extraction and evolving gold standards sets
  - – Textual Inference: RTE, paraphrasing
  - – NL Generation

Another dimension of the bootstrapping

- ☆ Two main approaches to useful applications on the way:
  - big powerful robust systems
  - Much more restricted precision-centered systems for special purposes

Big (noisy) data vs. a truly Semantic Web with inference and plausbility filters

Back to the Multlingual Web

- ☆ The Web is becoming multilingual, but eventually it will have to be translingual
- ☆ The Web is THE medium for sharing, storing, accessing und using knowledge and information
- ☆ This mission includes transforming knowledge and information to the information needs – this is the core argument of the SW and at the same time the core argument for the translingual web
- ☆ Should translation ever have to go through semantics, how could it bypass the SW?

META