Introduction to Localization

Localization World
Dublin 2014

Uwe Stahlschmidt
Daniel Goldschmidt

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Agenda

- Introduction
- The problem
- Definitions
- Project / process of localization
- Some words about Mobile
- Tools
- Standards, XLIFF
- QA/testing
- Vendor and project management

And maybe:

- Crowdsourcing
- Localization in Agile Development

Thanks

Thanks to Angelika Zerfass, Iris Orriss, Richard Sikes and Bodo Vahldieck for some of slides.

Agenda

Localization, internationalization, Globalization, translation, regionalization... too many "...ation" terms...

During the next sessions we will make sense of them for you.

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Welcome!

Tell us:

- Your name
- Where you are coming from
- Which company are you working with
- Your expectations from the workshop
- One thing that you did (unrelated to work / l10n etc.) which no one would believe you actually did

The problem

The Problem

- A known company developed a powerful product for CRM (Customer Relationship Management System)
- The first and main market was, as usual, the USA
- The board decided that it is time to penetrate new markets: Europe, Far-East, Middle East

    The R&D department claimed – no problem, we are fully UNICODE…let's go!

The Problem

Ouch…

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

The Problem
#1 – String Externalization

- All the GUI (graphical user interface) had to be translated to the target languages
- But lots of strings were hard-coded (written directly into the code)

The Problem
#2 - Sorting

- After translating the GUI, the first installation took place in Spain
- Some customers were unhappy: Many indexes and lexical orders were corrupted
- In Traditional Spanish, the letters "*CH* " and "LL" have their own positions in the sort order
- *A, B, C, CH, D...K, L, LL, M, ... etc.*
  - Curioso
  - Chalina
  - Luz
  - Llama

The Problem

The second installation in Germany had three problems:
- The search function didn't work
- The financial and numerical functions were buggy
- Many strings were cutoff in the GUI

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

The Problem
#3 –Collation

- Combining characters:
  Ü (Latin Small letter U with diaeresis 0x00DC)
  U¨ (Latin Small letter U 0x0055, Combining diaeresis 0x0308)

  ç (Latin Small letter with Cedilla 0x00E7)
  c¸ (Latin Small letter C 0x0063, Combining Cedilla 0x0327)

- fi=fi
- Case sensitive/insensitive
- Accent sensitive/insensitive
- Upper case ß (Latin Small letter Sharp S)= SS

The Problem
#4 – Numerical format

- 4.500 (UK) ≠ 4.500 (DE)
- 4,500 (UK) = 4.500 (DE)
- 4.500 (UK) = 4,500 (DE)

The Problem
#5 - Length

- German strings are usually longer than in most languages
- English: *Redo* German: *Wiederherstellen*
- English: *Skip* German: *Auslassen*
- English: *Exit* German: *Beenden*
- English: *Edit* German: *Bearbeiten*

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

The Problem
#6 – Date Format

- The client from Spain called after 2 months; the license had expired earlier then expected!

  Does 01/07/2012 mean:
  "July, first 2012"
  Or
  "January, seventh 2012"?

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

The Problem
#6 – Date Format, Calendars

- The first day of the week is Monday... or Sunday (weekend)
- Year length
- Week numbers (ISO? Other?)
- Last Monday

The Problem
#7 - Encoding

The installation in Russia was catastrophic:

- All imported data from the legacy systems was full of question marks.
- All data inserted by the user couldn't be retrieved from the database

- This was the first installation using a non "Western European" encoding!
- Cyrillic text is encoded using KOI8-U and other encodings.

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

The Problem
#8 – Look and Feel

The Chinese customer was confused:

- Why positive values were colored in green?

- Why negative values were colored in red?

- http://money.msn.com.cn/

The Problem
#8 – Look and Feel

	RED	Yellow	Green	White	Blue	Black
West	Danger Anger Stop	Caution Cowardice	Safe Sour Go	Purity Virtue	Masculinity Calm Authority	Death Evil
China	Joy Festivity	Honor Royalty	Youth Growth	Humility Mourning	Strength Power	Evil
Japan	Danger Anger	Nobility Childish Gaiety	Youth Future Energy	Death Mourning	Villainy	Evil
Middle East	Danger Evil	Happiness Prosperity	Fertility Strength	Purity Mourning		Mystery Evil

The Problem
#9 – Politics

The Hebrew website had some minor issues:

When localizing a website for Israel, which map shall we use:

- The one with Judea and Samaria
- The one with the Palestinian Authority
- The one without the occupied territories

"Judea and Samaria" vs. "occupied territories"

Map of Golan Heights

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

The Problem
#10 – Grammar

- Singular? Plural?
- Male, female, something else?

- How to translate concatenated strings?

The Problem
#10 – Grammar

String concatenation example:

The Winfax Installer has found %s.

- Case
  - Microsoft
    - S="Outlook"
  - Netscape
    - S="Netscape Mail"
  - Notes
    - S="Notes Email"
  - Else
    - that you have no email provider.

The Problem
#11 – Context dependent shapes

- Greek – Lower-case Sigma
- Hebrew: Kaf, Mem, Nun, Pe, Tsadi
- Arabic: most letters have up to four shapes
- Indic, Thai – even worse
- הלך/הלכה

Rendering? Search?

The Problem
#12 – Diacritics

Hebrew, Arabic

Rendering?

Search?

דניאל/דָּנִיאֵל

The Problem
#13 – Graphics & Symbols

The OK gesture:
- English-speaking: OK
- France: zero, nothing, worthless
- Mediterranean: a rude sign
- Japan: money
- Brazil & Germany: vulgar, obscene gesture

The Problem
more issues

- Time zone
- Paper sizes (A4 vs. Letter)
- Phone numbers
- Address format
- Temperature
- Measurements

Culture is Everywhere

"If I'm selling to you, I speak your language.

If I'm buying, dann müssen Sie Deutsch sprechen (then you must speak German)"

Willy Brandt

Definition

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Terms

Globalization

- Adaptation of product strategies to regional requirements of all kinds.

Internationalization

- Engineering of a product to enable efficient adaptation of that product to local requirements.

Localization

- Localization is the process of adapting a (software) product and accompanying materials to suit a target-market locale.

Terms

Locale

- A locale is a geographic region defined by a combination of language and cultural norms. "Locale" is not to be confused with "language." For example fr-FR, fr-CA, fr-CH.

CLDR definition (http://www.unicode.org/reports/tr35/#Locale)

- Set of user preferences that tend to be shared across regions of the world.
- Dates, times, numbers, and currencies; for measurement units, for sort-order (collation), plus translated names for time zones, languages, countries, and scripts.

Costs that are generated in one place become visible in another.
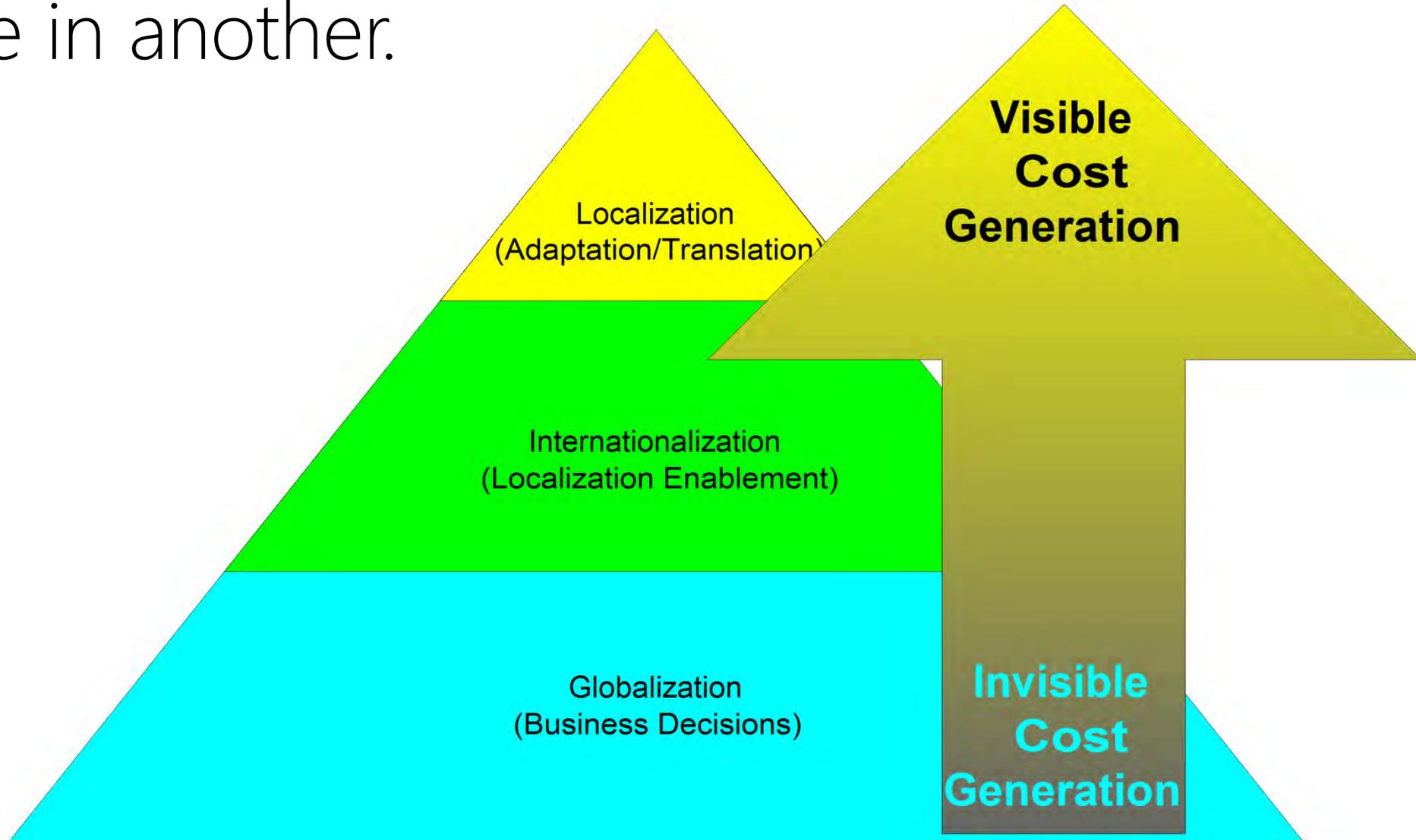
Localization

- Success

  Product appears to be produced for the target market
- Failure:

  We can easily notice that the product was adapted

The 3 layers approach

- String Externalization
- Collation (Luz Vs. Llama)
- Normalization (Ü Vs. U¨ , ç Vs. c¸)
- Numerical Format (10.5 Vs.10,5)
- String Length (Redo Vs. Wiederherstellen)
- Date / Time format (01/07/2006)
- Calendars
- The first day of the week (Monday? Sunday?)
- Year length
- Week numbers (ISO? Other?)

The 3 layers approach

- Encoding (Unicode, S-JIS, BIG5, Latin-X etc.)
- Shared messages
- Bi-Directional scripts
- Fonts (Unicode, Vertical vs. Horizontal)
- Color schemes (http://finance.cn.yahoo.com/
- Segmentation (the Japanese example – no white spaces!)
- Alphabets (Traditional vs. Simplified)
- Morphology (how to perform search in Hebrew, Russian?)
- Decompounding (how to perform search in German?)

The 3 layers approach

- Politics (Judea and Samaria? Palestinian Authority? occupied territories? Which City is the capital of Israel?)
- Time zone (Midnight? Where? When?)
- Paper sizes (A4 Vs. Letter)
- Measurements (Inch, CMs)
- Cultural issue (the OK gesture)
- Casing (Turkish uppercase of the Latin lower case (i) is "İ")
- IDN: www.דניאל.com
- Functional Localization

The 3 layers approach

The i18n and l10n problem is a mixture of:

- Technical Issues
- Cultural Issues
- Political Issues
- Language / Linguistic Issues
- Esthetical Issues

The 3 layers approach

Layer 1 – Data Integrity *("handle with care" sticker)*

      moving data from A to B
      Usually not locale dependent

The 3 layers approach

Layer 2 – Application

    doing something with the data (e.g. sorting, searching casing, date/time format etc.)
usually locale dependent

The 3 layers approach

Layer 3 – Display

Presentation layer
Localization readiness (resources externalization)
(In Microsoft we call it Localizability)
usually locale dependent

LOCALIZATION WORLD
CONFERENCE & EXHIBITS
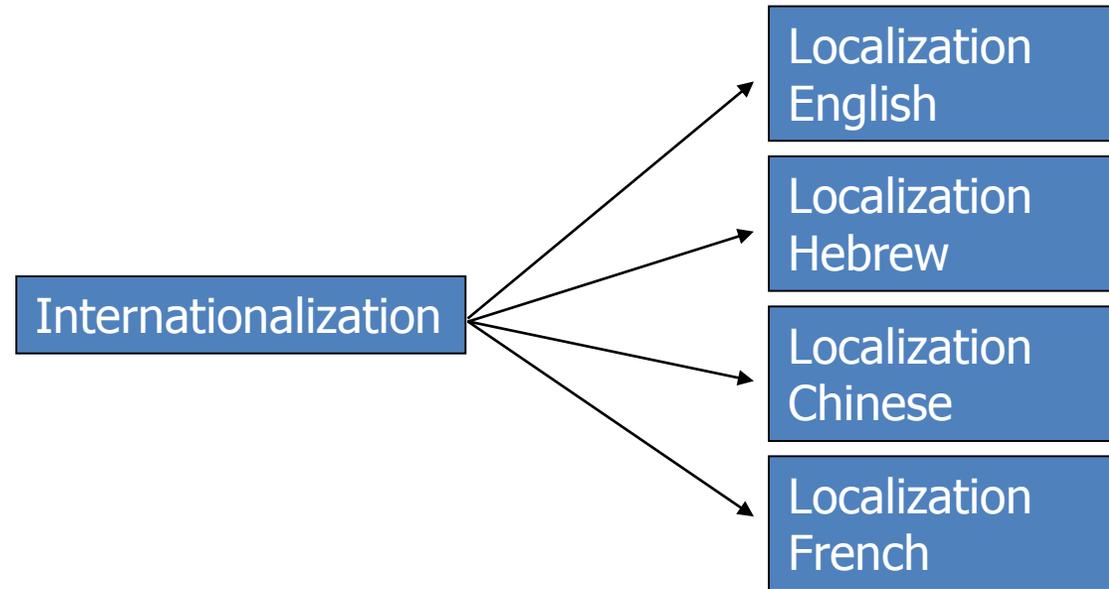
Localization vs. Internationalization

Localization vs. Internationalization

- You don't need to actually read and write 22 languages
- i18n is software engineering, not a linguistic process
  - There are cross-over concepts, however, such as:
    - Allowing sufficient white space for language growth in documentation
    - Not hard-coding page references in books
    - Planning website architecture to support multilingual content and navigation
- Project management is an essential part of Localization

Project / process of localization

Who's involved?

- Content providers (Editors, technical writers, R&D teams etc.)
- Localization project managers (on publisher side, on vendor side)
- Localization engineers (on publisher side or vendor side)
- Translators (In house, freelance, Single Language Vendor, sub contractors)
- Reviewers (In house, freelance, Single Language Vendor, sub contractors, regional office employees)
- Quality Assurance specialists (on publisher side, on vendor side)
- Finance personnel
- Program managers
- Product marketing managers
- Webmasters

Investment in preparation versus project execution

A short To-Do list

- Researching and gathering components to be localized
- Preparing the content (text segmentation, resource extraction etc.)
- Pseudo localization and proactive i18n QA on core code
- Leveraging against existing TMs
- Effort estimation, costing
- Management Approval
- Work assignment

- Translation and localization
- Proof reading / Editing / Reviewing
- Testing

- TM updates, maintenance of linguistic assets
- Delivery
- Billing

The Traditional Process

Content providers

Content providers

Content Repository

Preparing

Leveraging

Effort assessment

Linguistics assets:
TMs
Terms
Glossaries

Translating

Reviewing

Packaging and delivery

Updating Linguistics assets

Dublin, June 4th 2014
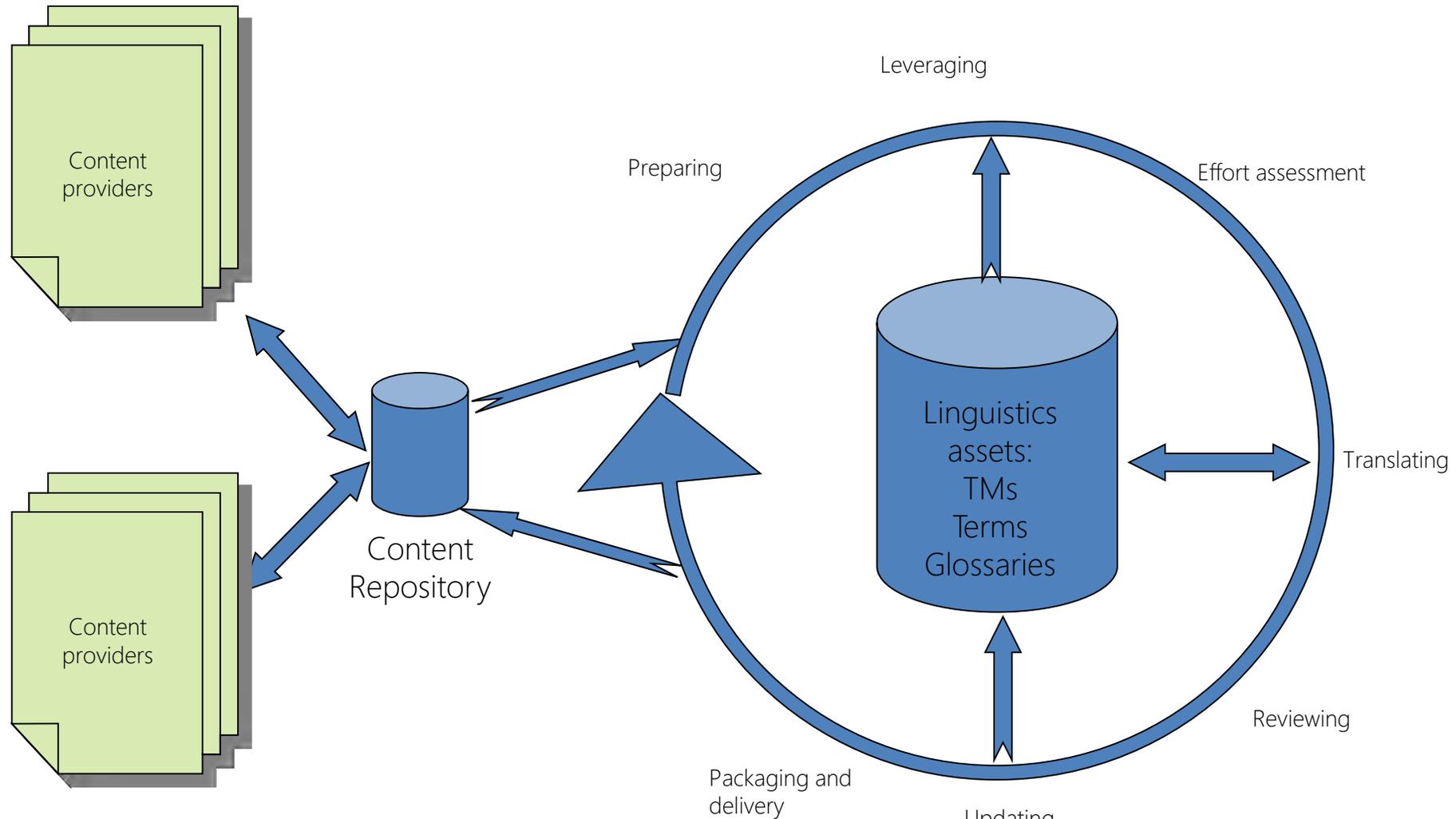
Preparation: The Vendor

- The vendor is your best friend!

- However, this friend sells words (for translation)!

Preparation: The Vendor

What to consider:

- Rate: xx-yy $cent/word
- Pace: 1500 words/day

Price should include:

- Translation
- Editing
- Proof reading

Not included:

- Project management
- QA cost
- DTP

Consider training the vendor's translators and the proof readers: it will give them insight into the product

Preparation: The Vendor

Be sure to establish the following:
- Processes
- Escalation process
- Location of translators
- Single focal point
- Localization material
- Deliverable
- TM ownership
- What are you paying for
- Bug fixing responsibility
- Service Level Agreement (SLA)
- How do we Communicate?
- Transferring data (Builds, accessing etc.)
- Usage of cloud
- Account / login

Preparation: The Vendor

Be sure to determine what you are paying for:

- Price per word
- Discount for repetitions
- Word counting in source language or target language?
- QA?
- Bug fixing?
- Penalties

Translation

Basic premises:

- Translation is expensive
  Example:
  - 1 million words = $250,000 per language
  - A 10 languages localization project easily could incur cost of $2.5M
- Glossaries are required
- Translation memories are required

Context in Translation

Translators need to know what to translate and what **not to translate**

(tags, code etc.) Expose only translatable content to them – don't run the risk of having your code broken

Translators need to know the context:

• Surrounding text, dialog etc.

• i.e. "display"

       German: anzeigen (to display)

       German: Anzeige (a display)

Project Wrap

- TM update
- Delivery
- Invoice Management
- Post-mortem

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

LOCALIZATION FOR MOBILE

Over 1.5B Windows/Phone/Xbox Customers

19%
USA

The Rest of the World

81%

Facts:

In 2015, the US will only account for roughly 25% of PC/tablet and 11% of smartphone shipments.

89% of 34 million phone install base is outside US.

Over 60% of our revenue comes from non-US sources.

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Dublin, June 4th 2014

Challenges for Mobile Localization
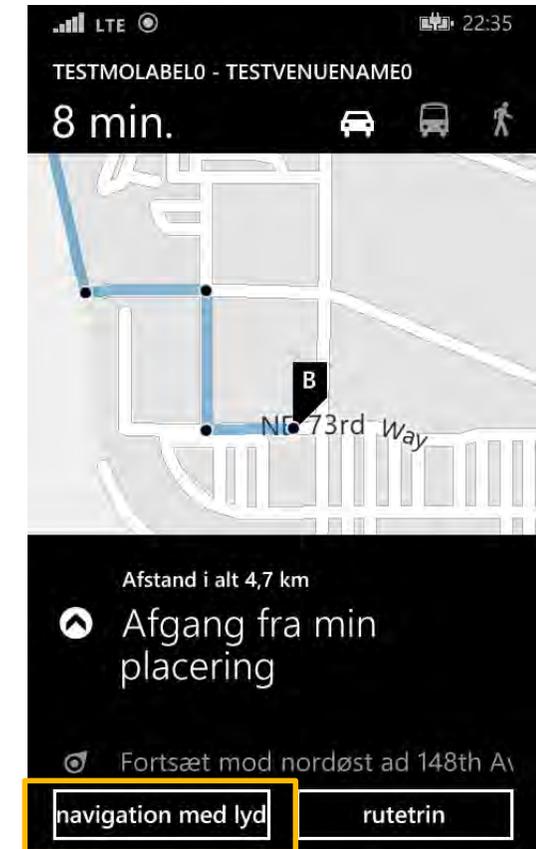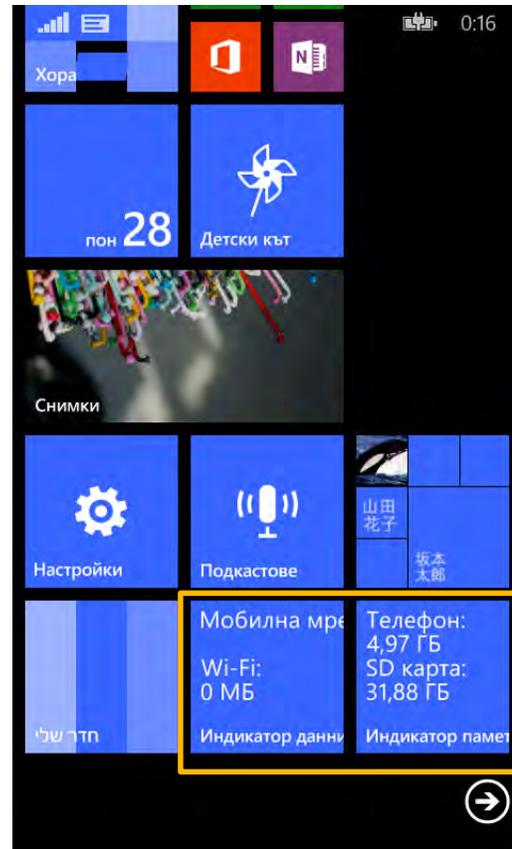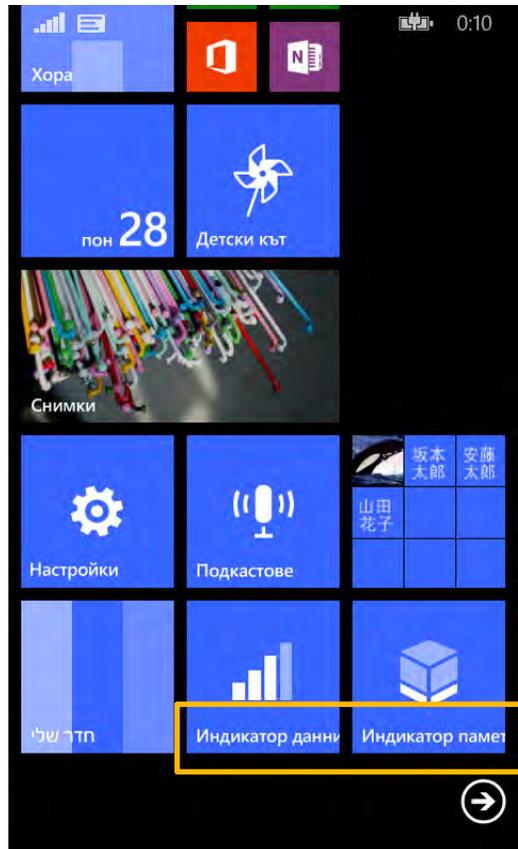
Consider

- ## Limitations of screen size
  - Some target languages will increase character and/or word count by as much as 35% over English
  - Languages like Chinese may require larger fonts to make characters readable

- ## Leave room for expansion
  - Allow for longer words/strings in design phase

- ## Font considerations
  - Special fonts, Italics, and/or bolding of text doesn't transfer well to non Latin scripts: don't use it

Also consider

- Context
  - Short strings in mobile lack context – this is often guesswork for the translator
  - Provide context info to translator!

- Terminology
  - Make sure translators know the domain/use mobile terminology
  - Example: Mobile Phone (US) – Handy (Germany)

- English Abbreviations
  - Don't use them

Tools

Angelika Zerfass wrote those slides

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Project Management / Workflow Management

Machine Translation Engine

Possible file preparation

Software Localization Tool

Self-developed tool

Source language files

Target language files

API

Terminology Extraction

Term base / term list

Translation Memory Tool

Alignment

Word count tool

QA tool

Macros

possible text extraction file conversion

Editor of TM tool

Creation of target language file DTP

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Software Localization Tool

- A tool to test the localizability of software
    - Pseudo Translation / simulated translation
- A tool to translate text in software applications
    - GUI (graphical user interface)
        - Menus, Dialogs
    - Error messages, system messages
- A tool to adapt the GUI to the translation
    - Resizing dialog boxes
    - Flipping contents of dialog boxes for right-to-left languages
    - Adaptation of icons, graphics

SDL Passolo 2011 Demo Version - [wordpad.lpu]

File Edit View Layout Project String List String Tools Window Help

Project German (Germany)

Resources

wordpad: Dialog 145 "Tabs"

- wordpad *
  - Menu
    - 4 "Edit"
    - 5 "File"
    - 6 "File"
    - 128 "File"
    - 142 "COLOR_POPUP"
    - 151 "MISC"
    - 2 "TEXT_POPUP"
    - "DEFAULTFORMATDIALC
    - "UNSUPPORTEDSAVEFOF
    - 143 "Paragraph"
    - 145 "Tabs"
    - 150 "Page Setup"
    - 151 "Date and Time"
    - 158 "Options"
    - 159
    - 160 "New"
  - String Table
  - Accelerator Table
  - Version

Navigation

Dialog view

Tabs

Tab stop position

OK

Cancel

e Inhalte lösch

Festlegen   Clear

Num...	ID	S..	English (United States)	German (Germany)
217	0		Tabs	Tabs
219	1022		&Tab stop position	&Tab stop position
221	1005	👁 f(	&Set	&Festlegen
222	1006		Cl&ear	Cl&ear
223	1		OK	OK
224	2		Cancel	Cancel
⚠ 225	1007	👁 f(	Clear A&ll	&Alle Inhalte löschen

Translation list

Dublin, June 4th 2014

Pseudo Localization
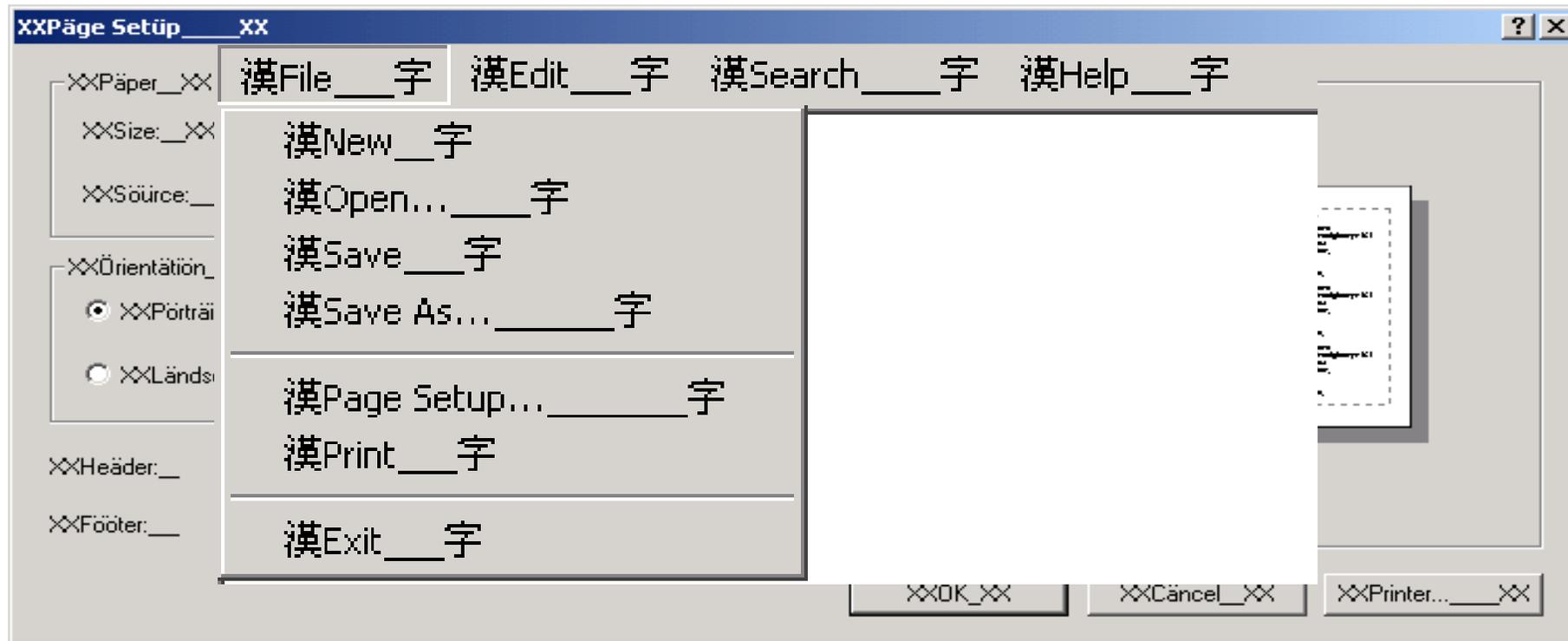
Translation Memory Tool

- A system (most often a database) that stores source sentence plus translation as a pair, a so-called „segment pair"
- During translation the translation memory compares the segment to be translated with the segments in the database.
- If a match is found (same or similar segment), the translation is offered as a suggestion
- The translator decides if the translation can be accepted or has to be changed.
- The TM system does NOT translate by itself, it is no machine translation system!

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

SDL Trados Studio

Navigation within file

Different tabs for term recognition, concordance window, comments, messages, searching the term base

Translation window

Parallel preview for Word, PPT HTML files

Program navigation

Floating windows for results from TM, termbase view

MemoQ

Sentence and terminology matches

Source and target language columns for translation

Differences between sentence and TM match

Localization Processes

Software Localization Tool

Pseudo translation

↓

File preparation

↓

Extract translatable segments

↓

Translation of software

Translation Memory Tool

Export of segment pairs for TM

Export of Terminology

Import into TM

Import into term base

↓

Translation of Online-help, readme files, manuals, web pages, packaging...

LOCALIZATION WORLD
CONFERENCE & EXHIBITS
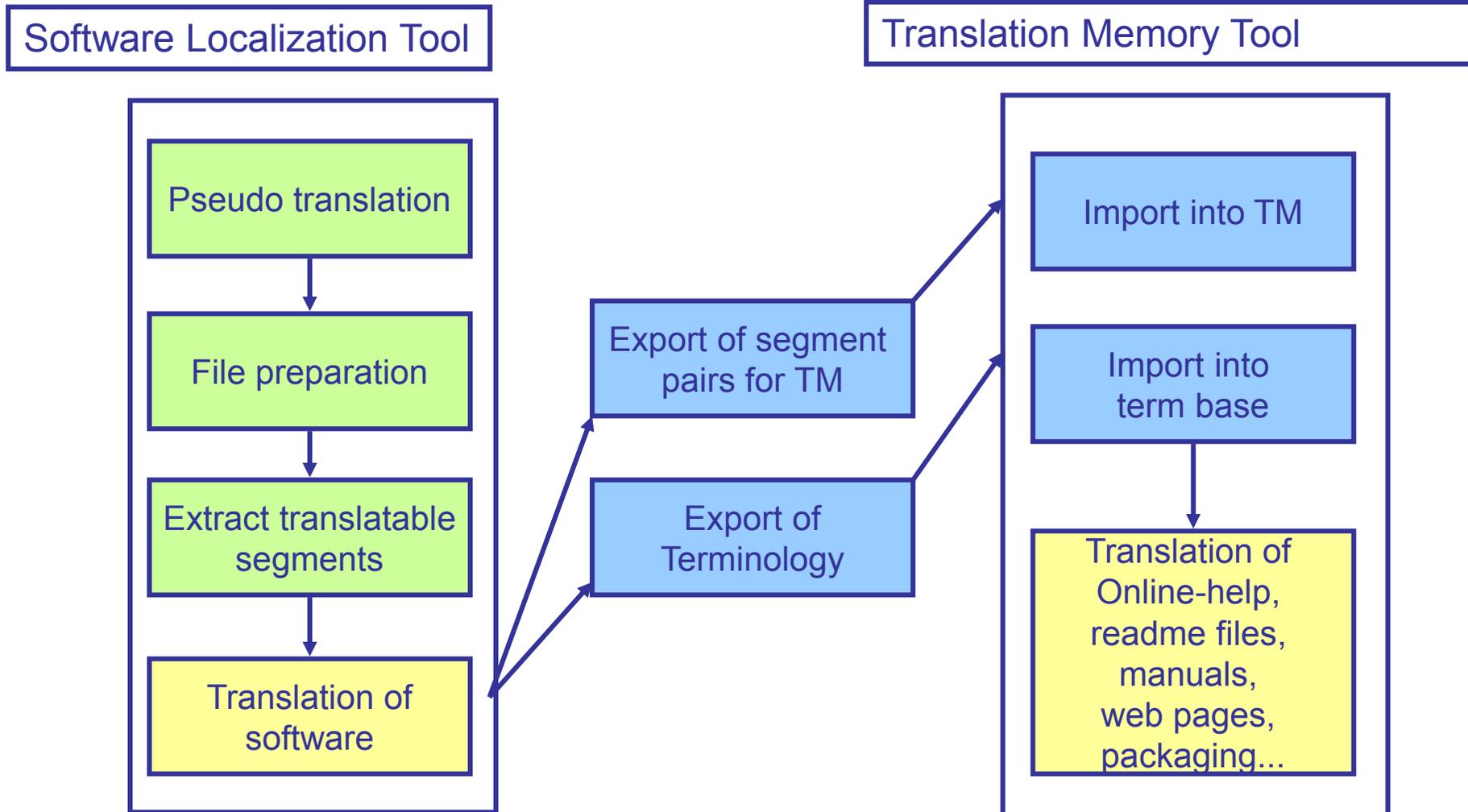
Terminology Management

- Components of TM tools or stand-alone solutions
- Connect to the TM systems and localization tools during translation
- Manage additional information like explanations, definitions, classifications and graphics
- Ensure the consistent use of terms over the whole project through term checks
- Term extraction (monolingual and bilingual)

Term Extraction

- Concordance tools
  - Extraction all words and word combinations up to x words from a monolingual document
- Statistical extraction
  - Extracting the most frequent terms from monolingual or bilingual sources
- Linguistic extraction
  - Extracting by rules and with the help of language analysis (e.g. all noun phrases up to 4 words)

Term Extraction

Alignment

- Old source and target language documents are read into the alignment component of the TM tool
- The tool segments the files and tries to connect the segments that belong together, thus creating segment pairs
- A translator checks the alignment
- Results are imported into a TM system for reuse with new translations

Project Management and Workflow Tools

- Project Creation in TM tool
  - packaging of project files
- Workflow Tool
  - Automation of processes (file conversion, pre-translation, packaging, sending out package to assigned translator...)
- Project Management Tools
  - Offers and invoicing

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

TM – MT

- Interactive translation

- interactive process

- almost all language pairs possible

- creation of a repository

- Recycling of translations independent of the format of the source document

- Machine translation

- fully automated process

- only works for the language pair the system was created for

- text is usually pre-edited and or post-edited

- good systems are relatively costly

- very fast

Machine Translation

- Today, machine translation is often added as an additional component in the translation process, to deal with text that has no match from a TM

- Online MT systems like Google Translate and Bing Translate can be easily attached to TM systems (most systems have the connection already prepared, it just needs to be activated)

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Standards, XLIFF

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Without standards

- High switching cost between LSPs
- De facto standards
- Higher cost of doing business
- Agility impact
- Tool & technology lock-in

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

With standards

- Consistent, reliable interchange mechanism
- Greater automation, improved tools compatibility
- Reduced localization friction
- Easy LSP switching, TM sharing
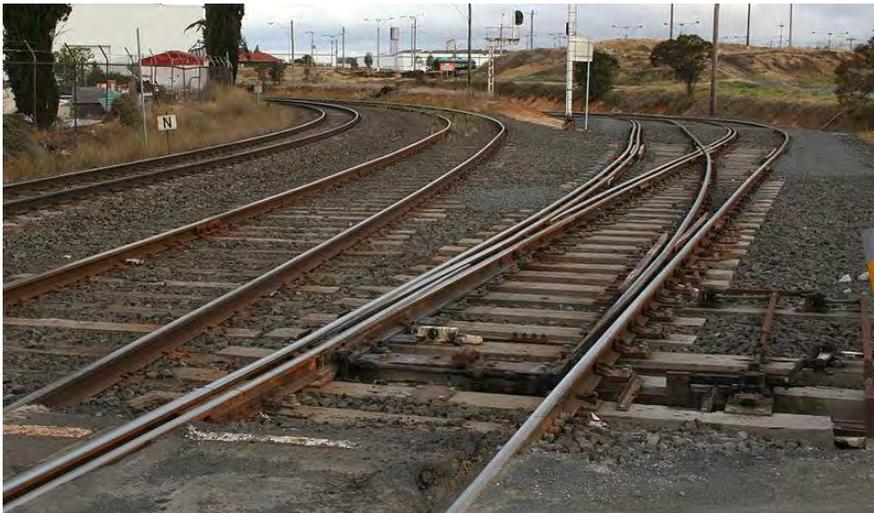
Some (Awful) Words About Standards in Localization

- TMX? TBX? GMX?
- We don't have to day and good and valid standard in the Localization industry
- Reasons:
  To accept a standard there is a good business need and / or regulation (which is, actually, a business need)
- We have good standards in the Localization, but no real business needs for them
- We need organization with deep-pockets (Microsoft, Google, Oracle, IBM, Apple etc.) to take part in defining and driving the standards

- Look at Unicode, Look how UTF-8 is not the major Unicode Transformation being used in the Internet. Think about organization with deep pockets.

What's wrong with XLIFF 1.2 today?

- Open to interpretation
- Patchy tool support
- Data loss during localization
- Customization = custom format
- Limited support for software & non-text

Tool Name	Elements									
	alt-trans	bin-source	bin-target	bin-unit	body	group	seg-source	source	target	trans-unit
Alchemy Catalyst 10	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes
Araya XLIFF Editor	Yes	No	No	No	Yes	No	No	Yes	Yes	Yes
Microsoft LBA 5.3/MLP 5.3	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes
MemoQ 5.0.64	Yes	No	No	No	Yes	No	Yes	Yes	Yes	Yes
MultiTrans Prism	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes
Okapi Framework M16	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
OmegaT 2.3.0	No	No	No	No	No	No	No	No	Yes	Yes
OpenTM2 V.0.9.5	Yes	No	No	No	Yes	No	No	Yes	Yes	Yes
SDL Trados Studio 2011 SP1	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Solas v1	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes
Swordfish Translation Editor	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Translation Factory 4.5.14	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes
Translation Workspace 1.10	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
XTM 6.2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

LOCALIZATION BUYERS
"Please localize my files"

LSPS
"Your files are translation-ready"

TRANSLATORS
"Which tool must I use?"

XLIFF 1.2* → Trados

XLIFF 1.2# → MemoQ

XLIFF 1.2++ → Catalyst

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Dublin, June 4th 2014
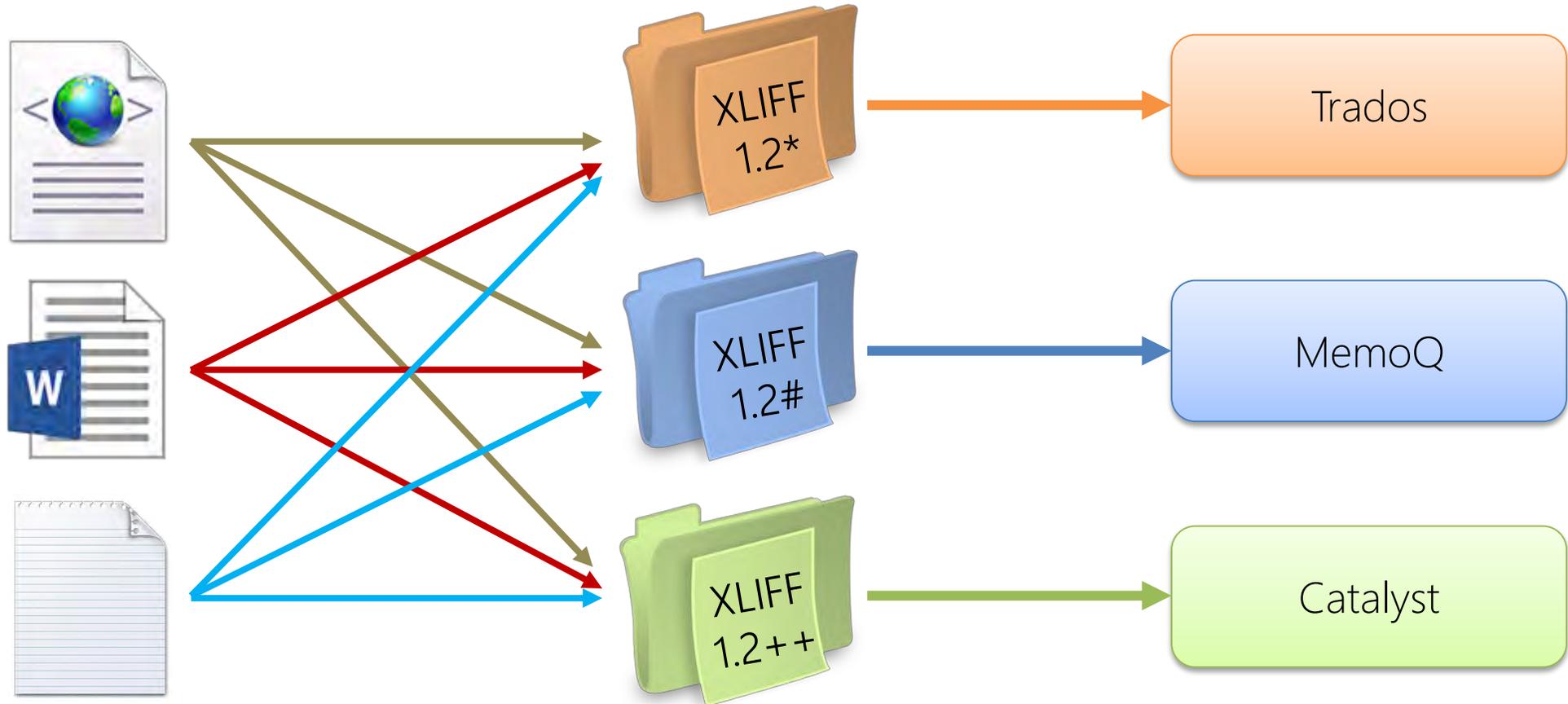
XLIFF 1.2 in practice

- Custom XLIFF converters
  - Benefits of process reuse, but with a cost
  - The price of XLIFF creativity
- Popular XLIFF implementations
- Interoperability Now!
- Microsoft-published XLIFF 1.2 implementation
- Walled garden approach
  - Lock clients into single eco-system
  - Use for internal projects

" **XLIFF 2.0** is at the core of a highly automated system for moving **localization data** from anywhere, to anywhere, with a minimum of cost and complication on the way."

Marc Levinson, The Box

What is XLIFF 2.0?

- A brand new XML format dedicated for localization
  - Not backwards compatible with XLIFF 1.2
- Designed for transporting localization between tools and systems
  - Supports both software & content
  - Binary data, images, reference files
- Guarantees data transfer and roundtrip
  - Rules for data preservation
- Designed for future expansion
  - Modular design
  - Room for controlled customization

Modular Design

Mandatory

Core

With XLIFF 2.0 you can…

- Include contextual screenshots with source text
- Mix reference language translations with source
- Bundle context & notes with your localization
- Use a single file format for all localization exchange
- Set validation rules for your target translation
- Use any compatible translation editor

Localization Buyers

- Supplier choice
- Simplified business model
- Use XLIFF as publishing format
- Promote XLIFF 2.0 among your LSPs

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Tool Makers

- Specialize in custom converters, XLIFF validators, etc.
- Innovate and compete against big players
- Develop and share custom modules
- Become the "best" XLIFF 2.0 editor

LSPs

- Bundle context with localization
- Provide reference language information
- Save on engineering costs for clients

Translators

- Choose your own CAT tool
- Utilize context and validation for more accurate translation
- Become expert in XLIFF 2.0 features and modules

More info

- http://multilingual.com/articles/201406-42/ #xliff!
- http://docs.oasis-open.org/xliff/xliff-core/v2.0/xliff-core-v2.0.html

Testing / QA

Testing / QA

5 types of testing:

- Before localization
  - i18n testing
  - l10n readiness testing (pseudo localization)
- After localization
  - Cosmetic testing
  - Linguistic testing
  - Functional testing

Testing / QA

Effort Estimations:

- i18n QA: the same timeframe as the original acceptance tests

- Pseudo localization: the same timeframe as the original acceptance tests
  Should be done by the dev team

- Cosmetic/ linguistic – one pass on all dialogs/ screens/ menus etc. Usually a matter of days.
  Should be automated as much as possible

- Functional testing - the same timeframe as the original full test cycle of the original product

Testing / QA

i18n testing:

- Is the software really locale independent
- Does your software know how to handle data in different languages (double-byte enabled?)

Development responsibility

Testing / QA

Cosmetic Testing:

- Check to see if the UI is broken
- Dialogs, buttons, menus etc. – have they been properly localized
- Chinese words are shorter, but the characters are higher!
- French words lengths...

- Automation
- On every build!

Testing / QA

Linguistic testing:

- Does the translation make sense in the context?
- Edite vs. Edition
- Share vs. Shares

Testing / QA

Functional testing:

- Full acceptance test of the product in target language
- Tests reports and log file for manual testing and test automation
- Monitor the progress versus monitor the end result

- Not always done due to cost and time

Testing / QA

In country reviewing:

- Resources in or from the country/market, who know the target market and target language to check if localization makes sense

- Opportunities for the crowd

Crowdsourcing translation- is it real?

YES
The end
Questions?

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Today's menu:

- History and Examples
- Definition
- Why do we need Crowdsourcing
- Examples in the translation world
- Where Crowdsourcing can work (and where not)
- Crowdsourcing HOWTO
- Q/A & Discussion

History

iStockPhoto
http://www.istockphoto.com/index.php

- The low cost of Digital Cameras
- Photo editing software
- The Internet

History and Examples

- Linux
- Firefox

And other open source initiatives

[Goodbye, "free software"; hello, "open source"](#)

"(Yes, we're aware of the specialized meaning "open source" has in the intelligence community. This is a feature, not a bug.)"

(Eric S. Raymond , 8 February 1998)

History and Examples

Amazon Mechanical turk

- www.mturk.com
- Breaking down tasks to little Human Intelligence Task (HIT)
- Micro payments

Interesting example: the search after Steve Fossett, 2006

Steve Fossett Missing: Help find him by searching satellite imagery

New examples and instructions! If this is your first time working on this task, please carefully review the instructions further down.

Review This Image

approx 85m / 278ft

approx 85m / 278ft

(c) This imagery is copyright GeoEye

Example

Example plane crash site

Example of an airplane and an airplane crash showing the size of object being searched for.

Need More Detail?

To view in Google Earth, load the KML file below then cut and paste:

38.290787,-118.732681

in the "Fly To" box found at the top left corner of the application.

For a similar viewing experience in Google Earth to the above image, navigate to an altitude of roughly 1,500 feet.

IMPORTANT: Please ensure that you've loaded the following KML file below in Google Earth before navigating to the co-ordinates. Otherwise, you risk looking at old and irrelevant images.

KML file for Google Earth Searching:
http://s3.amazonaws.com/Fossett-GE4/GeoEye_4/index.kml

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

History and Examples

Wikipedia

- The best source for any high-school student

"Wikipedia: The Faith-Based Encyclopedia" by Robert McHenry, Former Editor in Chief of the Encyclopedia Britannica, 2004

History and Examples

www.waze.com

- Traffic reports using your mobile
  (no need to read info from cells)

- Every mobile phone is a small contributor to the big picture

History and Examples

- Blogs and other user generated content
- Media and news (CNN iReport)
- Twitter?
- YouTube?

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

History

2006: The Rise of Crowdsourcing, Jeff Howe (Wired magazine)

http://www.wired.com/wired/archive/14.06/crowds_pr.html

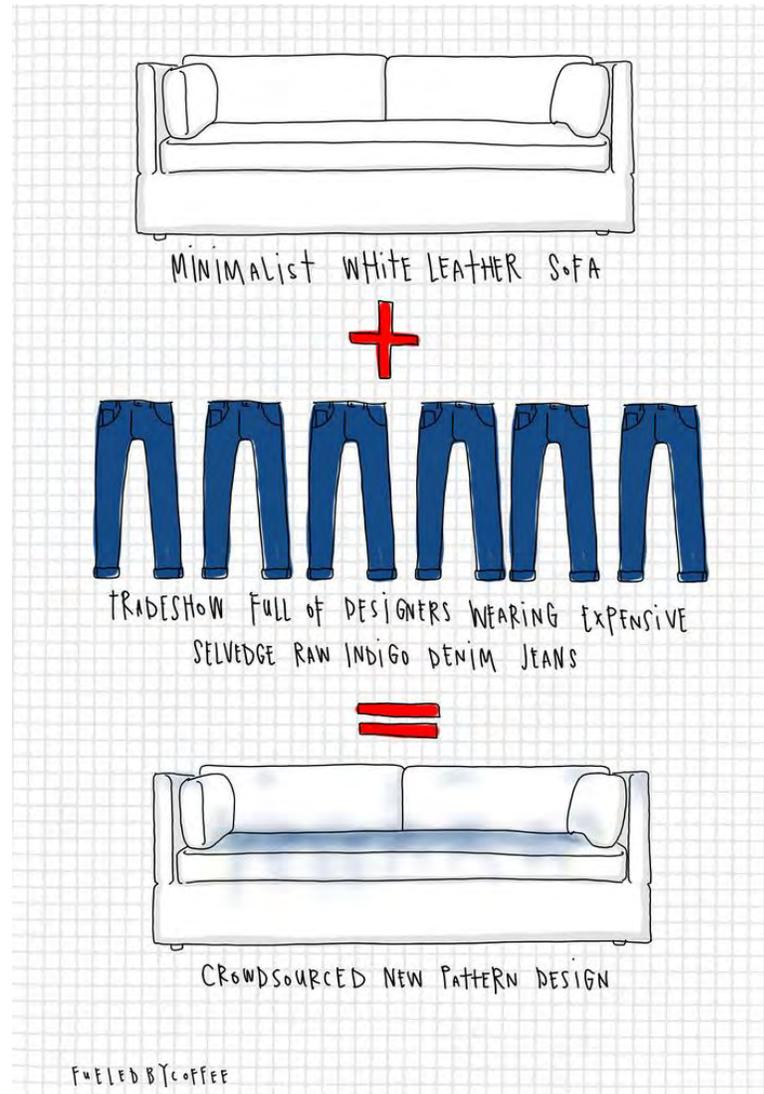http://www.youtube.com/watch?v=F0-UtNg3ots&feature=related

Definition

Jeff Howe:

Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an **undefined entity**, generally a large group of people, in the form of **an open call.**

A very good definition

MINIMALIST WHITE LEATHER SOFA

+

TRADESHOW FULL OF DESIGNERS WEARING EXPENSIVE
SELVEDGE RAW INDIGO DENIM JEANS

=

CROWDSOURCED NEW PATTERN DESIGN

FUELED BY COFFEE

Definition

Crowdsourcing doesn't mean free.
Crowdsourcing doesn't mean low quality.

Definition

Crowdsourcing changes business models:

- Customers and users become potential partners
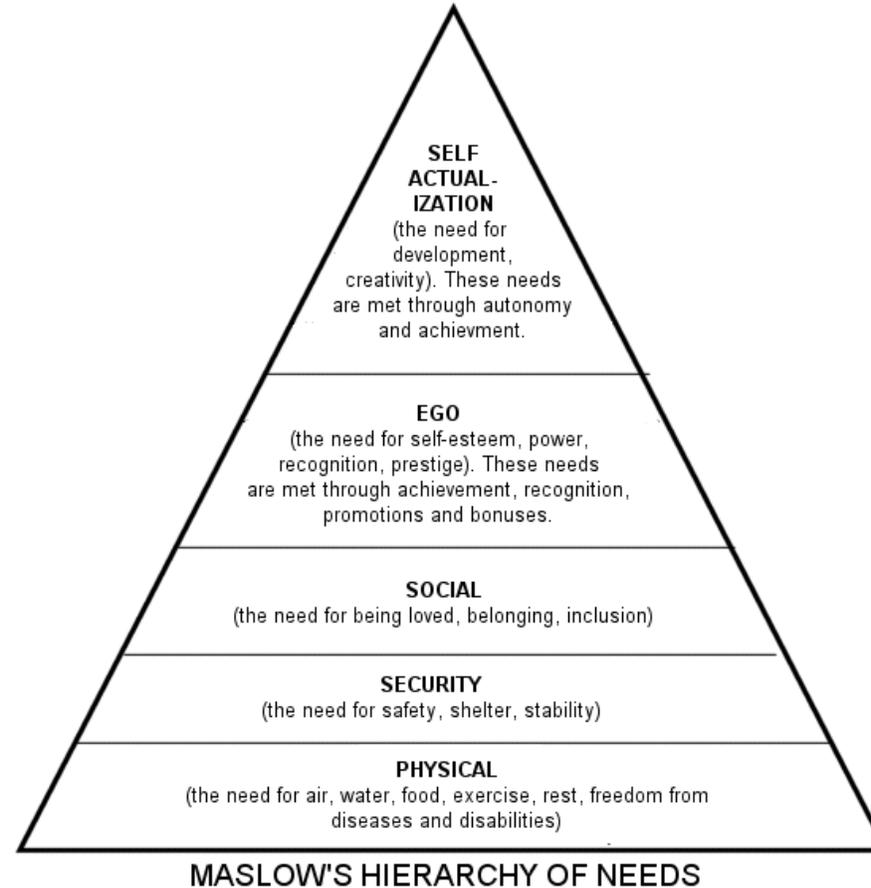- Not only buying or consuming, but participating

"Put simply, the economics of crowdsourcing involves leveraging a user community's skills and interest to deliver a specific service to the community and a (content) producer."

Francis Tsang, Adobe

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Definition

Technology helps us,
the individual,
to express ourselves and cross barriers.

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Why Crowdsourcing Works

SELF ACTUALIZATION
(the need for development, creativity). These needs are met through autonomy and achievment.

EGO
(the need for self-esteem, power, recognition, prestige). These needs are met through achievement, recognition, promotions and bonuses.

SOCIAL
(the need for being loved, belonging, inclusion)

SECURITY
(the need for safety, shelter, stability)

PHYSICAL
(the need for air, water, food, exercise, rest, freedom from diseases and disabilities)

MASLOW'S HIERARCHY OF NEEDS

Source:
http://www.omafra.gov.on.ca/english/rural/facts/96-001.htm
http://en.wikipedia.org/wiki/Maslow's_hierarchy_of_needs

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Why Crowdsourcing Works

Crowdsourcing Creates **communities**
with **common goals** and **passion**

Look at Seth Godin's talk on the tribes we lead
http://www.ted.com/talks/lang/eng/seth_godin_on_the_tribes_we_lead.html

Why Crowdsourcing Works

Individuals happily contribute to Crowdsourcing

The incentive is psychological, not financial

Why Crowdsourcing?

- Cost (not always)
- Efficiency
- Big range of talents
- Listen to the crowd!
- Involving the community: brand-building, ownership

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Why do we need translation Crowdsourcing?

(Crowdsourcing in our little puddle of translation)

- Too much content to be translated
- Too many languages

How to prioritize?

- A great way to build dedicated communities
- A great channel for marketing and getting market feedback

Why do we need Crowdsourcing?

2007

Internet usage by percentage of each country's population
Source: Wikipedia, of course

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Why do we need Crowdsourcing?

2012

100% 80% 60% 40% 20% 0%
No data

Internet usage by percentage of each
country's population
Source: Wikipedia, of course

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

http://www.bytelevel.com/map/ccTLD.html

Why do we need Crowdsourcing?

	WORLD INTERNET USAGE AND POPULATION STATISTICS June 30, 2012					
World Regions	Population (2012 Est.)	Internet Users Dec. 31, 2000	Internet Users Latest Data	Penetration (% Population)	Growth 2000-2012	Users % of Table
Africa	1,073,380,925	4,514,400	167,335,676	15.6 %	3,606.7 %	7.0 %
Asia	3,922,066,987	114,304,000	1,076,681,059	27.5 %	841.9 %	44.8 %
Europe	820,918,446	105,096,093	518,512,109	63.2 %	393.4 %	21.5 %
Middle East	223,608,203	3,284,800	90,000,455	40.2 %	2,639.9 %	3.7 %
North America	348,280,154	108,096,800	273,785,413	78.6 %	153.3 %	11.4 %
Latin America / Caribbean	593,688,638	18,068,919	254,915,745	42.9 %	1,310.8 %	10.6 %
Oceania / Australia	35,903,569	7,620,480	24,287,919	67.6 %	218.7 %	1.0 %
WORLD TOTAL	7,017,846,922	360,985,492	2,405,518,376	34.3 %	566.4 %	100.0 %

Internet User Distribution   - Source: http://www.internetworldstats.com/stats.htm

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Examples in the translation industry

- Google started(?) years ago with GIYL (UI)
- Sun (Oracle, UI)
- Mozilla (Firefox and more, UI)
- Facebook (UI)
- Adobe (Content)
- Remember the milk (UI)
- DotSub
- Microsoft (rating content translation)

Open source projects

- Mozilla Firefox
- Openoffice (Sun Microsystems)
- Ubuntu (https://translations.launchpad.net/)

- Highly structured organization: Contributors are selected, rated and monitored
- Very high quality (of both code and translation)

Google (GIYL)

- (mainly for the Search page)
- One of the first
- 100 languages including Klingon
- Practically no validation

Pros:

- Simple
- Low tech
- Minimal investment
- Little management

Cons: No real Quality control

Google (GIYL)

Issues

- Rating of translators by quantity and not quality
- No real TMS behind
- Not scalable
- Bad quality

Facebook

- Structured and controlled
- Clear process of translation rating
- Well defined quality process
- Automation

Pros:
- Quality
- User participation
- Very fast growing community(ies)

Cons:
- Management
- Technology support

Quality

- "But what about quality?"
- "But what about schedules and deadlines?"

Honestly, we can ask the same questions to some of the (paid) vendors we know...

Quality

- Don't patronize the users
- Let the user define and decide what quality is

Daniel's lemma: The 95% trash rule

Where Crowdsourcing can work:

- Companies with vision
- Using the right content
- With potential committed community
- Mutual benefits
- Automation
- Built for scale
- Quality is a need, but not a must (ouch, we need to define first Quality)

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Where Crowdsourcing will not work

- No mutual benefit (exploiting the user)
- Boring content (the user manual of a network switch)
- Quality is a must (health care device)
- Confidential content

Where to start from?

- Identifying the right products/ content
- Identifying what needs translating
- Defining quality
- Defining the process:
  - Scalable
  - Sustainable (how to deal with the maintenance?)
- Implementing automation

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Steps:

- Define problem
- Broadcast to the right crowd
- Crowd submits solutions
- Vet
- Rewarding (?)
- Profit (company? Community?)

Issues & Risk:

- Exploiting the crowd without building communities
- Unclear definition of quality will result in bad quality
- The Chicken and Egg problem

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Issues & Risk:

- Ethical
- Social
- Legal
- Economic

Summary

- The are different type of Crowdsourcing
- Crowdsourcing is here
- Crowdsourcing is not exploitation
- Crowdsourcing is not free
- Crowdsourcing quality can be high
- We need to adapt to it
- It is an opportunity, not a threat for our industry
- It is a huge opportunity for both individuals and communities

Integrating Localization in an Agile Process

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Agile development - SCRUM

Agile development:

- A group of Software Development Methodologies base
- Frequent inspection and adaptation
- Team work
- Self-organization and accountability
- Rapid delivery of high-quality software

Agile development - SCRUM

- An iterative incremental process of software development
- Each development cycle, called a "Sprint", takes 15-30 days
- The deliverable of each sprint is an increment of shippable software
- The scope of work for each sprint is derived from the product backlog, which is a prioritized set of high level requirements
- The tasks ("stories") and scope of work for each sprint are defined in the sprint planning meeting and the team is committed to complete those tasks
- No "stories" can be added during the sprint

Agile development - SCRUM

24 h

30 days

Product Backlog

Sprint Backlog

Sprint

Working increment of the software
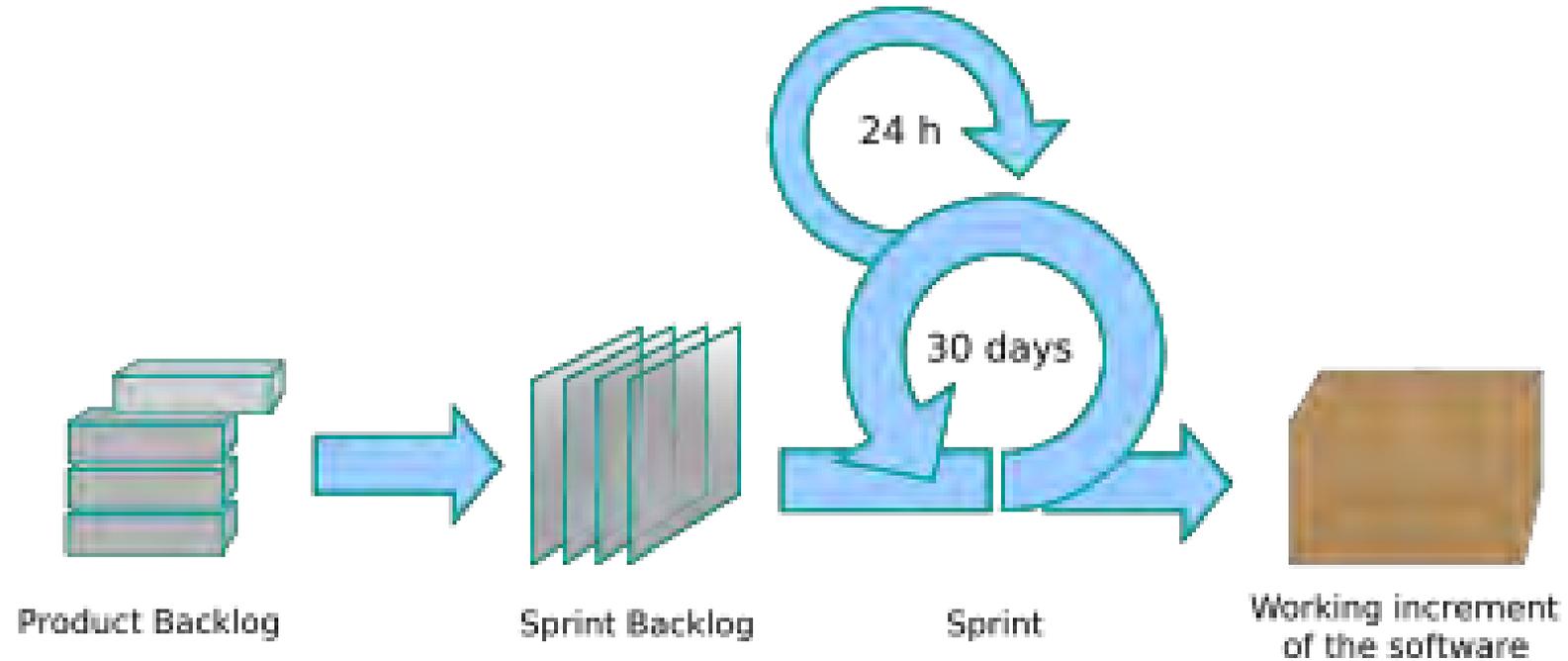
Source: Wikipedia, of course

Agile development - SCRUM

Roles:

A pig and a chicken are walking down a road. The chicken looks at the pig and says, "Hey, why don't we open a restaurant?" The pig looks back at the chicken and says, "Good idea, what do you want to call it?" The chicken thinks about it and says, "Why don't we call it 'Ham and Eggs'?" "I don't think so," says the pig, "I'd be committed, but you'd only be involved."

Pigs are committed to build the software.

Chickens are interested in the software.

Agile development - SCRUM

"Pig" roles
- Pigs
  The ones committed to the project and the Scrum process
- Product Owner
  Customer's voice
- ScrumMaster
  Facilitated the process
- Team
  5-10 people

Agile development - SCRUM

"Chicken" roles
- Users
- Stakeholders (Vendors, Customers etc.)
- Managers

Agile development - SCRUM

The Scrum meeting - standup
- Daily
- Starts on time
- 15 minutes
- "Chickens" are welcome, yet only "pigs" speak
- All attendees should stand
- Time and location are fixed

Agile development - SCRUM

The Scrum meeting - standup
* Yesterday's work
* Today's work
* Problems / issue preventing you from accomplishing your goal

* ScrumMaster might de-scope stories from the sprint

Agile development - SCRUM

Retrospective meeting
- In the end of each sprint
- What went well in the sprint
- What can be improved

Agile development - SCRUM

How localization works in the SCRUM model?

Constraints:

- Very short iteration, short deadlines
- Need to fit preparation, leveraging, translation, reviewing, QC and more
- Very small tasks
- Work and tasks are defined in a short notice

Agile development - SCRUM

Localization can be a sprint of its own or can be incorporated as a story in a sprint.

- Pros?
- Cons?

Agile development - SCRUM

Process

A localization representative need to be part of the team:
- Planning meetings
- Standup meetings
- Retrospective meetings

The development team needs to understand the constraints of localization:
- External vendors
- Rigid timelines (words/days, reviewing, QC etc.)
- Cutoff dates

Agile development - SCRUM

Process

Few approaches:
- One batch of localization / sprints
- Batch for each "done" story
- Daily batched

Agile development - SCRUM

How localization works in the SCRUM model?

Technologies:
- Use automation to extract resources ("continuous build") and integrate them back
- Use automation to leverage against your linguistic assets
- Some localization tools can integrate into the production environment
- Use a GOOD resource representation (XLIFF?) that contains all the needed info and reduces manual steps ("cut & paste" to Excel sheet)

Agile development - SCRUM

Pros:
- Good for maintenance of products (websites etc.)
- Good when approaching release dates

Cons:
- The development team, the localization team and the vendor need to be mature enough to implement it

QA/Discussion

Planning Tips

Planning Tips

- Kick off meeting
  - Touch on a all aspects of project, size, timeline, number of languages etc.

- Analysis of source meeting
  - Outline potential L10n/I18n issues with source code

- Scheduling and budgeting
  - Based on size, timeline, number of languages etc. schedule resources, quotes,

- Terminology setup
  - Create glossary leveraging existing glossaries, adding additional terminology by using tools such as SDL Trados TermExtract.

- Preparation of source Material
- and…..

Planning Tips

- Translation of Software
  - Translation, editing and proof-reading (TEP) of software

- Translation of documentation
  - Translation, editing and proof-reading (TEP) of documentation

- Testing the Software
  - Testing of software for functional, linguistic and cosmetic defects

- Screen Capture
  - Capture screenshots for documentation, help files

- DTP
  - Prepare the hard copy of the documents

Planning Tips

- Start planning from the end: focus on the release date
- Make sure that you work within a realistic timeframe – allow extra time, in case things go wrong (buffers, slippage, holidays)
- Check the required time for QA
- Estimate number of words, make sure what your are paying for (source/target)
- Rule of thumb:
  Number of words / 2000 = number of translator days for translation
  - Software = slower
  - Flowing documentation ~ faster
  - Diminishing returns as more translators added

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Planning Tips

- Keep in mind that translations can start before all resources are ready
- You can start translating your material once the GUI is frozen
- Think about running QA for several languages in parallel
- Remember that the process might require several iterations

Pitfalls

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Pitfalls

"We are not doing any localization nor translation. We will give our distributors in each country a discount, and they take care of it"

Careful – consider the following:

- Who is in the end responsible for quality?
- Who owns the Intellectual Property?
- No leveraging of handling the localization for all countries at once.

Pitfalls

"There is no need for a localization process, once we release the product, we will prepare Excel files with the strings to be translated"

Careful – consider the following:

- Has your software been prepared for localization?
- Be ready for surprises in the code
- Consider pseudo localization
- Translation out of context can result in errors and/or excessive project management time

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

Pitfalls

"Philippe, from engineering, speaks French fluently, lets ask him to translated the GUI of our product!"

Careful – consider the following:

- Languages are evolving – therefore best translations will be done using in-country translators
- What about localization?
- What about using translation tools?
- Leveraging, Terminology, Glossary?

Some final (and unrelated) words

- We need to bring content to those who don't have it
- We need to stimulate content creation in those area and provide them the infrastructure
- We need to do it without business incentives (but to combine it with those who have ones)
- We need to do it without patronizing anyone
  (please recall that we manage to destroy earth successfully. Maybe we can learn something from others...)

QA/Discussion

Thank you for your attention

Jargon

Jargon

- g11n
- i18n
- l10n
- Sim ship
- MLV
- SLV
- SLA
- Translation Memory (TM)
- Segment
- Matching (100%, ICE, Partial, Fuzzy)
- Leveraging
- Alignment
- Glossary, Glossary building
- Terminology management
- Machine Translation
- Localizing Marketing
- Translating Guideline
- NDA

- Software l10n
- Resource
- Resource ID
- Context
- Localization Tool
- QA
- Linguistic QA
- Cosmetic QA
- Functional QA
- Reviewing
- Proof Reading
- Localization Readiness
- Pseudo Localization
- Single source
- Word count
- CMS
- Publisher

LOCALIZATION WORLD
CONFERENCE & EXHIBITS