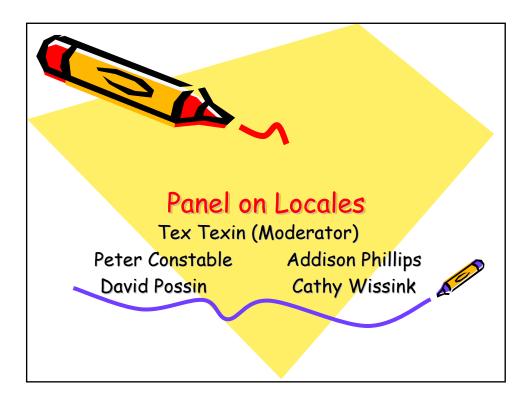
#### Panel on Locales



#### Panel on Locales

- Locale is a mechanism used in Web, Java and other technologies
- Many developers use locales and presume sound I18n behavior results
- · How well does it work in practice?



Panel on Locales IUC22 San Jose, California

#### Panelists Views





- · Cathy Wissink, Microsoft
- · Dave Possin, Welocalize.com
- · Peter Constable, SIL
- Audience Participation
- Tex Texin, XenCraft



Panel on Locales IUC22 San Jose, California

3

#### Problems with Locale?

- Locales solve many problems, but expectations may be too high:
  - -Desire for "magic I18N functionality"
  - -Failure to properly factor data structures.
  - -Lack of understanding of what a locale is.
  - -"Hubris" in solving internationalization problems.



Panel on Locales IUC22 San Jose, California

### Problems with Locale?

- Locales solve many problems, but expectations may be too high:
  - -Missing information can be added on a platform-by-platform basis.
  - -There are a lot of locale models...
    - Solve specific problems with each one instead of blanket rejection.



Panel on Locales IUC22 San Jose, California

5

## There are some problems...

- Interoperability and integration are the major problem!
  - -Internet lacks a "globalization architecture"
  - -Platforms and Environments differ on:
    - · Range of locales
    - Locale data sets
    - Locale standards
    - Specific behavior



Panel on Locales IUC22 San Jose, California

#### There are some problems...

- Interoperability and integration are the major problem!
  - -No mechanism for exchanging locale preferences.
  - -Limited mechanisms for exchanging language preferences.
  - Confusion about difference between locale and language.



Panel on Locales IUC22 San Jose, California

\_

#### Solutions

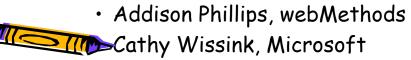
- Add a platform neutral locale tagging and negotiation system.
  - -E.g. See my proposal: *ULocales: Building Global Enterprise Web Services*
  - www.webmethods.com/whitepaper\_select/1,1332,58277,00.html
- Separate language from locale.
  - -Overloading of tags (for example, Accept-Language, xml:lang)
- Provide a locales metadata repository.

🕶 (Is this actually useful?)

Panel on Locales IUC22 San Jose, California

#### Panelists Views





- Dave Possin, Welocalize.com
- · Peter Constable, SIL
- Audience Participation
- Tex Texin, XenCraft



Panel on Locales IUC22 San Jose, California

9

### LCIDs (Locale IDs)

- Very old means of identifying locale (in use since at least Windows 3.1)
- Used on all subsequent versions of Windows (including Windows XP)
- Uses combination of language and sublanguage (= country/region), with sort ID and reserved value in high word (low word alone = language ID)



Panel on Locales IUC22 San Jose, California

#### Problems with LCIDs

- Difficult to decouple (can't separate language out from country/region; too tightly bound)
- Don't work well with multiple country regions ("Latin America")
- Don't take writing systems into account (Azeri, Kazak, Serbian)
- Difficult to change in response to changes in geopolitical situation (Serbo-Croatian → Serbian and Croatian; what does old LCID identify?)
- · Non-customizable
- Not standards conformant (proprietary format)



Panel on Locales IUC22 San Jose, California

1

## Response to LCIDs deficiencies

- Incorporated RFC 3066 codes into .NET Framework to identify locales
- Addresses issues of standards conformance, locale decoupling, GPS concerns
- Not perfect solution, either...
  - Must rely on ISO standards (slow to change compared to industry, not always complete data)
  - Doesn't solve multiple country region problem
  - Not customizable
  - RFC codes don't address writing systems or sort IDs

Panel on Locales IUC22 San Jose, California

#### Panelists Views



- Addison Phillips, webMethods
- Cathy Wissink, Microsoft
  Dave Possin, Welocalize.com
  - · Peter Constable, SIL
  - Audience Participation
  - Tex Texin, XenCraft



Panel on Locales IUC22 San Jose, California

1:

### Locales vs. User Preferences



- Abstract Concept to Identify the Location and Language for Computing Purposes ...
  - · Of the User?
  - Of the Computer?
- ... Based on Incomplete Standards

- User Preferences
  - A User has Preferences of how Data is presented
  - A User Speaks a Preferred Language
  - A User has more Preferences than the Locale offers



Panel on Locales IUC22 San Jose, California

## Why do we need a new standard locale model?



- 1. Computing demands for an improved model for new application areas.
- 2. Users demand a better representation of their culture, their region, and their language.
- 3. Mechanics are needed so both can negotiate their needs and find a satisfying compromise.



Panel on Locales IUC22 San Jose, California

15

#### What is in the Pot?



- Different Operating Systems are creating new incompatible models
- Web Software is stuffing all into one bucket: lang
- Standards Organizations are moving too slow and are missing the train
- The Users are the Losers



Panel on Locales IUC22 San Jose, California

#### What can We do?



- Create New Models
- · Prove their Usability
- Promote them to Organizations
- Integrate them into i18n Libraries
- Turn them into de-facto Standards
- Let Organizations adapt them



Panel on Locales IUC22 San Jose, California

17

#### Panelists Views



- Addison Phillips, webMethods
- · Cathy Wissink, Microsoft
- Dave Possin, Welocalize.com



- Audience Participation
- Tex Texin, XenCraft



Panel on Locales IUC22 San Jose, California

#### Language and locale

- Locales involve numerous parameters, which can be divided into two broad categories:
  - non-linguistic
    - currency symbol, decimal separator, postal address format, time zone, etc.
  - linguistic
    - language, script, orthography, sub-language variants (e.g. vocabulary)



Panel on Locales IUC22 San Jose, California

19

### Language and locale

- ·Linguistic parameters are relevant for locales
  - E.g. UI element: menu item 'file open'
    - presentation in UI affected by choices for language, script, orthography, regional vocabulary preferences, etc.



Panel on Locales IUC22 San Jose, California

#### Language and locale

- Linguistic parameters partially (in)dependent of one another
  - e.g. language vs. script
    - in principle, any given language can be written in any script
    - In practice, only certain combinations are likely to occur



Panel on Locales IUC22 San Jose, California

2

# Linguistic parameters and locale naming

- ·Locale identification: language-country?
  - Writing system (script) and orthography distinctions should not be confused with language distinctions
    - Azerbaijani (Latin) vs. Azerbaijani (Cyrillic): two writing systems, one language
    - English (American spelling) vs. English (UK spelling): two orthographies, one language
  - Cannot interpret "language" portion as conflation of all linguistic parameters

Panel on Locales IUC22 San Jose, California

# Linguistic parameters and locale naming

- ·Locale identification: language-country?
  - Country is not a good correlate for all linguistic parameters
    - may often work for orthography
    - not always a good correlate for script, vocabulary
      - Azerbaijani, Latin vs. Cyrillic: same language, same country
  - Linguistic parameters may correlate with a domain of usage other than country

Panel on Locales IUC22 San Jose, California

23

# Linguistic parameters and locale naming

- •Locale identification: language-country?
  - In principle, use of language and country not sufficient as identifying correlates for linguistic parameters.



Panel on Locales IUC22 San Jose, California

## Language codes in locale IDs

- •Some implementations use ISO 639-1 to designate *language* element in locale IDs
  - Inadequate!
    - · two-letter codes: not enough possible
    - ISO 639-1 has strict requirements
  - Implementers should start now in looking beyond ISO 639-1
    - ISO 639-2, RFC 3066, Ethnologue.



Panel on Locales IUC22 San Jose, California

25

#### Locale vs. language data

- Need to distinguish locale from linguistic attributes of user's data:
  - locale: collection of user / UI preferences
    - locale may imply some particular language / script / etc. for linguistic UI resources
  - independent of the language data users work with
    - regardless of UI pref's, user should be able to work with data in any language, script, etc.
    - buser pref's may determine *default* language, etc. of user's data, but no more

Panel on Locales IUC22 San Jose, California

#### Locale vs. language data

- Need to distinguish locale from linguistic attributes of data:
  - locale infrastructure for UI should be independent of infrastructure for supporting multilingual data
    - · e.g. input methods
      - calling an input method an "input locale" isn't a problem
      - issue is separation of support for user / UI preferences vs. multilingual data



Panel on Locales IUC22 San Jose, California

27

#### Panelists Views



- Addison Phillips, webMethods
- Cathy Wissink, Microsoft
- Dave Possin, Welocalize.com
- Peter Constable, STL



Tex Texin, XenCraft



Panel on Locales IUC22 San Jose, California