



Global Information Systems: Localization and Internationalization (5)

Prof. Dr. Jan M. Pawlowski
Autumn 2013



1934 2009

UNIVERSITY OF JYVÄSKYLÄ



Contents

- Introduction
- Definitions and Terms
- Design approaches
- Summary



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

The Open Unified Process – Disciplines

- Architecture
 - Architecture Notebook
- Configuration and Change Management
- Development
 - Design
 - Build
 - Developer Test
 - Implementation
- Project Management
 - Iteration Plan
 - **Project Plan**
 - Work Items List
 - **Risk List**
- Requirements
 - Supporting Requirements Specification
 - Vision
 - Use Case
 - Glossary
 - Use-Case Model
- Test
 - Test Case
 - Test Log
 - Test Script
- Roles
- *Artefacts / Support*

[Source: <http://www.epfwiki.net/wikis/openup/>]



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Samples

تتقري زجلا

Startseite (M) Feeds (J) Drucken Seite

ملخص الخبر

أعلن البنتاغون الأمريكي حالة التعبئة في صفوف جنوده للمساعدة في جهود إطفاء الحرائق الضخمة التي اندلعت في ولاية كاليفورنيا. وقد تسببت تلك الحرائق في مقتل خمسة أشخاص وإصابة 45 آخرين.

الأخبار الفضائية المعرفة الأعمال English البث الحي

روابط أخرى

الأربعاء 1428/10/12 هـ - الموافق 2007/10/24 م (آخر تحديث) الساعة 18:44 (مكة المكرمة)، 15:44 (غرينتش)

فصائل من متطرفي دارفور تصل إلى ليبيا للمشاركة في مؤتمر سرت وأخر

إغلاق كربلاء أمام الزوار ومصرع وإصابة العشرات ببغداد

فرضت السلطات العراقية حظرا على دخول كربلاء لغير سكان المدينة، في إطار سعيها لتهنئة الأوضاع هناك بعد مقتل وإصابة عشرين شخصا بينهم 14 من الشرطة خلال الأيام الماضية. وفيما لا زال التوتر يسيطر على البصرة، قتل ثمانية بانفجارين ببغداد صباح اليوم.

إصابة إسرائيلي بالضفة وتوصية بقطع الكهرباء عن غزة

أصيب إسرائيلي بعد أن أطلق مسلحون في سيارة النار عليه بينما كان يقف في موقف لحافلات بمستوطنة أرييل بالضفة الغربية، من جهة أخرى أوصى مسؤولون إسرائيليون بالبدء بقطع التيار الكهربائي عن قطاع غزة تنفيذا لقرار يعتبره كيانا معاديا.

العراق يلغي حصانة الشركات الأجنبية

قررت الحكومة العراقية إلغاء حصانة الشركات الأجنبية العاملة في البلاد والتي أصدرتها سلطة الائتلاف المؤقتة عام 2004. يأتي هذا

أزمة دارفور

الكردي.. دروب التاريخ الوعرة

هذه الصفحة برعاية

رمضان كريم


Fertig

Internet 100%

Samples

Al Jazeera English - Front Page

Startseite (M) Feeds (3) Drucken Seite



ALJAZEERA.NET

WATCH NOW
FRONT PAGE

- AFRICA
- AMERICAS
- ASIA-PACIFIC
- CENTRAL/S. ASIA
- EUROPE
- MIDDLE EAST


FOCUS
BUSINESS
SPORT
PROGRAMMES
WEATHER
YOUR VIEWS


SEARCH
ABOUT US
ARABIC
DOCUMENTARY

RSS
LOG IN


UPDATED ON:
WEDNESDAY, OCTOBER 24, 2007
19:04 MECCA TIME, 16:04 GMT

For every child
Health, Education, Equality, Protection
ADVANCE HUMANITY







Turkey 'hits PKK targets in Iraq'
Military sources say warplanes and troops attacked positions across the border.




Nato urged to up Afghan commitment



Israel set to sever power to Gaza




Sri Lanka furore over LTTE corpses



Roadside bombs hit Baghdad suburb


Now:
News
Next (in 10 min):
Witness


ALJAZEERA
TODAY'S SCHEDULE



ALJAZEERA Job Vacancies


IN VIDEO

Delhi's menacing monkeys 


Turtles under threat in Malaysia 

POLL

Who should act to stop Kurdish attacks in Turkey?



California fires force mass exodus
Nearly a million people abandon their homes as president declares an emergency.



BLACKWATER

Iraq clamps down on security firms
New law to revoke immunity of foreign security contractors from prosecution.

Fertig

Internet 100%

Definitions

- ❏ Internationalization (I18N) is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for redesign. Internationalization takes place at the level of program design and document development (W3C, 2007)
- ❏ Localization (L10N) is the process of taking a product and making it linguistically and culturally appropriate to a given target locale (country/region and language) where it will be used (W3C, 2007)

Definitions

- Globalization (G11N) defines a business strategy and business activities to act on a global market.
- A Locale is a geographic location and a language of a region (e.g., Germany, French-speaking Quebec, Central Finland) – classes based on a locale are locale-sensitive



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Types of internationalization

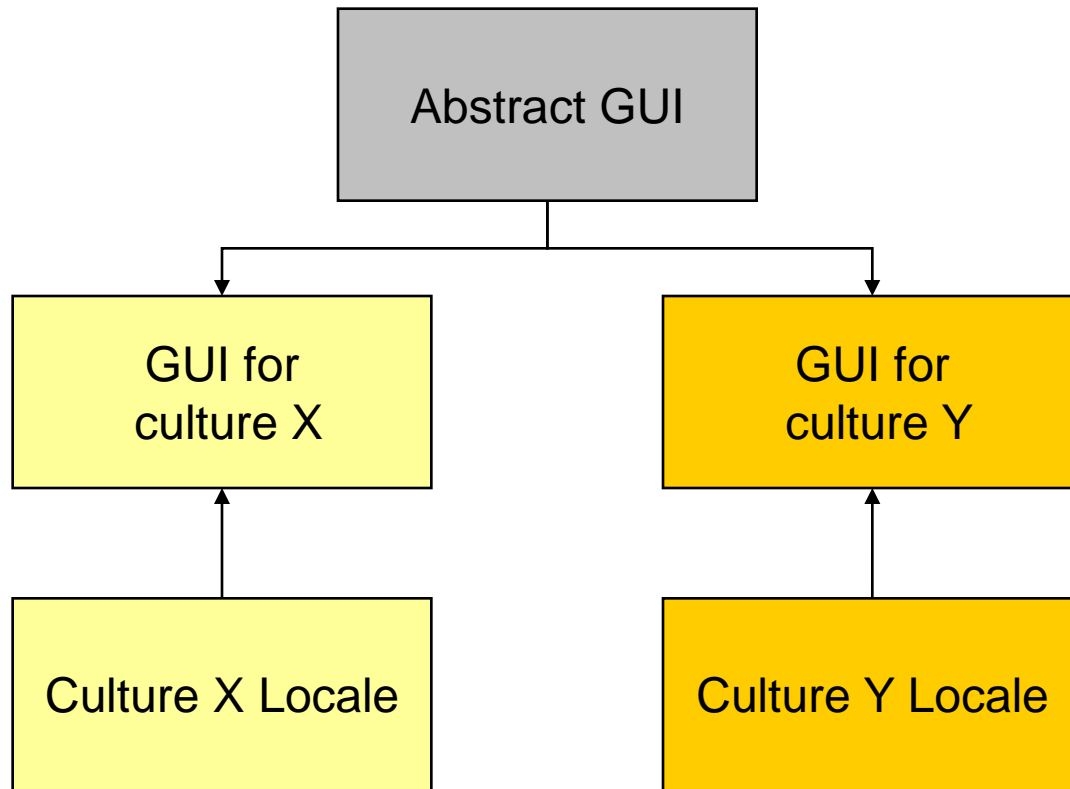
- Application development (business logic)
- User interface design (presentation logic)
- Time
 - Run-time
 - Compile-time
 - Design-time
- Aspects
 - Software
 - Documentation (process documentation, help, manual)
 - Web pages
 - Learning materials
 - Knowledge & experiences



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Types of internationalization



[Adapted from Kersten, 2002]



1934 2009

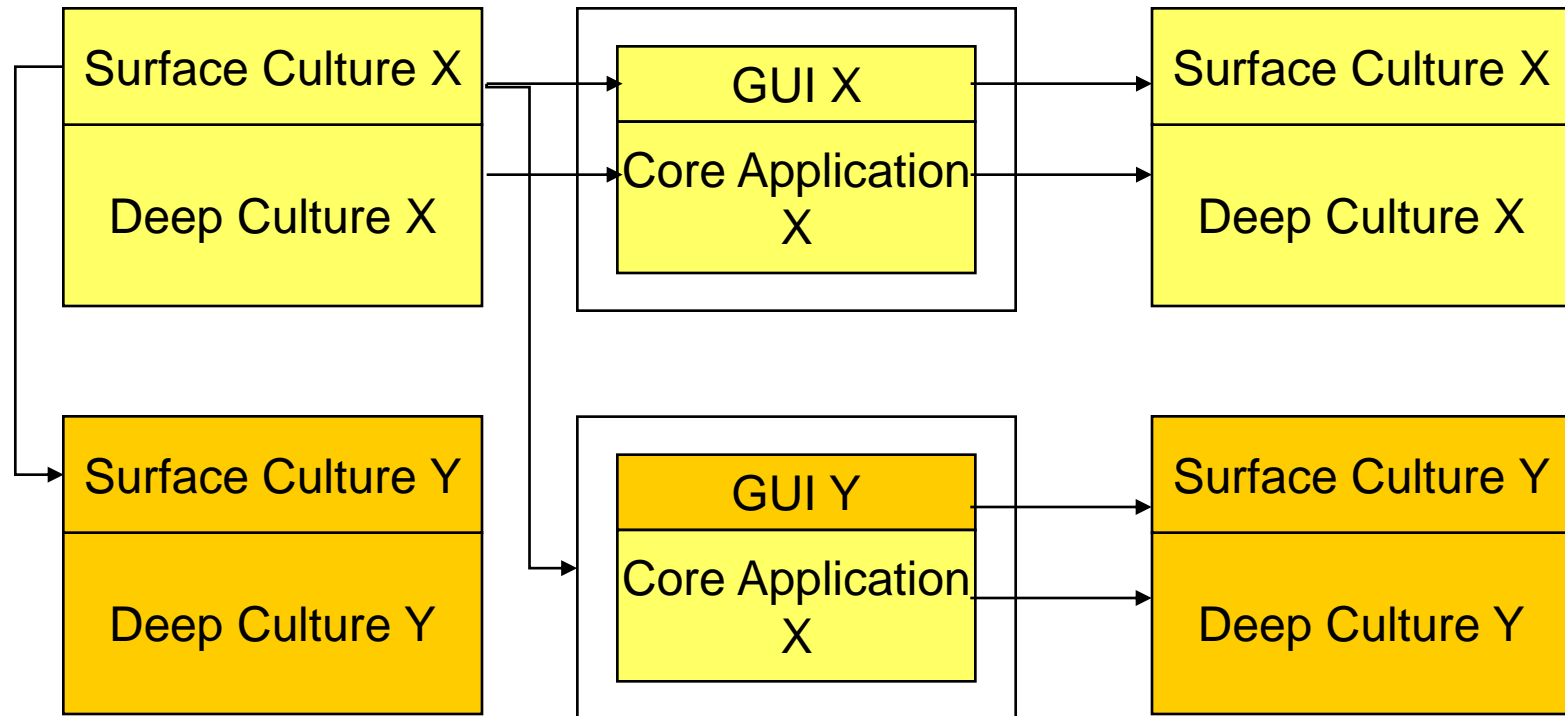
UNIVERSITY OF JYVÄSKYLÄ

Types of internationalization

Production

Product

Deployment



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Challenges in Localization

- Text string expansion
- Character sets and encoding
- Bidirectional text and vertical display
- Keyboard character layout, shortcuts
- Fonts
- Sorting order
- Placeholders
- Abbreviations
- Terminology
- And many more



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Aspects

- Formats
 - Date
 - Time
 - Currency
 - Addresses, Postal codes
- Symbols, icons, graphics, colors
- Language
 - Translation
 - Writing system
 - Characters
- Other
 - Contents...
 - Sounds
 - Messages
 - Measurements / Units



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Format samples

📅 Dates:

- 31.10.2007, 13:15:26 CET
- 10-31-2007, 01.15.26 am CET
- 31 OCT 2007, 13 h 15 CET
- ...

📊 Numbers




- 1 234 567,89
- 1.234.567,89
- 1,234,567,89

📅 Additionally: Other calendars, holidays

📅 Separate representation and presentation – using identifiers, string indexing

Localization by country

ISO 3166 Country Codes

Official country names used by the ISO 3166/MA	Numeric	Alpha-3	Alpha-2	Local ISO codes
 Afghanistan	004	AFG	AF	ISO 3166-2:AF
 Åland Islands	248	ALA	AX	ISO 3166-2:AX
 Albania	008	ALB	AL	ISO 3166-2:AL
 Algeria	012	DZA	DZ	ISO 3166-2:DZ
 American Samoa	016	ASM	AS	ISO 3166-2:AS
 Andorra	020	AND	AD	ISO 3166-2:AD
 Angola	024	AGO	AO	ISO 3166-2:AO
 Anguilla	660	AIA	AI	ISO 3166-2:AI
 Antarctica	010	ATA	AQ	ISO 3166-2:AQ
 Antigua and Barbuda	028	ATG	AG	ISO 3166-2:AG
 Argentina	032	ARG	AR	ISO 3166-2:AR
 Armenia	051	ARM	AM	ISO 3166-2:AM
 Aruba	533	ABW	AW	ISO 3166-2:AW

[Source: http://en.wikipedia.org/wiki/ISO_3166-1]



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Localization by language

ISO 639 Language Codes

am		Mapudungun; Mapuche	mapudungun; mapuche; mapuce
arp		Arapaho	arapaho
art		Artificial (Other)	artificielles, autres langues
arw		Arawak	arawak
asm	as	Assamese	assamais
ast		Asturian; Bable	asturien; bable
ath		Athapaskan languages	athapascanes, langues
aus		Australian languages	australiennes, langues
ava	av	Avaric	avar
ave	ae	Avestan	avestique
awa		Awadhi	awadhi
aym	ay	Aymara	aymara
aze	az	Azerbaijani	azéri
bad		Banda languages	banda, langues
bai		Bamileke languages	bamilékés, langues

Source: http://www.loc.gov/standards/iso639-2/php/code_list.php



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

But....the example of Khmer...

- Written from left to right, characters being placed also above and below the main line of writing
- Words are not separated by spaces. A space in Khmer is a punctuation sign similar to a comma

ជំរិតគឺជាផ្នែកមួយនៃព្រេងនិទានរបស់ខ្មែរយើង

- A word is composed of clusters, syllemes. They are not a proper syllable, as syllables are a unit of consonants and vowels pronounced in one stroke of breath. Consonants pronounced after a vowel are part of the syllable, but not part of the cluster or sylleme.

ជំរិត

- Source: <http://sourceforge.net/projects/khmer/>



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Formats

- ❏ **Unicode** is a universal character set, ie. a standard that defines, in one place, all the characters needed for writing the majority of living languages in use on computers. It aims to be, and to a large extent already is, a superset of all other character sets that have been encoded.
- ❏ A **coded character set** is a set of characters for which a unique number has been assigned to each character. Units of a coded character set are known as **code points**. For example, the code point for the letter à in the Unicode coded character set is 225 in decimal, or E1 in hexadecimal notation. (Note that hexadecimal notation is commonly used for identifying such characters, and will be used here.)
- ❏ The **character encoding** reflects the way these abstract characters are mapped to bytes for manipulation in a computer. (W3C, 2007)



Formats

- ❏ **Character:** The smallest component of written language that has semantic value; refers to the abstract meaning and/or shape (Unicode Glossary, 2007)
- ❏ **Visual rendering** introduces the notion of a glyph. Glyphs are defined by ISO/IEC 9541-1 [ISO/IEC 9541-1] as "a recognizable abstract graphic symbol which is independent of a specific design". There is not a one-to-one correspondence between characters and glyphs. (W3C, 2005)



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Formats: Recommendations (W3C, 2005)

- ❏ Specifications, software and content **MUST NOT** require or depend on a one-to-one correspondence between characters and the sounds of a language
- ❏ Specifications, software and content **MUST NOT** require or depend on a one-to-one mapping between characters and units of displayed text
- ❏ Protocols, data formats and APIs **MUST** store, interchange or process text data in logical order
- ❏ Independent of whether some implementation uses logical selection or visual selection, characters selected **MUST** be kept in logical order in storage
- ❏ Specifications of protocols and APIs that involve selection of ranges **SHOULD** provide for discontinuous logical selections, at least to the extent necessary to support implementation of visual selection on screen on top of those protocols and APIs



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Formats: Recommendations (W3C, 2005)

- ❏ Specifications and software **MUST NOT** require nor depend on a single keystroke resulting in a single character, nor that a single character be input with a single keystroke (even with modifiers), nor that keyboards are the same all over the world
- ❏ Software that sorts or searches text for users **SHOULD** do so on the basis of appropriate collation units and ordering rules for the relevant language and/or application
- ❏ Specifications, software and content **MUST NOT** require or depend on a one-to-one relationship between characters and units of physical storage
- ❏ More on characters and encoding:
<http://www.w3.org/TR/charmod>
- ❏



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Formats

- Different encodings for character sets

- ISO 8859-1

- Unicode

	A	κ	好	𠄎
Code point	U+0041	U+05D0	U+597D	U+233B4
UTF-8	41	D7 90	E5 A5 BD	F0 A3 8E B4
UTF-16	00 41	05 D0	59 7D	D8 4C DF B4
UTF-32	00 00 00 41	00 00 05 D0	00 00 59 7D	00 02 33 B4



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Recommendation samples

- ❏ Internationalisation Tag Set (W3C)
 - Used to develop localizable schemata
 - Identifying translation needs
 - Elements: Translate, localization note, terminology, directionality, language information, elements within text



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Recommendation samples

Internationalisation Tag Set (W3C)

```
<Res
  xmlns:its="http://www.w3.org/2005/11/its"
  its:version="1.0">
  <prolog
    its:translate="no">
    <revision>Sep-07-2006</revision>
    <its:rules version="1.0">
      <its:translateRule selector="//msg/notes" translate="no"/>
      <its:locNoteRule locNoteType="description" selector="//msg/data">
        <its:locNote>The variable {0} is the name of the host.</its:locNote>
      </its:locNoteRule>
    </its:rules>
  </prolog>
  <body>
    <msg id="HostNotFound">
      <data>Host {0} cannot be found.</data>
    </msg>
    <msg id="HostDisconnected">
      <data>The connection with {0} has been lost.</data>
    </msg>
    <msg id="FileNotFound">
      <data
        its:locNote="{0} is a filename">{0} not found.</data>
      </msg>
  </body>
</Res>
```

[Source: <http://www.w3.org/TR/2007/REC-its-20070403>]



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

“Culturalization” of applications

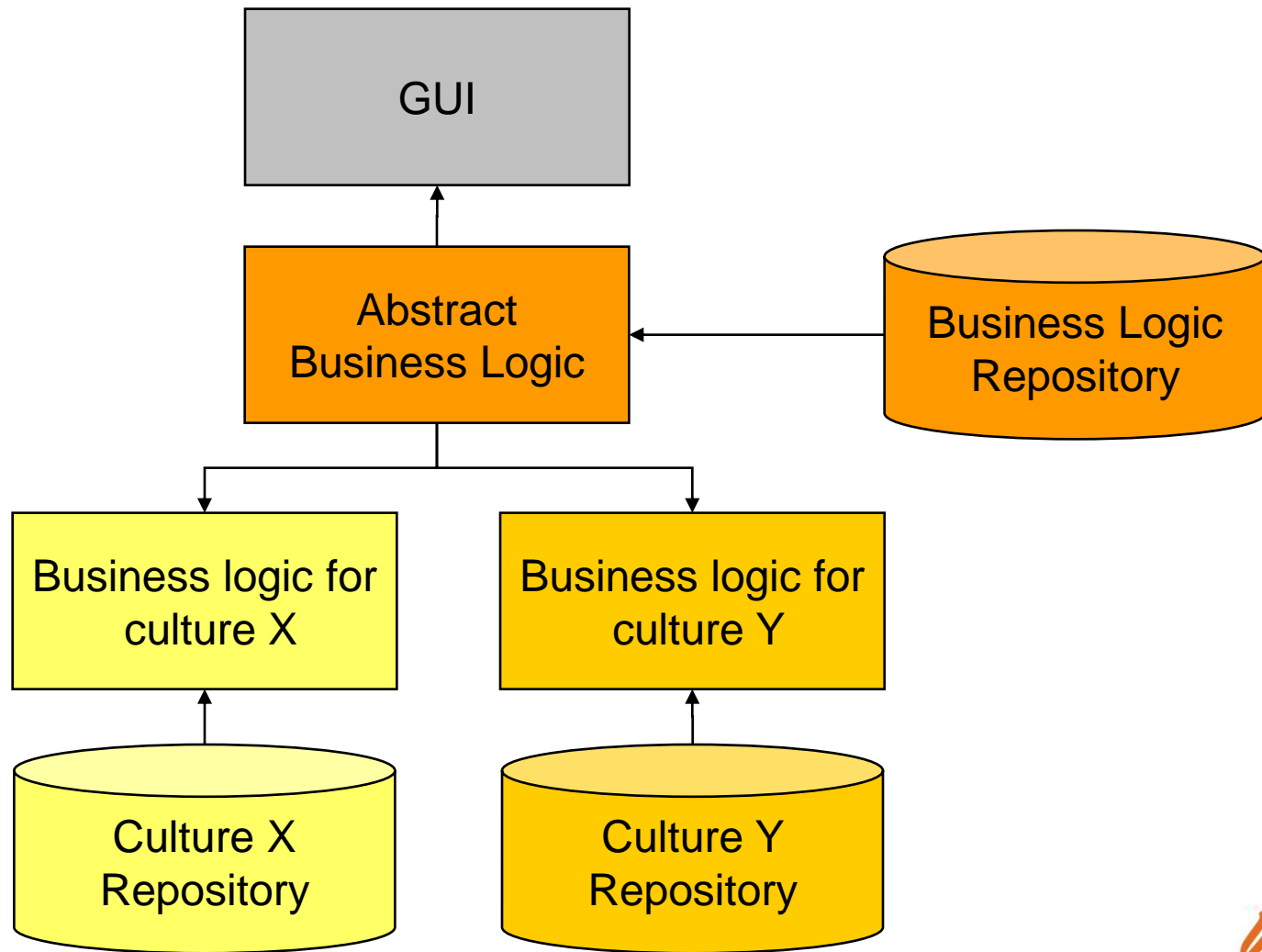
- Culture awareness
 - Adapting business logic
 - Adapting contents
 - Adapting user interfaces
-
- Samples for culturally adapted interfaces



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Types of internationalization



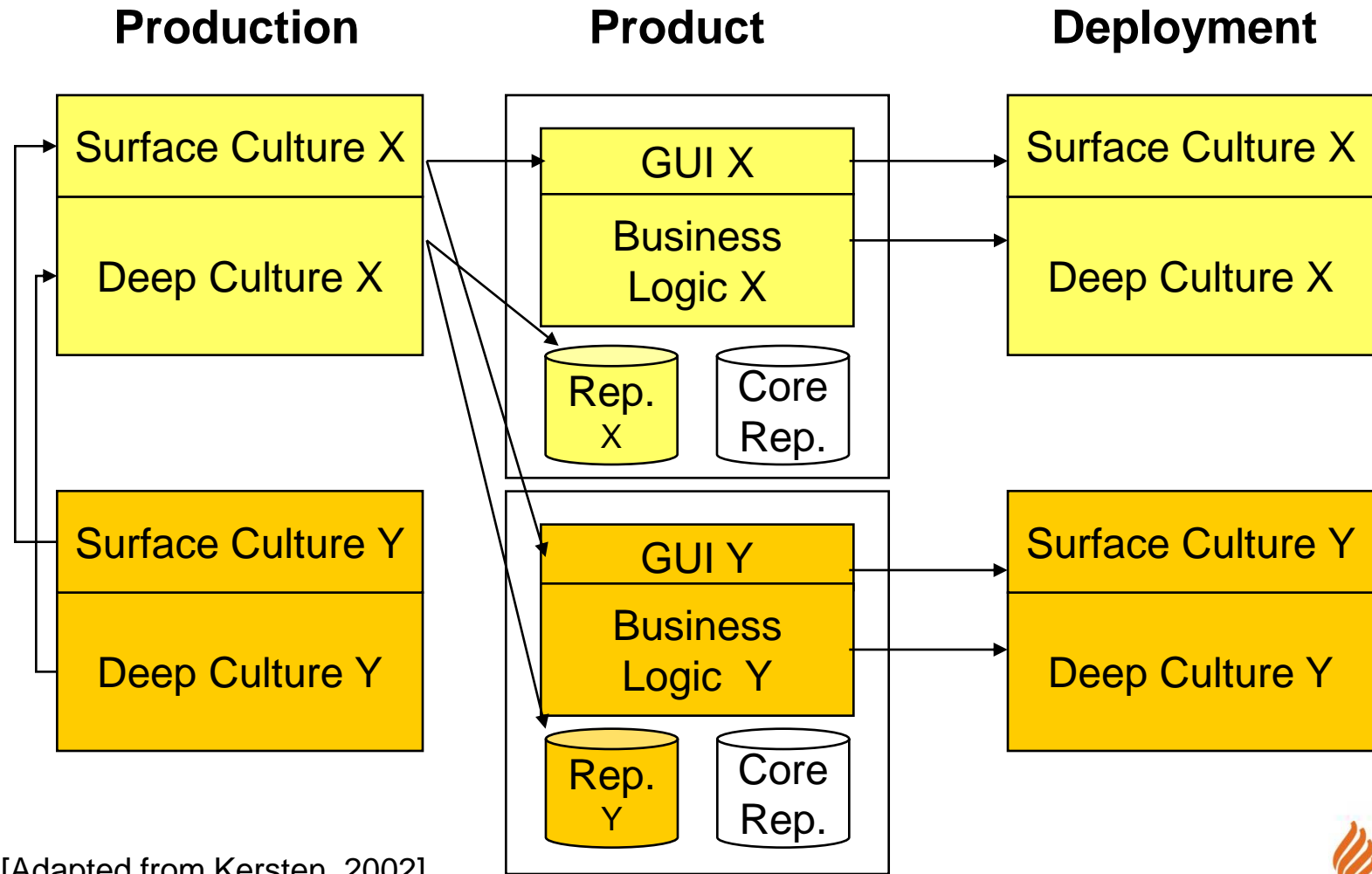
[Adapted from Kersten, 2002]



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Culture-aware internationalization



[Adapted from Kersten, 2002]



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

At the end of this phase, the following results should be ready:

- ▣ Strategy for internationalization & localization
 - Design planning
 - Architecture refinement
 - Standards, guidelines



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Summary

- ❏ There is no one-fits-all strategy for internationalization and localization
- ❏ Standards should be considered
- ❏ Based on a culture analysis, (internal) guidelines should be developed
- ❏ Prototyping and participating is essential
- ❏ Other individualization / personalization strategies should be considered



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Questions

- Describe the differences of globalization, internationalization, localization and adaptation.
- Which aspects should be considered when designing and developing international solutions
- Which guidelines can be applied for designing a website for a Finnish university?
- Which steps are necessary to develop an Asian marketing site for JYU?

References

- ❏ Hogan, J.M., Ho-Stuart, C., Pham, B. (2003): Current Issues in Software Internationalisation. Australian Computer Science Conference, Adelaide, May2003.
- ❏ Kersten, G.E., Kersten, M., Rokaowski, W.M. (2002): Software and Culture: Beyond the Internationalization of the Interface, Journal of Global Information Management, 10(4), 2002.



1934 2009

UNIVERSITY OF JYVÄSKYLÄ

Contact Information ITRI

▣ **Prof. Dr. Jan M. Pawlowski**

▣ **jan.pawlowski@titu.jyu.fi**

▣ **Skype: jan_m_pawlowski**

▣ **Office:**

▣ **Telephone +358 14 260 2596**

▣ **Fax +358 14 260 2544**

▣ **<http://users.jyu.fi/~japawlow>**



1934 2009

UNIVERSITY OF JYVÄSKYLÄ