Digital Preservation at INIS

Dobie Savić
Head, Nuclear Information Section (NIS)
Contents

1. INIS overview
2. INIS activities
3. Digitization
   - INIS microfiche collection
   - Print IAEA publications
   - New incoming NCL
4. Lessons learned
INIS Overview

INIS
- One of the leading worldwide reference collections of scientific literature on the peaceful uses of nuclear science and technology
- Established in 1970 in response to the IAEA’s mandate to “foster the exchange of scientific and technical information on peaceful uses of atomic energy”
- IAEA/NIS: INIS, Library & SDSG

International Cooperation
- Operated by the IAEA in collaboration with 128 Member States and 24 international organizations

INIS Collection
- Over 3.4 million bibliographic records
- 320,000 full text documents. In total 462,000 NCL
- Comprising of scientific and technical reports, conference proceedings, patents and theses not available through commercial channels
- Range of subjects: nuclear engineering and technology, nuclear safety and radiation protection, safeguards and non-proliferation, nuclear and isotopic techniques, sciences relevant to nuclear research, nuclear and high energy physics, radiation chemistry, legal aspects of nuclear power, environmental and economic aspects of nuclear and non-nuclear energy sources
- Free, open and unrestricted access is available at: www.iaea.org/inis using the INIS Collection Search, based on the Google search technology
INIS Activities

INIS Collection
- Web (online) and CD/DVD (offline) access

Building and maintaining IAEA-INIS/ETDE Thesaurus
- ar de fr en es ja ru zh

**Digitizing and preserving non-conventional nuclear literature**
- Over 462,000 NCL
- including full text reports from Member States
- IAEA publications and policy documents are available in electronic format

Capacity building and outreach
- Establishment and improvement of a national information infrastructure in Member States
- Staff training programmes for national INIS centres
Digitization of INIS Microfiche Collection

- Since its creation in 1970 until 1996, INIS collected and converted to microfiche over 312,000 non-conventional literature (NCL) reports received from member states and international organizations.

- The microfiche collection contains over 1 million items, with an estimated total of 25 million pages of full-texts.

- In 1997, the INIS Secretariat replaced the microfiche-based production system with an imaging system to process and to disseminate all NCL documents in electronic format.
Digitization of the INIS NCL Collection on Microfiche started in 2003
- Close to 80% of the microfiche collection digitized
- 3 million pages still remain on microfiche

- Part of the digitization done internally
- Large part or digitization outsourced, while internal quality control maintained

- Each digitization effort was preceded by efforts to avoid duplication
- Close collaboration with INIS Member States

<table>
<thead>
<tr>
<th>Year</th>
<th>PDF</th>
<th>Pages</th>
<th>Size (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>566</td>
<td>49,574</td>
<td>3.7</td>
</tr>
<tr>
<td>2004</td>
<td>19,962</td>
<td>1,325,217</td>
<td>36.5</td>
</tr>
<tr>
<td>2005</td>
<td>36,935</td>
<td>1,577,365</td>
<td>32.1</td>
</tr>
<tr>
<td>2006</td>
<td>23,163</td>
<td>1,367,637</td>
<td>33.3</td>
</tr>
<tr>
<td>2007</td>
<td>9,313</td>
<td>668,769</td>
<td>16.3</td>
</tr>
<tr>
<td>2008</td>
<td>25,675</td>
<td>1,228,057</td>
<td>29.7</td>
</tr>
<tr>
<td>2009</td>
<td>81,221</td>
<td>3,939,811</td>
<td>77.3</td>
</tr>
<tr>
<td>2010</td>
<td>33,881</td>
<td>1,969,110</td>
<td>45.9</td>
</tr>
<tr>
<td>2011</td>
<td>24,027</td>
<td>511,990</td>
<td>16.2</td>
</tr>
<tr>
<td>2012</td>
<td>20,434</td>
<td>843,579</td>
<td>40.7</td>
</tr>
<tr>
<td>Total:</td>
<td>275,177</td>
<td>13,481,109</td>
<td>331.8</td>
</tr>
</tbody>
</table>
Digitization of INIS Microfiche Collection (cont.)

**INIS Image creation Process**

- **Benchmarking** (probability for success, digitizing requirements)
- **Source material types** (printed, microfiche, drawings and formulas)
- **Preparation** (physical, structural, conditions/characteristics)
- **Scanning** (capture mode/bitonal/grayscale/colour, optical resolution 300-600 dpi bitonal, 200-300 dpi colour/greyscale)
- **Quality control** (accuracy and completeness, data integrity, metadata correctness, ‘checksum’ for authenticity and integrity of digital files)
- **Image enhancement** (despeckling, deskewing, noise reduction, black border removal; colour and tone adjustment)
- **File formats** (TIFF Group IV, PDF, PDF/A)
- **Compression** (JBIG2 for black & white web optimization, JPEG for color)
- **File naming convention** (required consistency throughout the Collection)
- **Optical character recognition (OCR)** (ABBYY FineReader and Adobe Acrobat)
- **Storage** (online, offline Blu-ray discs, and offsite storage)
- **Preservation planning** (migration, emulation, refreshing)
- **Metadata creation** (technical metadata and bibliographic metadata)
Digitization of INIS Microfiche Collection (cont.)

**INIS digitization software**
- Techsoft PixEdit v.7.11.18 (advanced image editing)
- ABBYY FineReader 11 Corporate Edition (OCR)
- Adobe Acrobat X Professional (conversion to PDF/A, OCR of Chinese Simplified, Japanese and Korean)
- Kofax Virtual ReScan (VRS) + Kodak Perfect Page (image enhancement)

**INIS digitization hardware**

<table>
<thead>
<tr>
<th>Scanner</th>
<th>Type</th>
<th>Paper size</th>
<th>Resolution (dpi)</th>
<th>Bit-in-depth</th>
<th>Speed (A4, 200 dpi)</th>
<th>ADF Page capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fujitsu fi-5750c</td>
<td>Colour; ADF/flatbed</td>
<td>A8 – A3 Up to 34 inches</td>
<td>50 to 600</td>
<td>24</td>
<td>110 p/min (simplex) 55 p/min (duplex)</td>
<td>200 p</td>
</tr>
<tr>
<td>with VRS Pro Kodak i1440</td>
<td>Colour; ADF/flatbed</td>
<td>A5 – A3 Up to 34 inches</td>
<td>50 to 600 Optical Resolution 300</td>
<td>16 48</td>
<td>50 p/min (simplex) 100 p/min (duplex)</td>
<td>150 p</td>
</tr>
<tr>
<td>SunRise 2000</td>
<td>Microfiche scanner</td>
<td>A0-A4 reductions 7x-50x</td>
<td>CCD 3600-8800 True Resolution</td>
<td>Up to 2500 frames/hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SunRise Apollo</td>
<td>Microfiche scanner</td>
<td>A0-A4 reductions 7x-50x</td>
<td>Up to 3600 frames/hr</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Digitization of Printed IAEA Publications

Completed projects
• IAEA Bulletin
• IAEA General Conference
• INDC Technical documentation
• IAEA Technical Reports Series (TRS)

On-going projects
• Out of Print IAEA Publications
  • Proceedings Series
  • Safety Series (IAEA/NS)
• Photo Archival Project
Digitization of New Incoming NCL

- NCL coming from 50 countries or organizations
- Full-texts control
  - Input mostly in PDF => Quality control
  - NCL input in hard copy => Full digitization

**Digitization** = Scanning + Image Enhancement + OCR + PDF output + QC
Lessons learned

- Serious planning
- Substantial funds
- Staff well qualified
- Standards’ awareness
- Searchable metadata
- Well defined purpose
Thank you!