Using Oracle Clinical and RDC with Multibyte Languages without deploying the Oracle NLS option

Presented by Sunil G. Singh of DBMS Consulting

Acknowledgements

- Thanks to OCUG and the OCMU focus group for this opportunity to speak and present, and their infinite patience in accepting and reviewing this presentation.
- My sincere thanks to Academia Sinica in Taiwan, and all of the members of Data Coordinating Group there, including Jeffery Tsai, Jim Chen and Dr. Cathy Fann whose input was invaluable.
- My sincere thanks to Martin van Lanen at Akzo Nobel whose kind insights also made this presentation possible.
- Thanks to the audience members for attending.

Goals

- Examine a specific use of Multibyte character use in OC and RDC 4.5 Classic Mode where the OPA NLS option was not installed.
- Observe working and non-working functionality in this type of environment
- Discuss potential validation implications
- Examine the implications of extended European Characters in UTF8 databases as a de-facto partial NLS deployment
- Suggest some "partial" deployment options for OPA NLS in the future.

Current Prerequisites for Deploying OPA NLS Option

- Requires UTF8 RDBMS Characterset
- Currently requires UNIX backend running Japanese OS HP-UX or Sun Solaris
- Currently requires Windows 2000 Middle Tier running Japanese OS for OPA NLS application with OPA on the same Middle Tier
- Two functional requirements drive the requirements for a Japanese OS on the UNIX Server:
  - Batch Data Load
  - DX View Generation/SAS Dataset output

The need for other Multibyte Languages in OPA

- Currently, the primary emphasis of the OPA NLS option focused on the use of Japanese in OPA.
- This emphasis is driven by customer demand and market forces, and there are several OPA customers in Japan using the NLS option.
- However, other customers in the greater APAC region also have requirements to use multibyte charactersets in OPA.
- Specifically, OPA customers in Taiwan, Singapore and India have all expressed interest in using OPA in languages other than English or Japanese.

The need for other Multibyte Languages in OPA

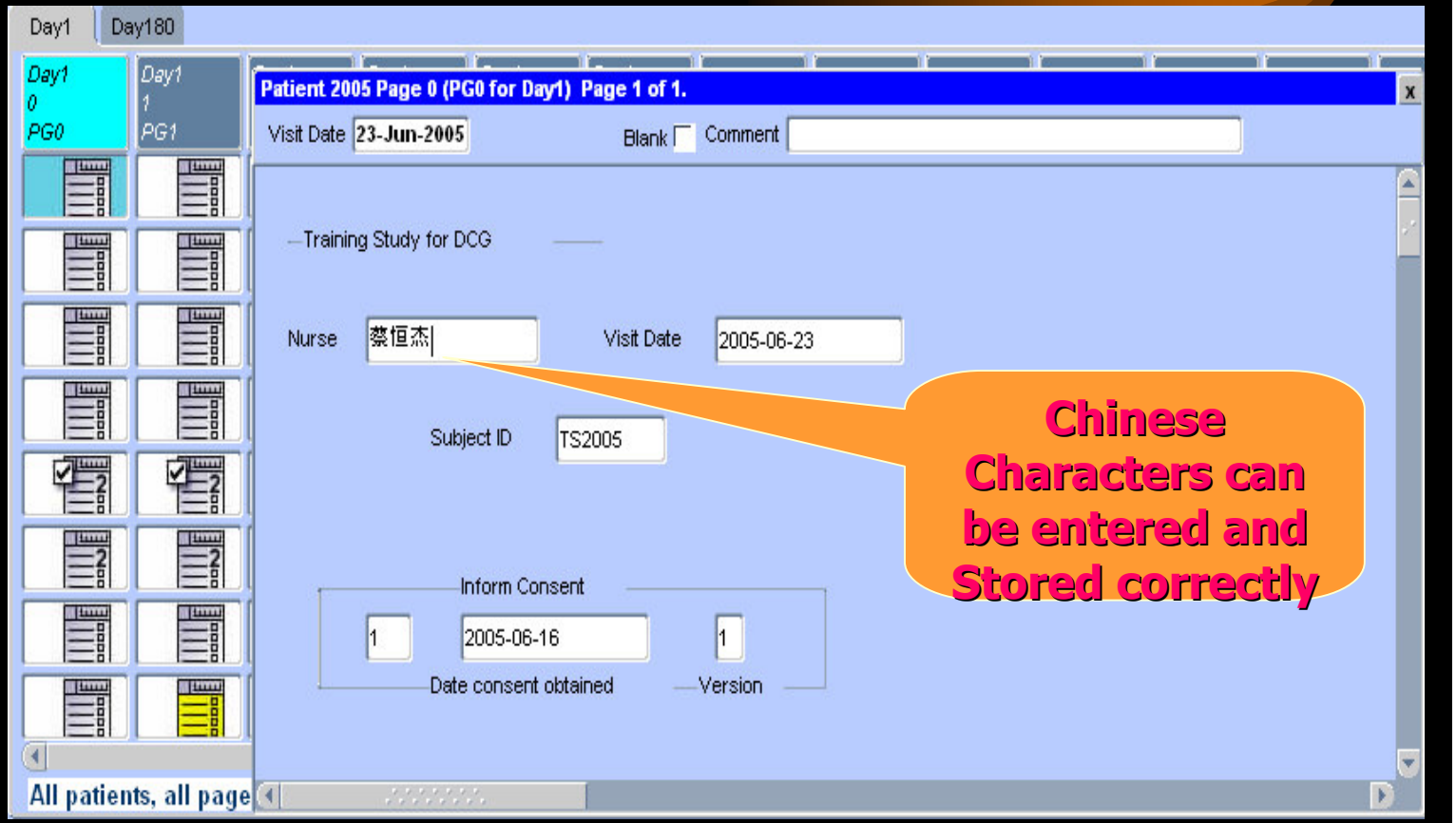
- There is an enormous growth potential for OPA in the markets with Chinese-based multibyte languages, such as Taiwan, Singapore, and mainland China.
- In these regions, translation to English may not always possible before data is entered into Oracle Clinical. Additionally, using RDC requires that some type of Data Entry in a multibyte language is required.
- Furthermore, if OPA is deployed in an RDBMS using UTF8 characterset, then the European (accented) characters are also multi-byte. This means that languages treated previously as single byte character sets must now be treated as partial multibyte languages.

Case Study: Using Chinese Characters without NLS

- Environment:
  - Windows 2000 Middle Tier and RDBMS Server running Chinese OS Windows
  - English and Chinese OS PCs
- Goal: run OC and RDC Classic Mode and accept both English and Chinese characters without deploying the OPA NLS option

Multibyte Data Entry without OPA NLS option installed

- Data Entry is possible in Chinese without having to install the entire NLS option



Multibyte Default Repeats Displayed without OPA NLS option installed

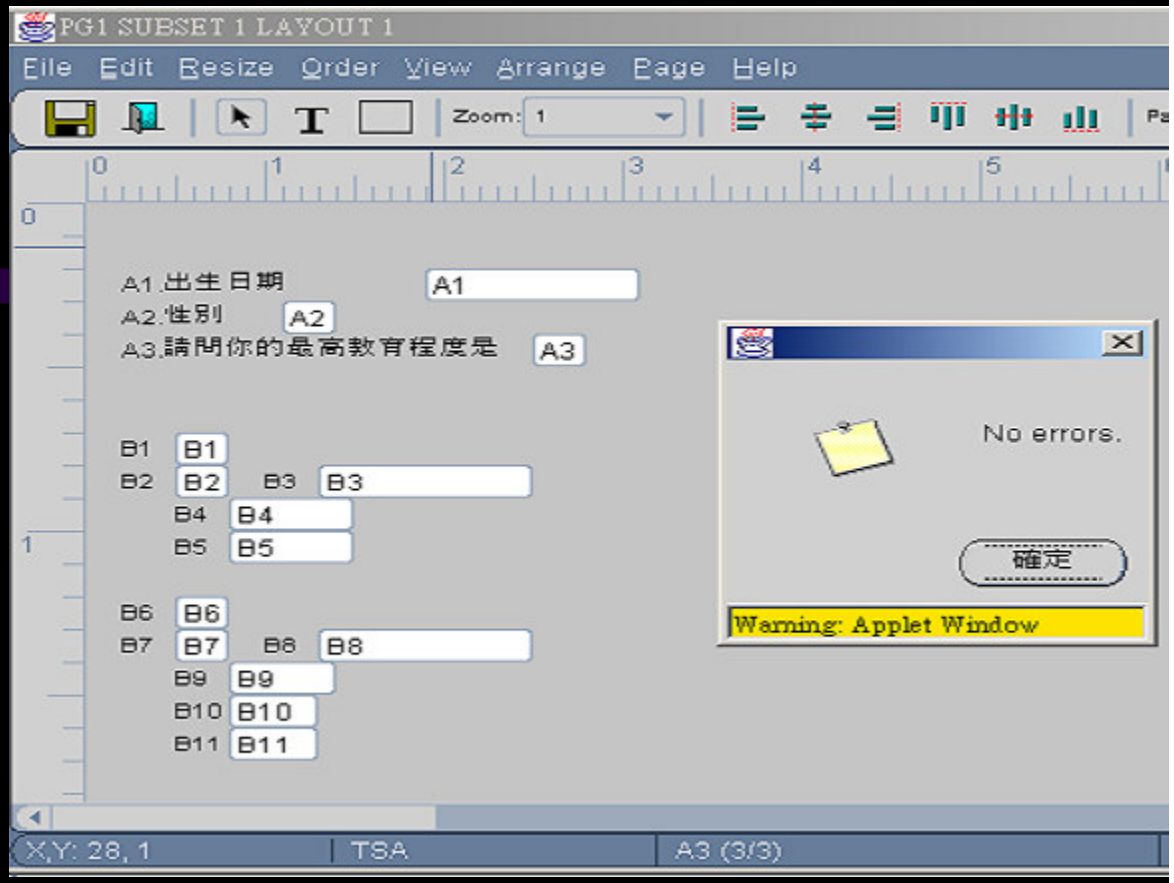
- Default Repeats can be displayed in Chinese in OC and RDC 4.5.0 (Classic Mode)

Simple Layouts can be created with Chinese Prompts

- A Character layout can be saved with prompts in Chinese only if
  - 1. It is the first DCM page
  - 2. Some specific Chinese characters are entered (not all characters seem to work)
  - 3. The number of all Chinese characters is less than 20.
- Complex Layouts can be created, but require the Layout to be generated in English first, with back-end updates and translations performed on the RXC.DCM_LAYOUT_TEXT table to allow the display and Data Entry in these layouts.

Character Layouts in Chinese

- Character Layouts appear to validate correctly in OC (see upper right), but do not save due to the additional bytes required by multibyte characters (see lower right).



Comments on other OC and RDC functions

- These functions were initially tested but produced errors. However, there may be some workarounds available which I am not currently aware of:
  - Batch Data Load with Chinese characters
  - Graphical PDF Mode Layouts in OC
  - RDC Graphical PDF Mode Data Entry
    - Chinese Characters appear as periods
    - Chinese Characters are re-displayed are then redisplayed incorrectly
- Data Extracts in SAS **do** work, as the OS of the SAS Server is also in Chinese.

Support and Validation using Multibyte data without the NLS option

- Any backend updates to create or extend the current OC functionality would not be supported by OPA and therefore, would cause a validation issue.
- However, while installation of OPA on a Middle Tier and RDBMS server which is not running in either English or Japanese is not officially supported by OPA, it is possible to test and prove that the core components of OC and Classic Mode RDC to work on servers installed in other languages.

NLS_LANG Settings

- NLS_LANG must be set to AMERICAN_AMERICA.UTF8 to view previously entered multibyte data and also to allow OC/RDC to start correctly without generating ORA-01843 “invalid month” errors.
- File %ORACLE_HOME%\806\FORMS60\SERVER\opa45.env on the Middle Tier MUST be updated to contain
  - NLS_LANG=AMERICAN_AMERICA.UTF8
- as the default for non-NLS setting is WE8ISO8859P1, which will override all registry key settings in the Oracle Forms environment and cause failures in the display of any previously entered multi-byte characters.

Consideration of European Languages as "Partial" Multibyte Charactersets

- UTF8 is now becoming the characterset of choice in new OPA 4.5.x implementations, and is required for some products such as Oracle AERS
- However, accented European Characters which were one byte in WE8ISO8859P1 now occupy 2 bytes in a UTF8 database.
- Also, by default, Windows uses the characterset WE8MSWIN1252, which can generate characters not compatible with UTF8.

Consideration of European Languages as "Partial" Multibyte Charactersets (2)

- This leads to some potential issues with the two parts of OC which are very OS dependent
  - Batch Data Load
  - DX View Creation/Creation of SAS Datasets
- Since European characters with accents will now use two bytes instead of one byte, and since the source of the files used for Batch Load can come from a different characterset, two types of problems can occur:
  - Data which is expected to "fit" into Oracle Clinical can no longer "fit" because the actual number of bytes used is more than the column size (such as 200 characters for value_text)
  - Data can be incorrectly converted (usually as an upside down question mark) if the source file contains a character incompatible with the UTF8 characterset.

Consideration of European Languages as "Partial" Multibyte Charactersets (3)

- For both BDL and DX View/SAS Dataset creation, there must be consistency with the characterset used by the OS as well as the database
- Since HP-UX, Sun Solaris, and Windows all support UTF8 as a characterset, the OS can easily be set to use UTF8.
- The programs which generate the Batch Load files (for example, SQL*Plus or MS-Excel) must have their environment set to use UTF8. Otherwise, dynamic (incorrect) conversion or truncation of characters can result.

Consideration of European Languages as "Partial" Multibyte Charactersets (4)

- Similarily, SAS must also be configured to use UTF8, otherwise the same dynamic (incorrect) conversion or truncation of characters could result.
- So a "chain of consistency" in handling European characters where a UTF8 OC database is in use must be established to avoid illegal conversions and truncated characters, which are the same issues that also must be considered when using Chinese or "traditional" multibyte language charactersets.

Implications of More Common Multibyte characters in UTF8 databases

- Since many more OPA customers now potentially will face the same types of NLS-related considerations when using European characters in a UTF8 database, the issue of how multibyte characters are handled in an OPA environment is **no longer isolated** to OPA customers using Chinese or Japanese languages.

Implications of More Common Multibyte characters in UTF8 databases (2)

- Since it is possible for existing OPA customers to perform BDL and DX View/SAS Dataset creation today in UTF8 database environments with English-based database server OS using UTF8 charactersets, then it must be true that both of these core OC application functions, which are heavily dependent on the OS, can work without having to have a database OS in running in a specific multibyte language (such as Chinese or Japanese).

Suggestions for a Supported "Partial" NLS option

- Provide an enhancement to Oracle Clinical to double the size of all data-related character fields, as was done previously for Oracle AERS
- Provide explicitly written and documented support for customers who would like to use multibyte languages throughout some additional parts of OC, such as the layout editor, BDL and DX View/SAS Extract, without making the purchase or deployment of the entire NLS option a prerequisite, or without requiring a language specific OS on the RDBMS server
- Extend the current NLS option support to other multibyte languages with growing demand, such as Chinese.

Conclusions

- Use of other languages within OPA will become more prevalent as OPA's use continues to grow globally.
- Some multibyte language functionality on OPA, such as data entry can work without the entire NLS option.
- BDL and DX Views/SAS Extracts might work if all OS and tool environments are set to use UTF8 and if all characters in the data are UTF8 compatible.
- It may be possible to provide users of OPA a "partial" NLS option, where some functionality works for multibyte characters within OPA, without having to install and deploy the entire NLS component of OPA.

Additional Questions ?

- E-mail: singh@clinicalserver.com
- Call: US 001-860-983-5848
- Electronic copies will be posted on the OCUG Intranets Site and www.clinicalserver.com