

Is your global business at risk?

Unicode Conference IUC 42



1

Presenters

Tex Texin

- Globalization Architect, XenCraft

Michael McKenna

- World-Ready Architect, PayPal, Inc.

Unicode Conference IUC 42



2

Abstract

There are many requirements to create a high quality global application that customers value.

Failing to meet these requirements may hurt sales. More egregious failures may damage your brand.

Some mistakes can result in legal action, penalties and even expulsion from a market.

Two of the industry's seasoned experts will describe globalization requirements, key performance indicators, and potential business risk.

If you are unsure of the risks that errors in globalization can cause to your organization and how to avoid them, attend this session.

This presentation will include, among other topics:

Assessing the risk that your product isn't ready for global distribution.

Capturing metrics that quantify quality and assess whether your organization is improving.

Automating the collection of key performance indicators via automated systems where possible.

Unicode Conference IUC 42



3

Globalization is a process

- It is not turnkey
 - One input has many possible outputs
 - It can be complex, with many points of potential failure
 - Working processes can degrade
 - Once successful results, can be unusable at a later time
- Failure can be catastrophic

Unicode Conference IUC 42



4

A simple catastrophic example



A large company rolls out a worldwide marketing campaign associating their new product for “heroes”
A fireman is featured as an example of a hero.

Unicode Conference IUC 42



5

Fireman as a hero

- Campaign failed outside North America
 - In USA & CA, firemen are brave and run into danger to save lives.
 - Elsewhere, firemen just clean up the rubble.
 - Besides materials cost, market momentum, relationships with distributors hurt
- Could have **easily** been prevented by in-market testing.

Unicode Conference IUC 42



6

Translation is not turnkey

- A company's knowledge DB is called an "encyclopedia" to distinguish it in market
 - Competitors use the term "repository"
 - Their manual was translated and printed
 - "Encyclopedia" translated to the French equivalent of "repository"
- \$1Million of publication was tossed

Unicode Conference IUC 42



7

Translation is not turnkey

- This failure should have been prevented
 - There was a review before printing
 - Were reviewers thorough?
 - Did reviewers know what to look for?
 - How do you establish a review is successful?
- How would you reduce risk for a simple translation project?

Unicode Conference IUC 42



8

Risk reduction

- Glossary of approved brand and key terms
 - Include a list of terms to avoid
 - Scan translation for terms to avoid
 - Hold translators and reviewers accountable to use key terms
- Inject known problems into translations
 - so reviewers can be graded
- Metrics- reasonable? in line with history?
 - Error count, translation wpm, review wpm, et al



9

Reducing international risk

- Many ways products fail in global markets
- Many points in development where you can evaluate the potential for failure.
- Catching and correcting problems early is significantly less expensive than later.

Unicode Conference IUC 42



10

Don't "go through the motions"

- All points of the process need monitoring
- Executing "reviews" does not assure proper checking
- Quantitative measures allow evaluation of quality of both products and process
 - KPI = Key performance indicators
 - Comparing KPI over time shows problem areas and whether quality and efficiency are improving

Unicode Conference IUC 42



11

Naïve Localization Process

- Source language/locale in, destination out
- Presumes well-defined, error-free process
 - translation is 1 to 1 conversion
 - translators (or Machine Translation) are perfect
- No QA; Market judges quality (and you)



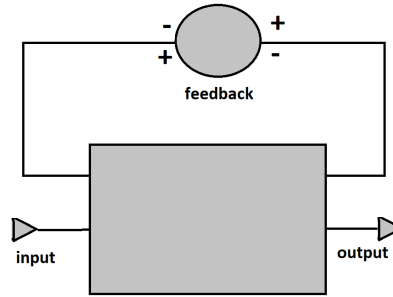
Unicode Conference IUC 42



12

Final stage QA

- Rerun the process if changes needed
 - Translation, and/or internationalization
 - Expensive, unreliable delivery
 - Fixes often pushed to next release



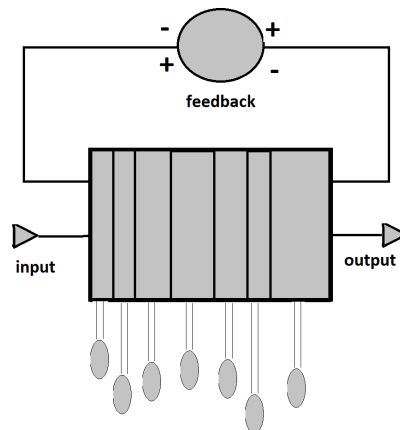
Unicode Conference IUC 42



13

Continuous monitoring

- Early, less-costly correction
- Systemic, not just spot fixes
 - E.g. string externalization, plurality



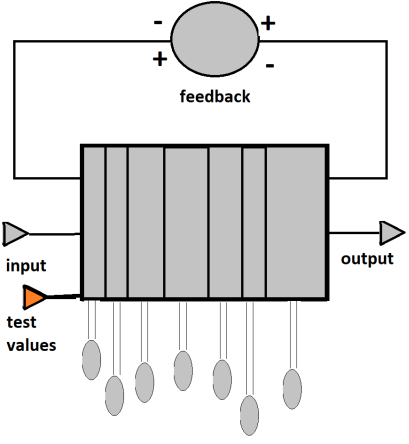
Unicode Conference IUC 42



14


Test the process

- Proof points
 - Pseudolocalization
 - Error injection
 - Illegal utf-8
 - Impossible date-times
- Repeat periodically
 - detects degradation
- Automate testing and data collection
 - For consistency, reliability



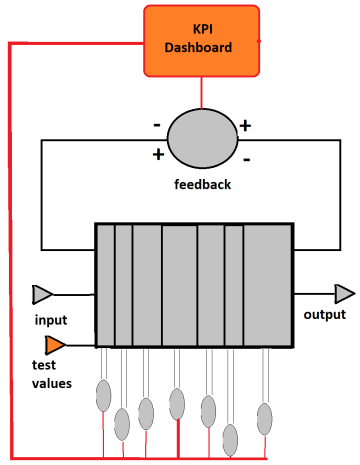
The diagram shows a process represented by a grey box with vertical bars. It has an 'input' on the left and an 'output' on the right. A 'feedback' loop is shown as a circle with '+' and '-' signs, connected to the process. Below the process, several vertical lines lead to grey oval shapes labeled 'test values'.

Unicode Conference IUC 42 15




KPI Key Performance Indicators

- Quantitative values
 - Objective comparison of process changes
 - Intentional or not
- Diagnostic
- Analysis
 - by team, code, phase, release, market, et al
- Dashboard
 - Transparency
 - Urgency: green, yellow, red
 - Automated data collection



This diagram is similar to the one on slide 15, but it includes an orange box labeled 'KPI Dashboard' at the top. A red line connects the 'test values' to the dashboard. The feedback loop and process components are identical to the previous slide.

Unicode Conference IUC 42 16



Root cause analysis

- Wow- we have many translation errors
 - How many is a lot?
 - 100 errors in 1000 or in 100,000 words?
 - Translator mistranslated?
 - Delivery or timing or versioning error?
 - Inadequate context (orange fruit vs color)
 - Poor code (word order, gender, plurals)
 - UI/Text design (pun/ phrase with 2 meanings, metaphor, invented term -heads up display)

Unicode Conference IUC 42



17

KPI Candidates

- # Errors, %total, errors/1000 lines, coverage, severity
 - Translation by topic, language, translator, reviewer, file type
 - Code: by type, developer, technology, file, module
 - QA: # tests, %total by language, type,
 - Design: flaws, #reviews, by language, etc.
 - Process: #code scans, results, scan misses
- Timing, duration (rushed, planned, conflicted)
- Training: %attendance, coverage
- Compliance: #exceptions to defined process
- Correction: time to investigate, design, fix, test, age, cost
- Support: #calls, #returns, field reports, sales,

Unicode Conference IUC 42



18

Defining process risk

- What do you expect to achieve?
 - Globalization goals and requirements
 - Priorities: quality, cost, turn-around time, customer satisfaction, ROI, et al
 - Rank priorities
 - Roles and accountability
 - Internal, external (LSP, CDN, et al)
 - Process definition
 - What can be measured and how?

Unicode Conference IUC 42



19

Process definition

- A well-defined localization process
- A balanced, strategic QA process
- A well-defined correction process
- Clear ownership and accountability
 - For people, teams
 - For stages
 - go, no-go- avoid escalation, crisis management

Unicode Conference IUC 42



20

Tactics

- Problem avoidance (e.g. training, guides)
- Prevention (source scans)
- Detection (testing, QA)
- Correction process
 - track, priority, severity, reoccurrence, root cause analysis
- Dashboard/metrics and Analysis
 - Status: Track progress, forecast completion
 - Insight: Reward positive behavior, fix problem areas
 - Futures: Lower bars to compliance, automation

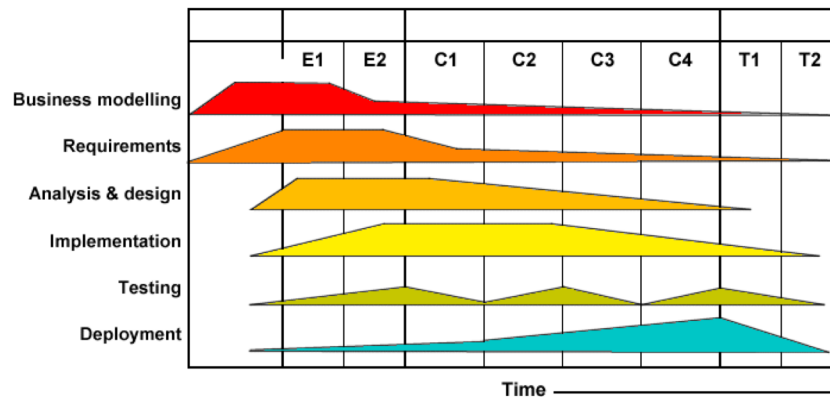
Unicode Conference IUC 42



21

Product Lifecycle and Risk

Epics, sprints, and scrums →



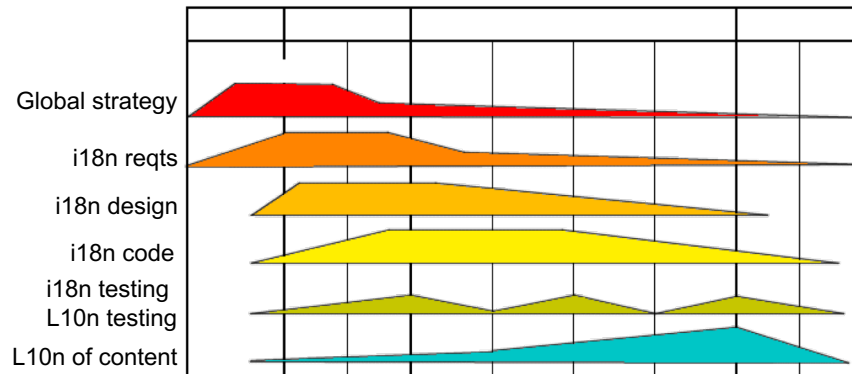
Unicode Conference IUC 42



22

Product Lifecycle and G11n

Different KPI depending on stage



Unicode Conference IUC 42



23

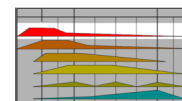



Risk KPI measuring g11n

- Market
 - Are international targets set?
 - Actual vs Expected revenue
 - A:B testing and international results
- Regional support
 - Supported languages
 - Regional usability

KPI
Yes/no
Target
Actual
Yes/no
Current/planned
Regional tickets

Unicode Conference IUC 42




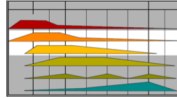



Global Requirements

- Locale Identification
- Text externalization
- Grammar rules
- Data format
- Layout/Form factor
- Graphics, Media
- Search, Sort, et al
- WAN Performance
- Financial
- Regulatory
- Privacy, GDPR
- Security
- Accessibility
- Cross-locale usage
- Domain requirements
 - Eg health, e-commerce

Numerous G11n checklists on the Web

Unicode Conference IUC 42




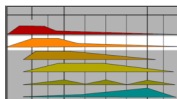
Assessing Risk


ready for global distribution

- Regulatory and Legal
 - Financial flows are legal
 - Regional servers where needed for data privacy and regulatory
 - GDPR
 - Advertising and Content are legal

KPI
Yes/no
schedule

Unicode Conference IUC 42



Assessing Risk

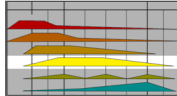
ready for global distribution

- Code – I18n Scans
 - Code scans for good, bad i18n patterns
 - Message scans for bad L10n patterns
 - Embedded dynamic content
 - Fragments and incomplete sentences


KPI


Yes/no

tickets



Unicode Conference IUC 42





Risk KPI

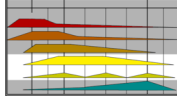
measuring g11n

- I18n Code Scan Analysis
 - I18n Areas
 - Use scans to determine code areas where i18n is/should be used
 - Bug Density
 - Issues/lines code
 - Issues/number i18n areas
 - Trends
 - Code change cause less/more issues?
 - Repeat problems?


KPI

Issue density

tickets



Unicode Conference IUC 42





Critical Test Values

- Values that are more likely to find errors
 - Boundary conditions
 - Supplementary characters
 - Date-time eg Dec 30-Jan 2, Leap yr Feb 28-Mar 1
 - Data transformation
 - Upper case changes length: Upper(eßen)=ESSEN
 - Replace changes byte length: resume->résumé
 - Turkish “i İ ı I” breaks ASCII case-insensitivity
 - ISO 8859-1 control characters 0x80-0x9F e.g. ™ €

See I18nguy.com for previous conference talk: “Critical values for I18n testing”

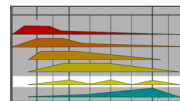



Assessing Risk

ready for global distribution


- i18n/L10n Testing
 - Functional testing – product doesn’t break
 - Pseudo-Localization – pre translation testing
 - Linguistic – are they the right words?
 - Esthetic – does it look right in context
 - Prioritize found issues
 - Functional show stopper
 - product won’t work
 - Linguistic show stopper
 - customer confused/offended

KPI
 Schedule
 Tickets
 Category
 Priority/severity
 Age



Unicode Conference IUC 42






Risk KPI

measuring g11n

- Test results
 - How many language/regions tested
 - Matrix of tests should cover all major functional differences
 - Don't need to cover every possible combination
 - Time to test
 - Should not adversely affect time-to-market
 - Automation
 - Build time – e.g. through Jenkins
 - Check in tests – e.g. through esLint

Unicode Conference IUC 42

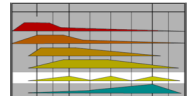



KPI

Coverage

Test time

Automation %






Assessing Risk

ready for global distribution

- Deployment testing
 - Last mile tests
 - Regional latency testing
 - CDN testing – content delivery
 - Mobile carriers testing
 - “Friends and Family” testing
 - Regional presence

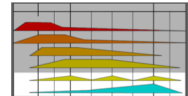
Unicode Conference IUC 42




KPI

Latency

Tickets



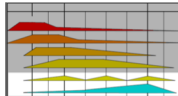


Assessing Risk


ready for global distribution


- Customer Support
 - Help in local language
 - Local contact/phone for region

KPI
Yes/no tickets



Unicode Conference IUC 42



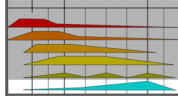


Risk KPI


measuring g11n

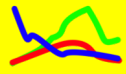
- Released Product tracking by Region and Language
 - Customer support issues
 - Rejection rate
 - Engagement
 - NPS – “Net Promoter Score” by Region
 - Sentiment analysis by Region

KPI
NPS
Tickets
Sentiment scores



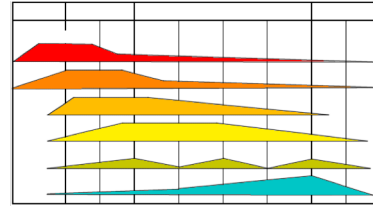
Unicode Conference IUC 42





Reporting Risk

- Convert to scores
 - Easy to compare
 - Reflect ROI or cost/benefit



Organize

- Overall
- By team, by org
- By KPI
- By age of code
- By Market
- By age of issue

Unicode Conference IUC 42



35

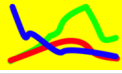
Putting it all together

KPI and dashboards

Unicode Conference IUC 42


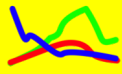


36




Dashboards


Use I18n Assessment Tools menu to		Enablement		Role			Overall
		World-Ready	Localizati	Planner	Designer	Engineer	
Scores on a scale of 0 to 10							
Product ABC	average	6.3	5.7	6.1	4.8	6.3	5.6
Details by category							
~Planning	Market Requirement	2.0	na	6.0	1.0	na	6.3
	Dependency	6.0	1.0	8.0	10.0	8.0	5.1
	Process	8.0	6.0	6.0	3.0	na	6.6
	Content	9.0	4.0	na	1.0	na	6.1
~Design	Interaction Design	2.0	3.0	8.0	7.0	4.0	5.7
	Text Processing	8.0	na	10.0	5.0	7.0	4.6
	Locale formatting	9.0	na	1.0	na	2.0	6.1
	Search	2.0	na	3.0	7.0	7.0	1.0
	Language switching	9.0	na	8.0	10.0	10.0	7.6
	Editorial guidelines	10.0	7.0	9.0	3.0	10.0	5.2
~Engineering	Unicode	8.0	8.0	2.0	na	3.0	4.8
	API/SDK	7.0	10.0	2.0	na	7.0	6.4
	Interoperability	2.0	4.0	9.0	na	2.0	6.6
	Standards compliance	7.0	1.0	8.0	na	9.0	6.4
	Deployment	6.0	2.0	4.0	8.0	4.0	4.8
	Test automation	5.0	9.0	9.0	1.0	7.0	6.3
	UI Extraction	9.0	9.0	10.0	3.0	10.0	4.0
	UI Layout	8.0	6.0	7.0	8.0	7.0	7.1
	UI Implementation	2.0	7.0	4.0	2.0	3.0	5.4
	Build and Deliverables	8.0	5.0	4.0	7.0	7.0	6.9
UI Modularity	6.0	9.0	4.0	1.0	6.0	4.4	

Dashboards

3.3	Globalization contact person	Globalization contact person 'expert/champion' identified in team.
3.4	Globalization checking tools	Regularly run globalization checking tools on source code and resource files.
3.5	Standard defect tracking	Globalization defects are tracked in consistent manner and assigned to related feature team.
4.0	Content	Product or user generated content is designed for international requirements
4.1	Worldwide content availability	Any user anywhere can access any globally licensed and jurisdictionally appropriate content. Easy and effective method to select and obtain content (by language/region/topic).
4.2	User generated data and moderation	Consult with legal requirements for jurisdiction based abuse, decency, privacy and data security requirements. Design and implement features to enable this content moderation and if needed blocking.
4.3	Multi-lingual content	Numerous countries officially support multiple official languages. Consider requirements for multi-lingual content library creation capability and delivery.
4.4	Content library localization process	Localizable content library data resides in localization standard file formats or efficient extraction or automation tools exist
5.0	Interaction Design	User interface must support regional specific input and output formats and market needs
5.1	Regional identity requirements	Support local ID requirements (drivers license, passport, residence number (Korea), etc.) for creating accounts or internet service.
5.2	Regional payment systems	Support necessary regional payment systems, some countries favor giro payment system rather than credit cards.
5.3	Culturally correct name and title use	Some regions place special emphasis on awareness and use of title and salutations. Others reverse order of typical first and last name, and never refer to someone by first

Unicode Conference IUC 42  38



Issues Measures Code Dashboards ▾

Lingoport Overview

Globalyzer Summary

Concatenations	12
Embedded Strings	13
Locale-Sensitive Methods	26
General Patterns	4
Static File References	0

Resource Manager Summary Repo


Default Locale:
Base Resource Files:
Base Resource Words:
Target Locales
% Complete
Last Prepped Kit Version
Last Prepped Kit Date



Unicode Conference IUC 42

Licensing

LRM License End Date-valid:
LRM Version:
LRM License Company Name: 39
LRM-Number of Projects Allowed



Hi, Mike McKenna

International Components for Unicode (ICU) Global Readiness

Global readiness grade: **B** based on overall score: **67**

Market Entry

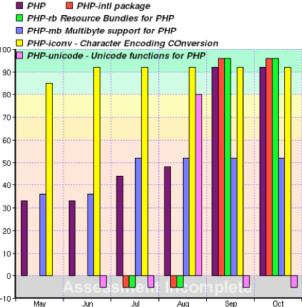
- Global UED
- Data Representation & Exchange
- Locale sensitive behavior
- Content management
- Deployment

Data Representation & Exchange

The Data Representation & Exchange capability's grade is **A** based on the score of 93. This capability is worth 52% of the total global readiness grade.


The readiness grade for this capability can improve by doing the following:

- Standards-compliant metadata tagging
- Represent languages using BCP 47 codes
- ICU uses older posix style. It needs to be updated to support BCP-47. Maybe in ICU 4.4



Assess Complete

Unicode Conference IUC 42



40

Dashboards


Grade: 81.78%
Status: Graded

Categories	Grade	Wt%	Status
Western European language support	10.00	5.00	Graded
CJK language support	10.00	5.00	Graded
Eastern European language support	9.50	5.00	Graded
Unicode Support	9.52	3.00	Graded
GB18030 certification	-	5.00	Not Applicable
Full locale awareness	4.69	5.00	Graded
Linguistic support	-	5.00	Not Applicable
Single code base	10.00	5.00	Graded
Complex Script Support	4.50	5.00	Graded
UI resources externalized	9.00	5.00	Graded
Efficient build process	10.00	5.00	Graded
Streamlined kit creation	9.00	5.00	Graded
Language expandability	10.00	5.00	Graded

Western European language support
Grade: 100%
Status: Graded

Filter Question: Do you accept input, display, import/export, copy/paste, or manipulate the navigation and editing of textual characters? yes no

Requirements	Grade	Wt%	Status
Typing characters	10.0-Full support	2.00	Graded
Keyboard switch input	Not Graded	.00	Graded
Non-ascii chars in filenames	Not applicable	.00	Graded
Keyboard shortcuts	0.0-No support	.00	Graded
Backspace and delete usage	2.5-Minor support	.00	Graded
Cursor positioning	5.0-Intermediate support	.00	Graded
Text selection	7.5-Major support	.00	Graded
Multiple click selection	10.0-Full support	2.00	Graded
Character display in UI	10.0-Full support	20.00	Graded
Printing	10.0-Full support	20.00	Graded
Interoperability	10.0-Full support	3.00	Graded
Cross-language compatibility	10.0-Full support	10.00	Graded
Text layout	10.0-Full support	5.00	Graded

Unicode Conference IUC 42

41

Dashboards


API Localization Ready, unweighted category scores

Selected project overview

Product	Release cycle	Requirements	Score 1.0	Score 3.5	Score 5	Score 7.5	Score 8	IB at 1.5	IB at 5.0 or 0
Site	P111	28	67%	20%	13%	0%	0%	67%	13%
Tool	P111	28	57%	21%	13%	0%	0%	79%	27%
Prod	P111	27	55%	14%	0%	0%	16%	70%	30%

World Ready, unweighted category scores

Unweighted score ranges

Unicode Conference IUC 42

42

Process data collection benefits

- Confident of outcome based on data
 - Meets specs (product, G11n)
 - Forecast customer acceptance, known pitfalls
- Early detection allows mid-course correction
- Reliability: Final QA confirms not corrects
- Historic data allow process improvements
 - Efficiency (Cost, Time to market, ROI)
 - Quality (Accuracy, Functional, Rejection rate)
 - Insight (robustness, automation, reuse, training)

Unicode Conference IUC 42



43

Questions



Unicode Conference IUC 42



44

References

- W3.org
 - <https://www.w3.org/standards/webdesign/i18n>
 - <https://validator.w3.org/i18n-checker/> - checks web pages for i18n correctness
 - G11n checklist
 - I18nguy – list of assessment tools
<http://www.i18nguy.com/guidelines.html>
 - Multidimensional Quality Metrics (MQM) Definition, *German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad*. “translation-oriented quality assessment systems”
<http://www.qt21.eu/mqm-definition/>
 - Lingoport WorldWare Conference
https://lingoport.com/wp-content/uploads/2012/03/I18nConference2012_I18nAssessments.pdf
 - Search terms:
 - “i18n assessment tool”
 - “Localization acceptance tests”
- Unicode Conference IUC 42



45

Category/Requirement Overall		Requirement description	average
1.0	Market Requirement Plan	Product requirements document specifies which language groups are supported by the product, and when. All market regions should be addressed, with expected release phase.	
1.1	Market Requirement Plan	The product requirements document specifies which language groups are supported by the product, special linguistic or locale features needed, which languages required for UI translation, and when. Regional Product Management and experts are consulted to verify that the functionality meets the needs of their markets. International support and features clearly indicated in product datasheet or specs to potential customers.	
1.2	Conduct market research	Conduct market research or gather data to understand market conditions.	
2.0	Dependency Management	Management of the project should consider dependencies on other products and have a clear goal for time-to-market for new markets.	
2.1	Inventory (ID and list) all dependencies	All third party products or other components that this product depends on are World Ready Enabled, localized versions available or Localization Ready; or their deficiencies are well understood, workarounds acceptable and handled appropriately (enumerate in the metadata table). Consider some webservices or social sites aren't used in certain countries (blocked or not popular).	
2.2	Common framework usage	Application uses applicable globalization and internal common platform libraries. Common framework resources are integrated unmodified or in modular fashion - new/updated strings added in separate module.	
3.0	Process	Workflow and tools are setup for international success	
3.1	Milestones defined and agreed	UI specification, globalization testing complete date, translation update frequency and/or final master delivery milestones defined and agreed upon.	
3.2	Plan for localization	Plans for tasks such as text translation, screen capture, software building, project management, post-localization functional testing (feature functionality still works), linguistic testing (correct and appropriate translations) and cosmetic testing (layout, spacing, readability, flow, truncation).	
3.3	Globalization contact person	Globalization contact person 'expert/champion' identified in team.	
3.4	Globalization checking tools	Regularly run globalization checking tools on source code and resource files.	
3.5	Standard defect tracking	Globalization defects are tracked in consistent manner and assigned to related feature team.	
4.0	Content	Product or user generated content is designed for international requirements	
4.1	Worldwide content availability	Any user anywhere can access any globally licensed and jurisdictionally appropriate content. Easy and effective method to select and obtain content (by language/region/topic).	
4.2	User generated data and moderation	Consult with legal requirements for jurisdiction based abuse, decency, privacy and data security requirements. Design and implement features to enable this content moderation and if needed blocking.	
4.3	Multi-lingual content	Numerous countries officially support multiple official languages. Consider requirements for multi-lingual content library creation capability and delivery.	
4.4	Content library localization process	Localizable content library data resides in localization standard file formats or efficient extraction or automation tools exist	
5.0	Interaction Design	User interface must support regional specific input and output formats and market needs	
5.1	Regional identity requirements	Support local ID requirements (drivers license, passport, residence number (Korea), etc.) for creating accounts or internet service.	
5.2	Regional payment systems	Support necessary regional payment systems, some countries favor giro payment system rather than credit cards.	
5.3	Culturally correct name and title use	Some regions place special emphasis on awareness and use of title and salutations. Others reverse order of typical first and last name, and never refer to someone by first name. Ensure order and use meets user expectations in email communications, welcome screens, and screen names.	
5.4	Support regional holidays	The design allows calendars and timelines to be localized for each country, since different countries observe different holidays.	

5.5	Support non-Gregorian (non-Western) calendars	Some regions expect calendars that have different years or different months from the Western Gregorian calendar. Examples are Taiwan, whose calendar started in 1911; Japanese official calendars, which are indexed from the beginning of the emperor's rule.
5.6	Timezone handling	Time and time zone is handled in intuitive and correct manner. Storage allows for including time zone Olson IDs, indicators and offsets. Support countries with multiple timezone. Keep daylight savings data separate and easy to update.
5.7	Printer and paper friendly design	Design supports ISO and ANSI paper and envelope sizes. Defaults remembered. For printer-friendly displays, the design limits the width of printable page content to 630 px. If the content needs to fit on a single page, the height is limited to 905 px.
5.8	Worldwide feature set and processes	Features and functionality meet worldwide needs. Default visibility of features may be different by user profile/location/language, however any user can access any feature worldwide
5.9	Technical constraints	Technical constraints, such as screen resolution, average page loading speed, etc., are considered in layout and feature design. Some markets don't have the bandwidth to load lots of graphics, others tolerate long load times.
5.10	Search Engine Optimization (SEO)	Plan and implement SEO compatible design. Use localized URL names to improve results and usability.

6.0 Text Processing

6.1	Character input	Country specific keyboard layouts are supported. Standard methods for international characters are supported - AltGr key, ALT+numeric keypad, dead key combinations. Asian Input Method Editors (IME) are fully integrated for in-place Asian character composition and input. Product specific hotkeys don't interfere with OS shortcuts for changing keyboard layouts or operating IME.
6.2	Text editing	Correct cursor movement, backspace, insertion and delete key operation, multiple click selection. (See comment on region field).
6.3	Text wrapping	Correct detection of word boundaries based on language of text and use of punctuation.
6.4	Text truncation	String truncation cuts off only at word or grapheme boundaries, not at code point nor byte boundaries.
6.5	Vertical layout	User choice for page layout, input and display of text in top to bottom and right to left layout.
6.6	RTL layout	User choice for left to right or right to left overall text layout
6.7	Complex shaping	Support for Arabic cursive scripts. Support for combining and stacking sequences used in Indian, Vietnamese and African languages using tone and vowel diacritics and shapes. (See comment on region field.)
6.8	Collation	Comparison and sort order of text depends on locale. Use appropriate locale for this collation.
6.9	Case conversion	Support language and locale specific rules (for example Turkish I, partial use of accents on French uppercase letters, Eszett German character, lack of case in Asian languages)
6.10	Linguistic services	Provide language specific spell checker, dictionary, grammar checking, thesaurus, text-translation services

7.0 Locale formatting

7.1	Support regional postal address formats	Components of postal addresses must be selected and arranged according to the local conventions of the country of the address.
7.2	Support regional telephone number formats	Telephone number can be input in intuitive manner, and is displayed correctly for the regional context of the number itself.
7.3	Support regional date/time formats	Support input and display of date/times in regional format. Handle timezone when needed.
7.4	Support regional number formats	Use host and/or user defined setting for number display and input (thousand separator, decimal separator changes)
7.5	Support different measurement systems	Metric and US measurement formats supported, appropriate default values for each.
7.6	Support regional punctuation and separators	Punctuation (question marks, exclamation marks) and separators (non-breaking spaces, colons, semi-colons, quotes) vary by locale.
7.7	Support different currency formats	Currency symbol and placement, negative format changes by region. Ensure currency type carried with data to avoid currency display of data switching to host setting automatically.

8.0 Search

8.1	Language-insensitive search	Search or Find features have capabilities that work effectively worldwide - similar to common 'case insensitive', feature has 'accent/diacritic insensitive', 'full width/half width insensitive', ligature support, etc.
8.2	Language aware stemming	Support language specific stemming knowledge (root word forms, gender versions, plural/singular forms)

8.3	Language aware thesaurus and spelling	Support language specific common interchangeable words, common spelling mistakes
8.4	Pattern-matching	Search patterns for international data defined and effective for international data (any character includes ideographic, numbers include full width numbers, etc.)
9.0 Language switching		
9.1	UI language selection	Users can easily select UI language from a list of installed or available languages. Server provides API for language selection. Preferred language choice passed in full detail to child applications and processes.
9.2	Default language selection	Application will use user profile information (when available) to default to user preferred language. Otherwise last used language for user, or host environment (browser or OS preferred/current language) determines initial UI language.
10.0 Editorial guidelines		
10.1	Consistent terminology	Establish and use corporate and product glossaries that highlight new and changed terminology.
10.2	Linguistic style	Style guideline established and followed. Avoid slang.
10.3	Culturally neutral terminology, examples	Use up to date country/region names and politically neutral terminology or graphics for geo-politically sensitive areas (for example 'country or region' when referring to list of country names). Use culturally neutral examples and images. Examples and explanations compatible with different international hardware and keyboards. Sensitive issues such as religion, politics, sex, alcohol, and drugs must be carefully addressed in all content.
10.4	Acronyms and abbreviations defined	Abbreviations used in one region may not be understood in another.
11.0 Unicode		
11.1	File/user names	Product supports any valid Unicode and file system allowed characters for file/folder paths installation, open/save locations. Any valid OS user/login account name supported to install and operate product.
11.2	Printing	All characters from supported language groups print properly to electronic or hardcopy formats.
11.3	Basic Unicode	European, Asian and Eastern Europe characters can be copy/pasted and displayed in the same string/document/objectname/layemame/tooltip/palette regardless of 'Language for non-Unicode programs settings' or current language or codepage of OS.
11.4	Data input and output	All data input and output streams use Unicode encoded stream by default. For non-Unicode streams the character encoding must be known, and then data converted to/from Unicode.
11.5	Data stored in Unicode (or stored with character encoding tags)	Data must be stored in Unicode. BLOB data, data that may not change (e.g., external email), and data with an uncertain encoding must be stored with encoding tags to facilitate conversion to Unicode.
11.6	Text normalization	Standard (and recommended) normalization and canonicalization algorithms are used (NFC, NFKC before toupper or tolower).
11.7	Multilingual sorting	Multilingual data, when sorted, is collated according to the Unicode Collation Algorithm (UCA).
11.8	Ensure correct Unicode metadata in Web pages	The HTTP header (HTTP Content Type) is set to UTF-8. For HTML, the meta tag http-equiv attribute charset set to UTF-8. The @charset declaration is problematic because when style sheets are concatenated, imbedded @charset declarations cause breakage.
11.9	International domain name (IDN) support	The product must support International Domain Names (IDN) and International Resource Identifiers (IRI) that may appear in feeds, user generated content, indexing, tags and email addresses. Applications/products should process, store, and display IDNs and IRIs correctly. IDNs/IRIs display to users in human readable format. Use machine readable format to access networks resources in AJAX/cURL calls.
12.0 API/SDK		
12.1	Globalized APIs	Product and Webservice APIs support Unicode strings, and are independent of locale settings, independent of localized product UI and enable 3rd party developers to create their own internationalized and localized applications. Parent applications pass locale information to child applications that have no other access to preferred locale.
13.0 Interoperability		
13.1	Character data	No file, data corruption, or data loss viewing or exchanging product files or output with mixed language strings with product on other codepage OS or data between backend systems, databases and frontend. Cut/paste of international data with other applications works fine.

13.2	Region sensitive data	Date/number and other data with regional sensitive input and output formatting stored in files or transmitted between systems or servers is not corrupted or misinterpreted, regardless of regional settings or format of input system, storage system or destination system.
13.3	User interface data	Files created on one localized product are properly interpreted on other language product. Default style names and other default UI object names display in language matching product UI.
14.0 Standards compliance		
14.1	Represent character encoding names with standard IANA names	Character encoding names in protocol and data must use MIME-preferred IANA names. Best name chosen for conversions.
14.2	Represent countries using ISO 3166 codes	Countries must be identified following standards on Country Representation, ISO-3166.
14.3	Represent languages using BCP 47 codes	Standards on Language Representation (BCP-47) must be followed when working with language tags in any context.
14.4	Represent date/time stamps using RFC 3339 (where possible)	All time data must be in RFC 3339 format when shared with or transmitted to other applications independent of storage format. Note that this format is not allowed in email and HTTP headers, and is not supported in SQL statements.
14.5	Represent time zones using Olson identifiers	Time zones must be identified using the Olson identifiers from the public TZ database.
14.6	Represent currency units using ISO 4217 codes	Currency units must be designated by the 3-letter code from ISO 4217.
14.7	Specify geo-locations using standard meta data	Exact geolocations must be specified using LAT/LON.
14.8	GB18030 support	Product is certified by CESI for use in China (product has full Unicode support and user selectable fonts).
15.0 Deployment		
15.1	Single codebase	Single codeline, source code and (if applicable) binary used for worldwide release. All bug fixes made to the same codeline.
15.2	Perform last mile testing	Perform last mile testing to ensure performance. Must meet an agreed upon SLA for page load time, latency, etc. Used experience QA labs from remote continents to test latency, load time, and perceived performance.
15.3	Support dial-up and slow connections	Users in some countries have only dial-up or shared-line access, and therefore have slow bandwidth. These users have an alternate experience that works in low-bandwidth settings.
15.4	Multilingual or multi-locale installation or deployment	Installation (for desktop) and deployment (for web-based) in multiple languages or for multiple locales is possible.
16.0 Test automation		
16.1	International test data	QA automation regularly runs with international data - accented and ideographic characters. Regional format data (number/dates) also used in test data.
16.2	Globalized test automation	QA automation is designed to work with any language OS and product.
17.0 UI Extraction		
17.1	UI is separated from code	Localizable resources (dialog controls text, error messages, etc.) are extracted from source code. All UI can be translated.
17.2	UI in standard resource format	All the localizable binary resources are stored in industry standard file formats. (such as Win32 res-only DLLs, .NET resource assemblies, .MSI, WPF .XAML files, Resource Bundles or Java property files) or A simple process or a command line tool that can be automated is made available to handle the localization of every "non standard" file types.
18.0 UI Layout		
18.1	Loc-friendly static layout	UI with no support for automatic-layout uses generous and well designed layout that needs minimum localization resizing work. No overlapping/hidden dialog controls. Ability to change tab order.
18.2	Automatic layout	UI uses automatic dialog layout which adapts (typically at run-time) to translated dialog and control contents.
18.3	Font handling	Non western european languages require a different font than US English, and minimum size is sometimes bigger than US. Ensure font and default size adapts automatically for all required languages or is extracted for change. Ensure layout adapts for different height and width fonts. Web pages using named fonts have fallbacks to serif or sanserif.
18.4	Highlighting and emphasis configurability	Highlighting and emphasis are configurable so they can be appropriate for the target languages.
18.5	Color configurability	The design is flexible as to how colors are used and where. For example, red means "lucky" in China.
18.6	Embedded fields in sentences	The design avoids embedded images, fields, and controls within sentence structures. Sentence structure varies by language and there is no guarantee that embedded elements will maintain their position in the translated version.

18.7 UI Mirroring enabled

In order to support Arabic and Hebrew localized UI, product recognizes UI languages that require mirroring and sets Window, custom control and bitmap Right To Left (RTL) styles as appropriate. The text fields and text areas within the UI function as right-to-left text. For web pages, the text direction is properly specified, tags and controls are in place, and other adjustments are made to enable presentation of bidirectional elements. Phone numbers are displayed left to right in RTL layout. Fields for phone numbers should default left to right. Universal symbols and icons, such as the 'Play' button for media, do not change direction based on context. They should remain as they are. The cursor in text fields and text areas should default to the right (have a RTL-context). The functionality of next/previous buttons should be switched for a RTL-context; their icons can often be reused.

19.0 UI Implementation

19.1	No must-match or delicate strings	All strings in resource files may be translated. Different translation of identical source strings is acceptable. Code allows strings with spaces, quotes, etc. even if English string did not contain these. A different order of translated strings in a sorted combobox is permitted.
19.2	String formatting and concatenation	Sentence fragments (just beginning or ending of strings and other sentence parts concatenated to others) and strings with multiple variables but no numbering of arguments do not exist. Account for sentences in other languages that have additional words that change and must match for gender and plurals.
19.3	Styles are externalized	Styling is separate from code. Use CSS or cultural context for styling, colors, fonts. Instructions for translators or PMs on how to change the styles are provided.
19.4	Stable IDs	Resource IDs are stable between builds or versions.
19.5	Use unique strings	Identical strings that appear in different contexts have different instances of the string available for translation.
19.6	Globalized images	Images are culturally neutral (not using symbols, colors, gestures, attire specific to a culture). Wherever possible, design image without text that needs to be translated. When text is required, provided layered image format (text in separate layer) to localization team.
19.7	Consistent light source for RTL	In graphics, such as buttons, light sources, shade angles, and gradients should be kept the same as those in a left-to-right context.
19.8	Text files encoding	Text files needing localizing are saved in Unicode (UTF-8 or UTF-16) format with byte order mark (BOM) for clarity.
19.9	Ordered lists	Ensure applications function with sorted comboboxes and lists. Ensure browser apps UI logic can sort multiple choice elements such as pull down menus or other elements required to be ordered.
19.10	Remove obsolete/unused resources	All obsolete resources or resources files are deleted from product.

20.0 Build and Deliverables

20.1	Efficient localized build/engineering	Portable and minimal size localization kit. Possible for localization to work without access to source code files. UI to localize available in UI/resource only files. English is just another language and can serve as a template for directory/file organization for other languages of UI.
20.2	Global service pack	No localization engineering work necessary for Service Pack. Service Pack releases are global and apply to any language product. Only the SP readme needs to be localized.
20.3	Context information	Supporting files to give symbolic ids, comments, and context to translator are regularly delivered with English builds. File formats that include symbolic ids and comments (WPF/XAML) in native format do not need extra files.
20.4	Pseudo-translated build	Pseudo-translated build regularly created as part of build process for feature and system testing.

21.0 UI Modularity

21.1	UI files grouped into language folder	Global and localized files are split in distinct language specific directories in build and install structure.
21.2	Install downloaded pack	Method exists to install additional language for product using downloaded language pack. Core product doesn't need change to support new language UI request.

Copyright 2011 Leandro Reis, Mike McKenna, Paul-Henri Araud. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of

conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

THIS SOFTWARE IS PROVIDED BY Leandro Reis, Mike McKenna, Paul-Henri Araud "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL Leandro Reis, Mike McKenna, Paul-Henri Araud OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

The views and conclusions contained in the software and documentation are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of Leandro Reis, Mike McKenna, Paul-Henri Araud.
