

### Founder's Archive Digitalization Solution: Preserving Human History One Character at a Time



*Throughout the very rich history of China, many ancient texts and scrolls have been passed down for generations. However these artifacts are not in the best of shape. Therefore to preserve the original artifacts and at the same time, share them with the world, the Chinese National Library wishes to have them digitized. Founder is honored to be engaged to undertake the very important task in preserving the history of China for future generations.*

#### Description

The project was aimed to digitize half a million ancient regional history chronicle scrolls. Project requirements include:

- ❑ Complete chronicles' digitization, preserving the contents in digital form.
- ❑ Digitization of text uses Song and Kai fonts. Han character code uses GB18030-2005 and interchangeable with UniCodes5.0 standard.
- ❑ For special characters that do not exist in GB18030-2005 standard, limited substitution should be made and substitution table should be created.
- ❑ For characters that cannot be substituted, a standardized character creation process should be used.
- ❑ The digitized version of the materials should reflect as much as possible the composition of the original scrolls. The composition process should be standardized.
- ❑ Digitized output would be stored in XML and PDF format. Tool would be provided for the XML to PDF conversion.
- ❑ Ability to substitute and create characters in batch process.
- ❑ Error rate of character digitization should be below 0.03%.

#### Solution Highlights

Targeting the characteristics and requirements of the digitization of national library's chronicles scrolls, an integrated operating and management system is developed. Aiming to increase digitization process efficiency, the key process employs artificial intelligence technology for automatic analysis of material's category and composition. The scroll composition is automatically restored based on the result of the analysis.

Digitization system has following features:

- ❑ Analysis tool targeting ancient text scanning enables the automatic recording of the composition information.
- ❑ System uses Founder Big Font and GB18030-2005 standard. It also has UniCode5.0 conversion tools and employs WuBi and Founder character entry method. It uses WuBi entry method for regular characters while Founder entry method is used for special characters. This realizes the combination of efficiency and functionality.
- ❑ Employment of automatic comparison and horizontal and vertical proof readings to enforce high quality. System lowers error rate while ensuring high operating efficiency.
- ❑ Batch replacement of special and substitute characters to allow continual improvement of quality.
- ❑ Automatic page composition restoration based on analysis result largely increases operation efficiency.
- ❑ An integrated operating system manages everything from front end processing to end result output ensures overall system efficiency.



#### FOUNDER'S SUCCESS FACTORS

- Founder's extensive experience in publishing and document management
- Capable technical skills in developing effective and powerful applications and tools for paper-based document digitalization
- Disciplined and efficient management of the archive conversion team



Partner

#### Project Snapshot

##### Industry

- Government

##### Client Location

- Beijing/Suzhou, China

##### Services

- Archive Digitization

##### Implementation Details

- 18 months
- Team of 70-85

##### Technologies

- .NET
- C# and C++