the national archives

The Open Source Evaluation Project at the National Archives

Jo Pettitt

24 June 2003

Stance

- All new IT systems in government departments must be e-GIF compliant
- E-GMS compliant as well
- Assumes XML wherever (in any layer) practical
- Government wants departments to investigate OSS wherever practical

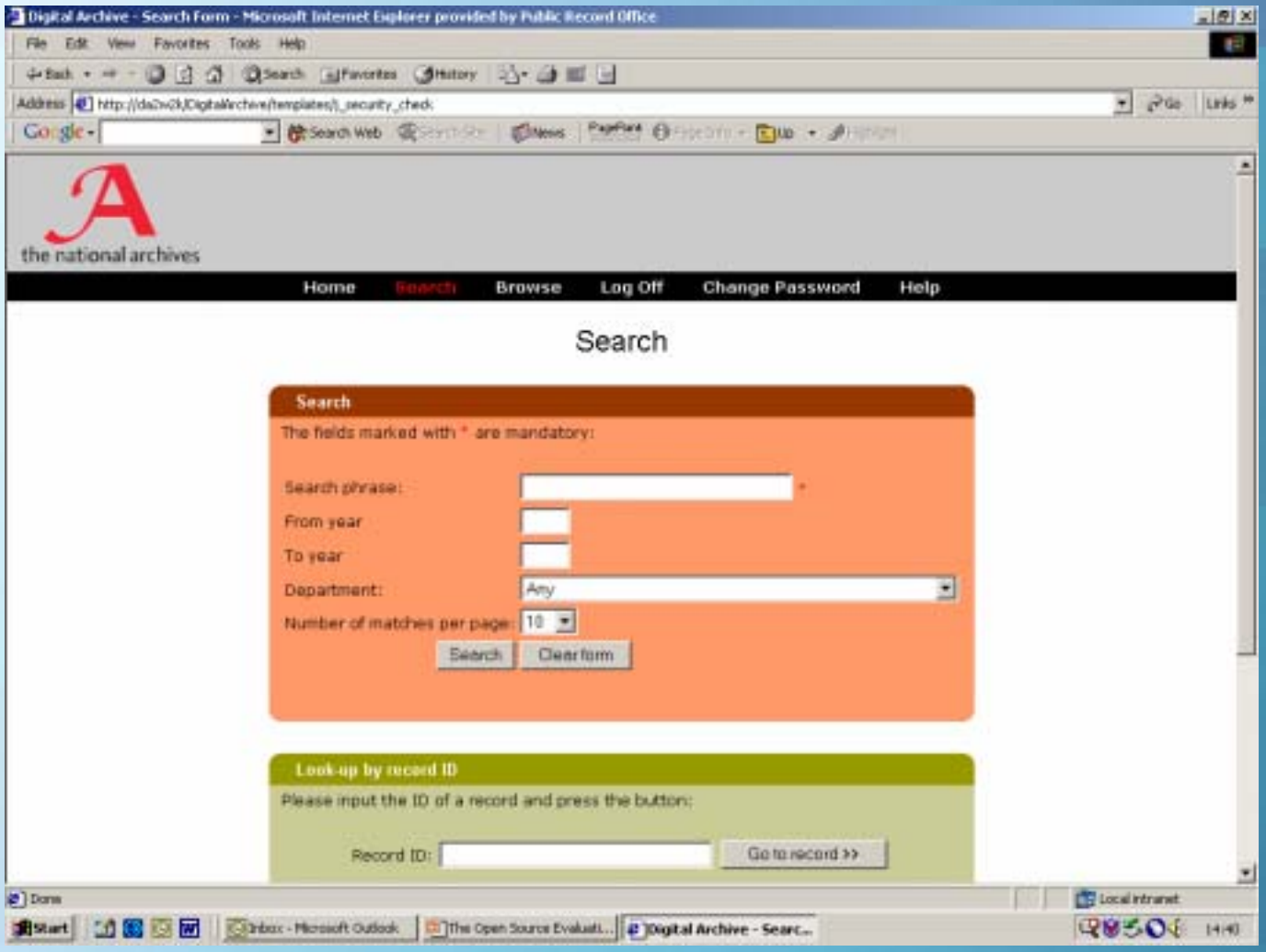
Archivists Proactive Role

- Archivists need to anticipate technology dependencies for future transfers
- Need to influence (if possible) technologies adopted by creators so that…
- Costs controlled and speed of availability supported

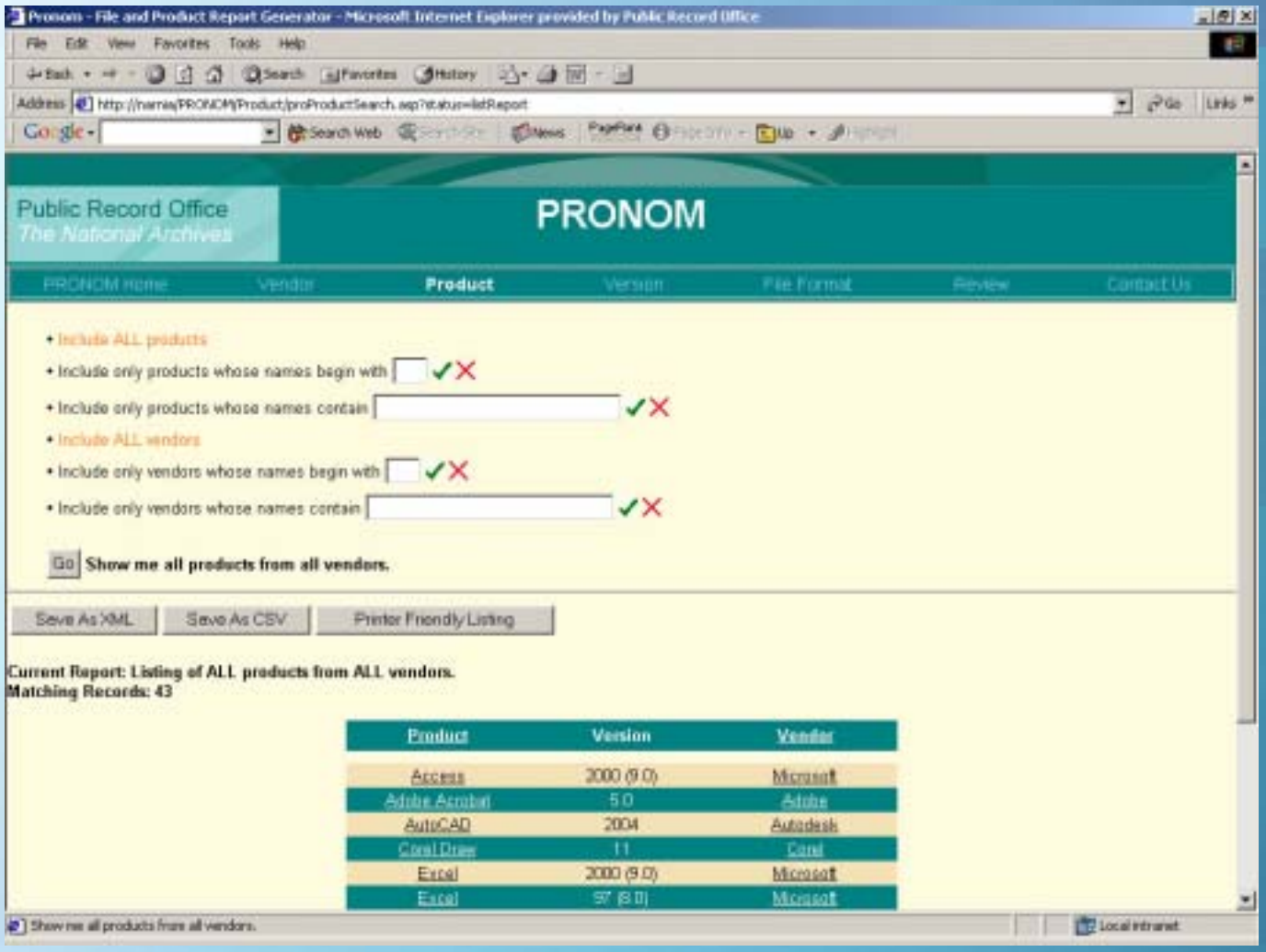
Our Approach To Digital Preservation

- Practical approach taken
  - Digital Archive built
  - PRONOM released internally
- Both are e-GIF compliant
- Both use XML
- And both have thrown up issues relating to file format longevity

Archive Search Screen



PRONOM Search Screen



PRONOM XML Schema

```
<?xml version="1.0" encoding="UTF-16" ?>
- <PRONOM-Report xmlns:rs="urn:schemas-microsoft-com:rowset" xmlns:z="#RowsetSchema">
  - <report_product_detail>
    - <Product>
        <ProductID>20</ProductID>
        <ProductName>Excel</ProductName>
        <VendorID>1</VendorID>
        <VendorName>Microsoft</VendorName>
        <Version>2000 (9.0)</Version>
        <SupportName>Microsoft</SupportName>
        <SupportID>1</SupportID>
        <ThirdPartySupportID />
        <ThirdPartySupportName />
        <DefaultExtensionID>132</DefaultExtensionID>
        <DefaultExtension>xls</DefaultExtension>
        <FileTypeID>32</FileTypeID>
        <FileTypeDescription>Microsoft Excel 2000 Workbook</FileTypeDescription>
        <Category>Spreadsheet</Category>
        <CategoryID>3</CategoryID>
        <MediaFormat>CD-ROM</MediaFormat>
        <MediaFormatID>1</MediaFormatID>
        <ReleaseDate>10 Jun 1999</ReleaseDate>
        <WithdrawnDate>30 Jun 2006</WithdrawnDate>
        <CopyrightOwner>Microsoft</CopyrightOwner>
        <CopyrightStatus>Copyright Owner</CopyrightStatus>
```

Trials And Pilots

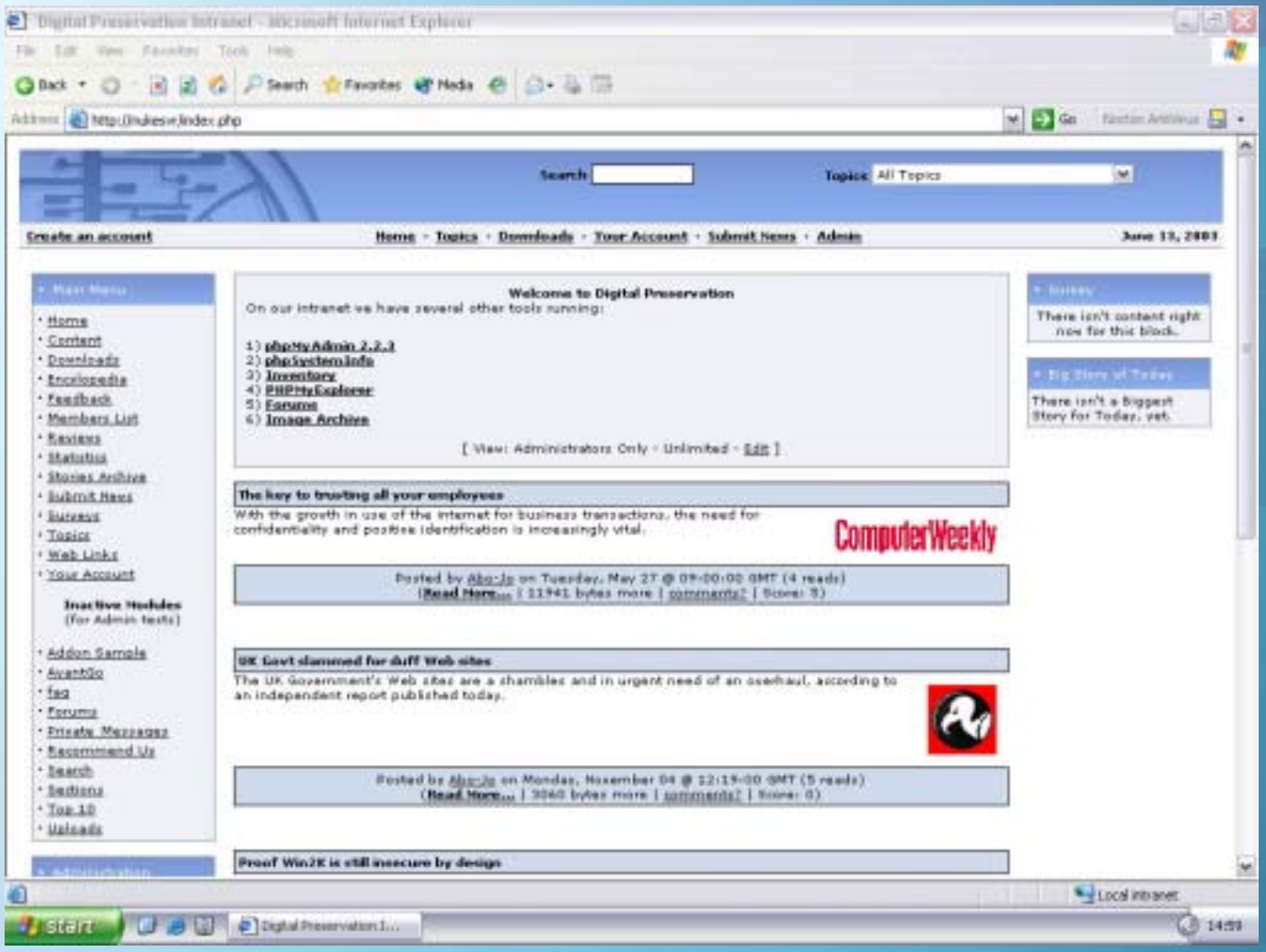
- The need to resolve
  - File format longevity
  - And other issues discovered
- Have led to:
  - Trials and pilots programme
  - Work on open source products have been included

Open Source trials

- PHP Nuke
- Post Nuke
- MySQL
- Dspace
- E-Prints
- Image Magick
- Open Office
- SuSE Linux 8.2
- PHP MySQL asset tracking application

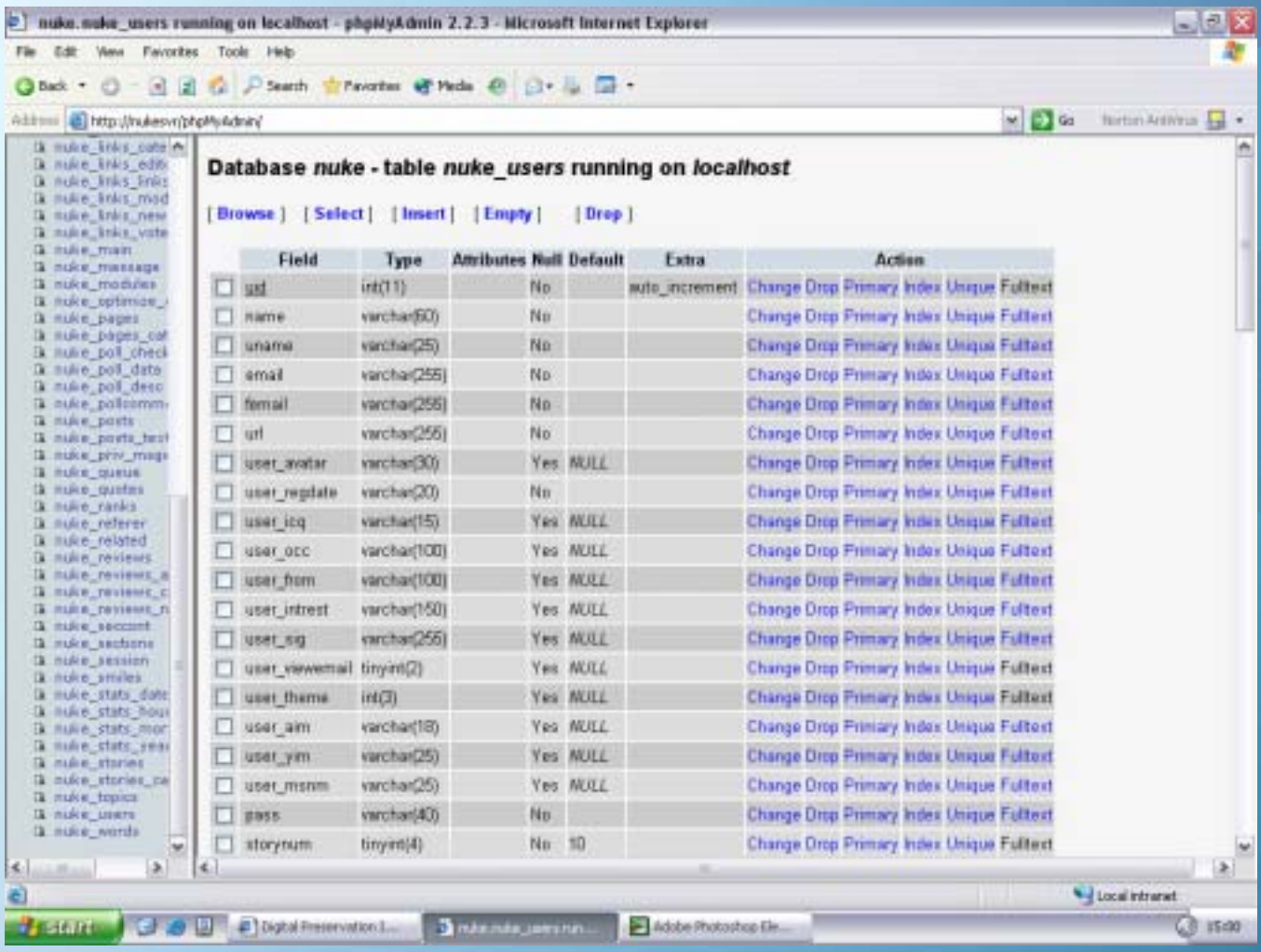
the national archives

PHP Nuke & Post Nuke



the national archives

MySQL Database

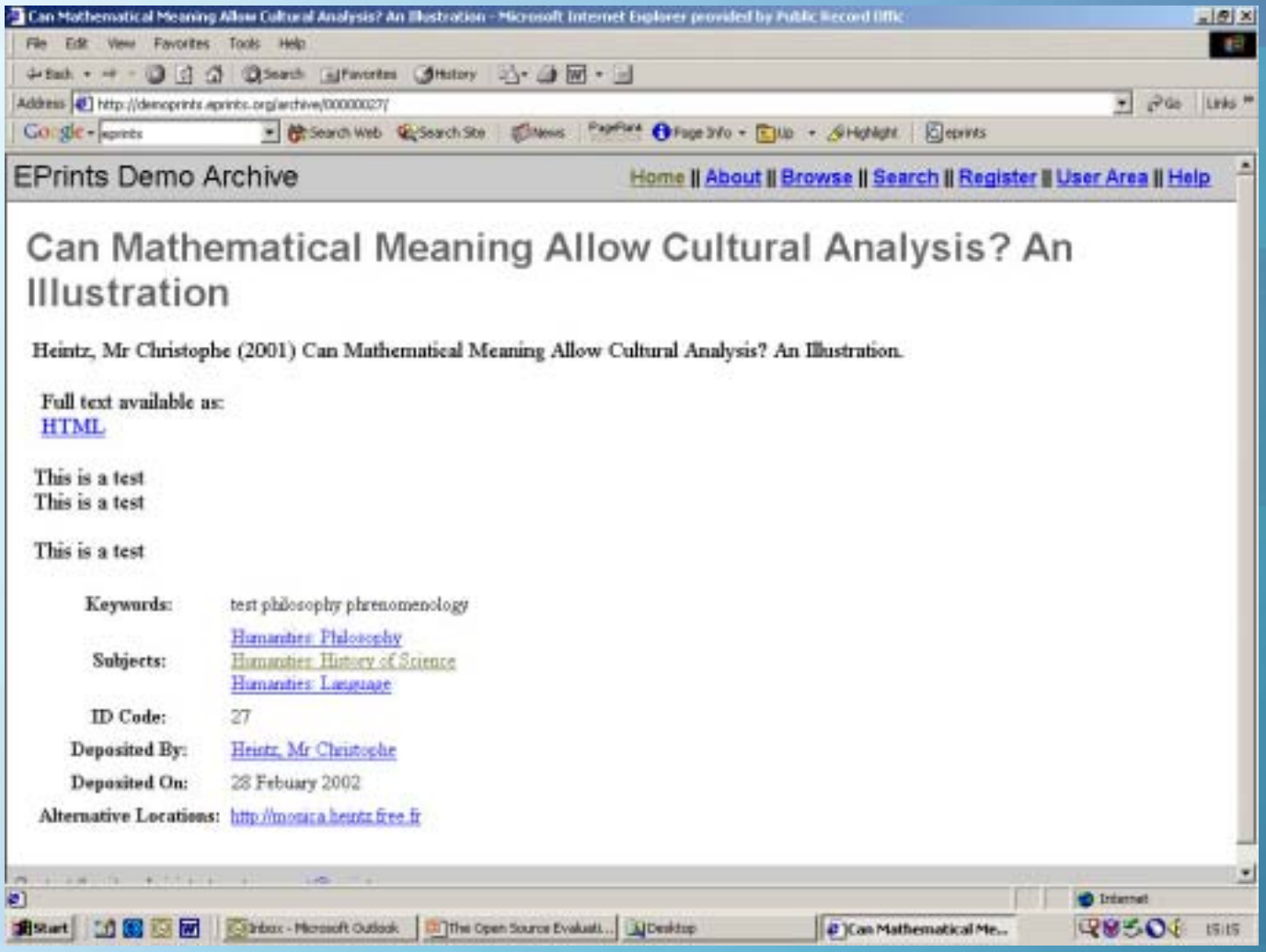


the national archives

DSpace

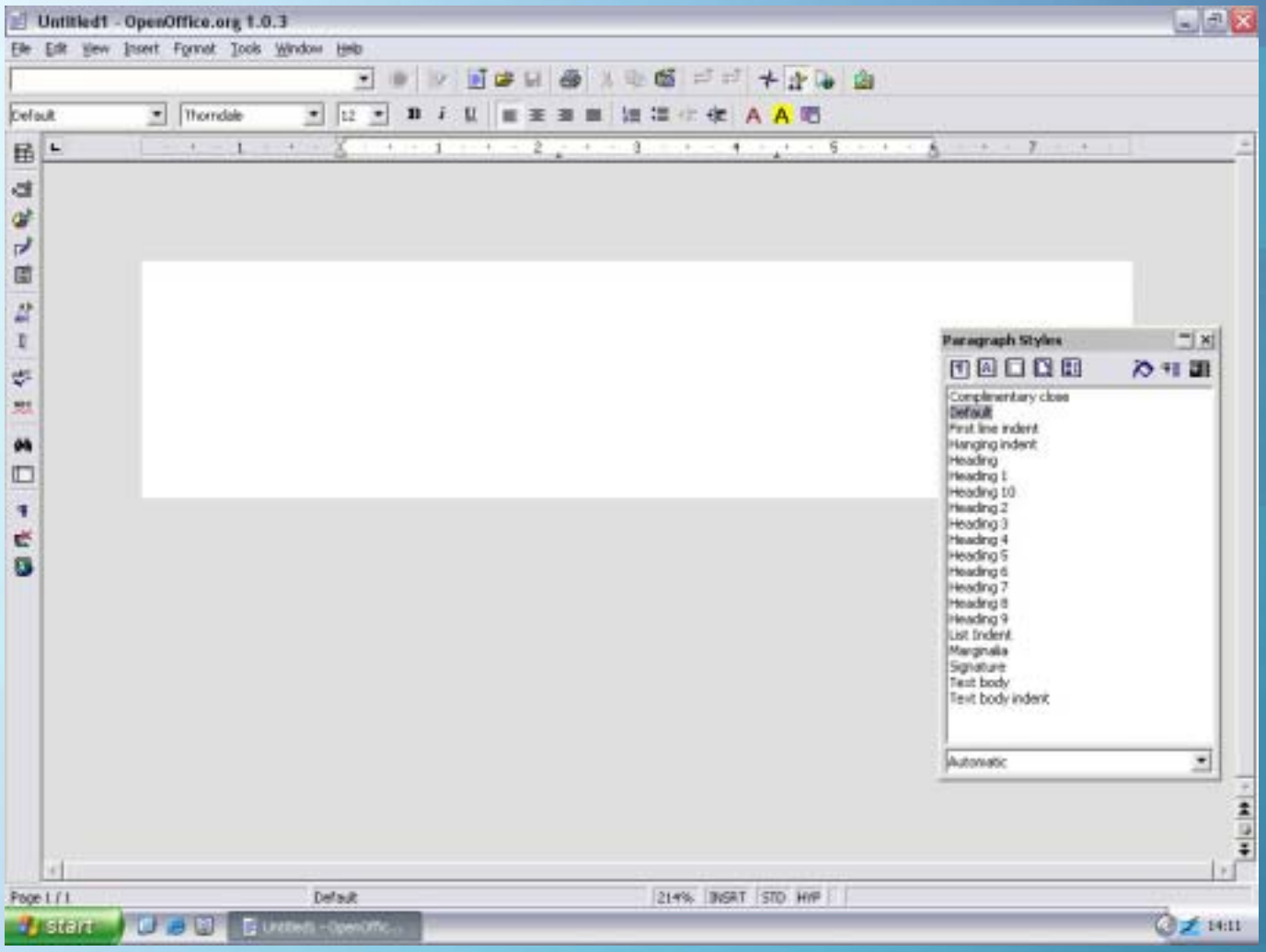
the national archives

EPrints



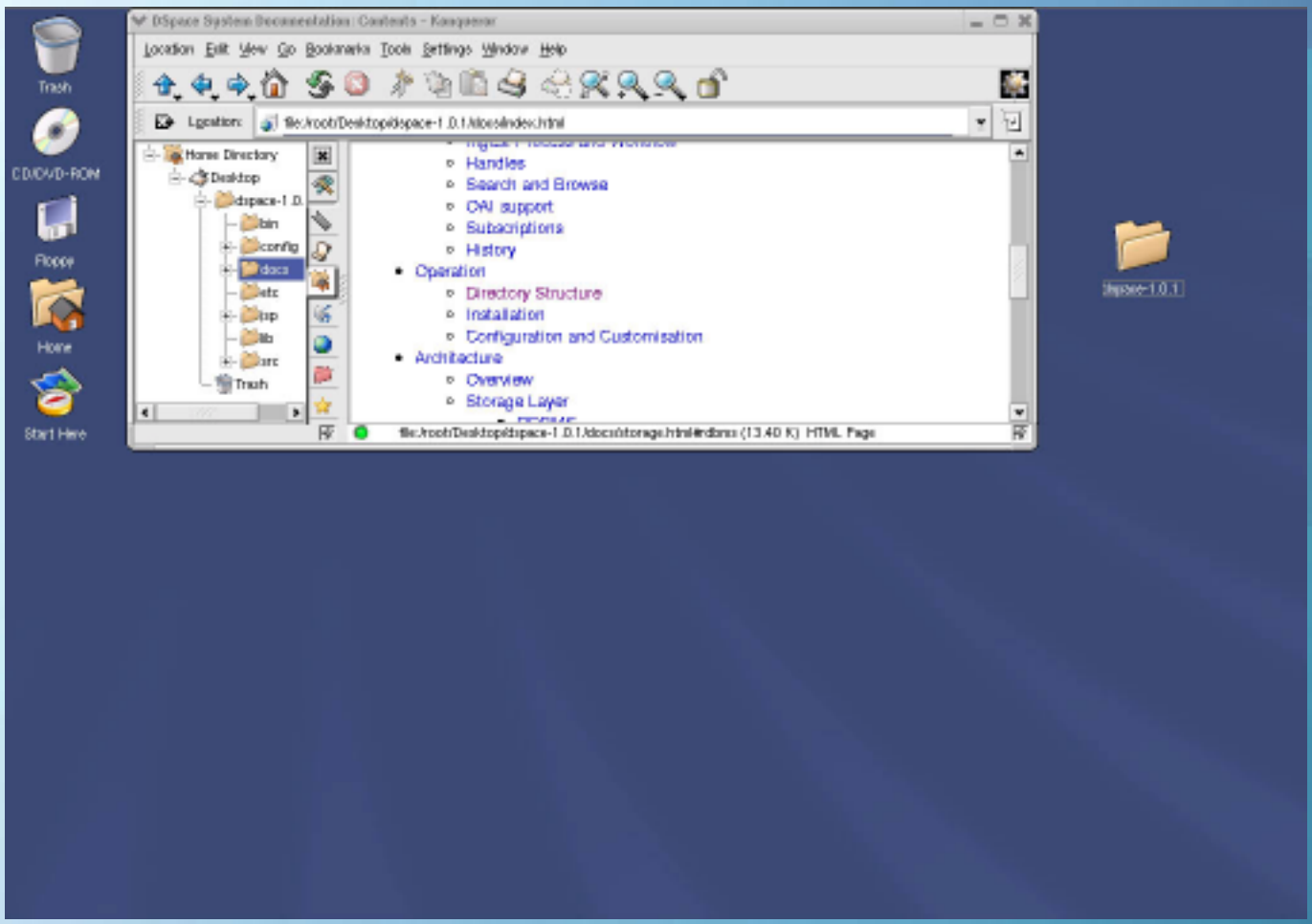
the national archives

Open Office

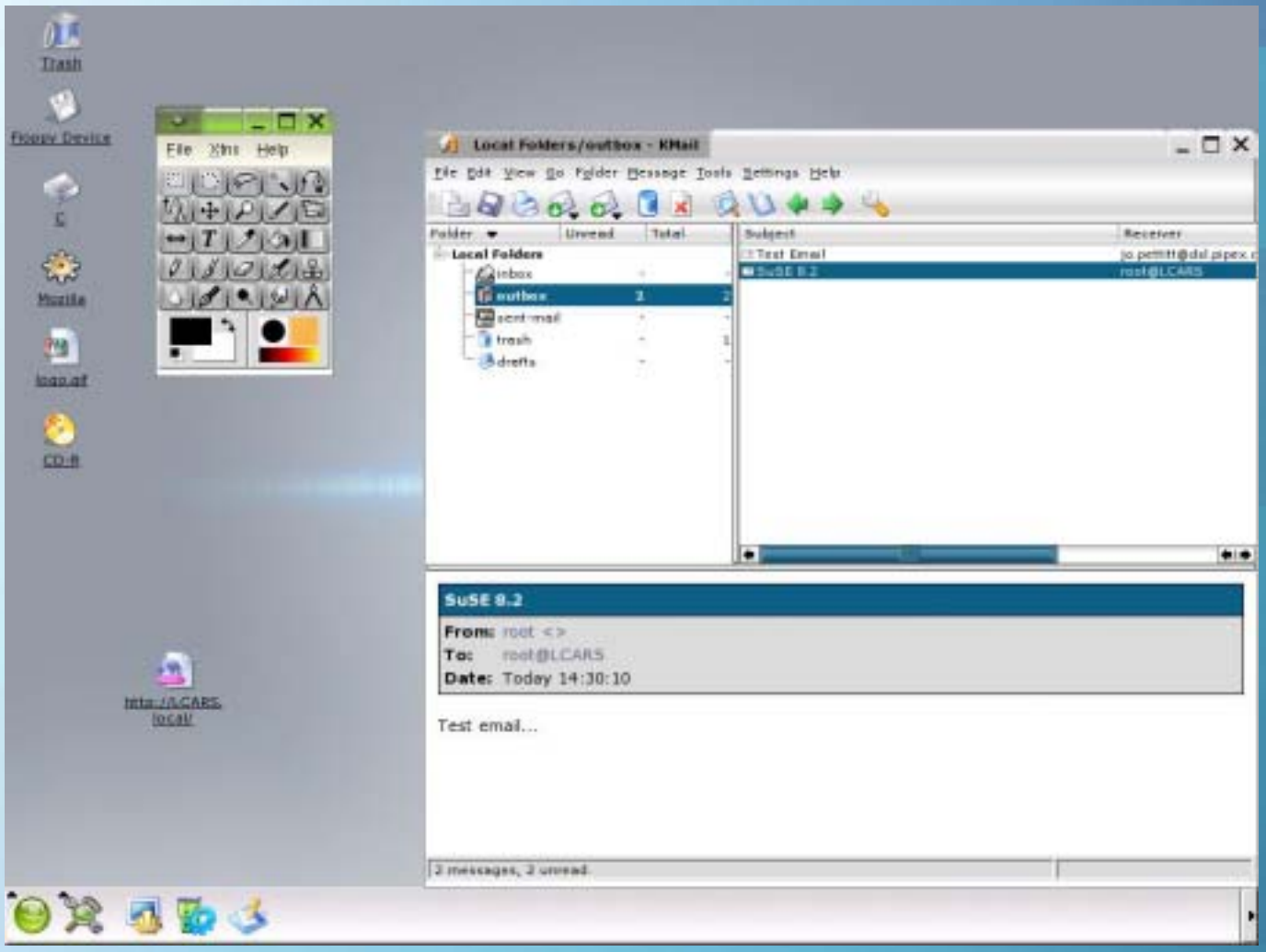


the national archives

RedHat 9.0 Linux

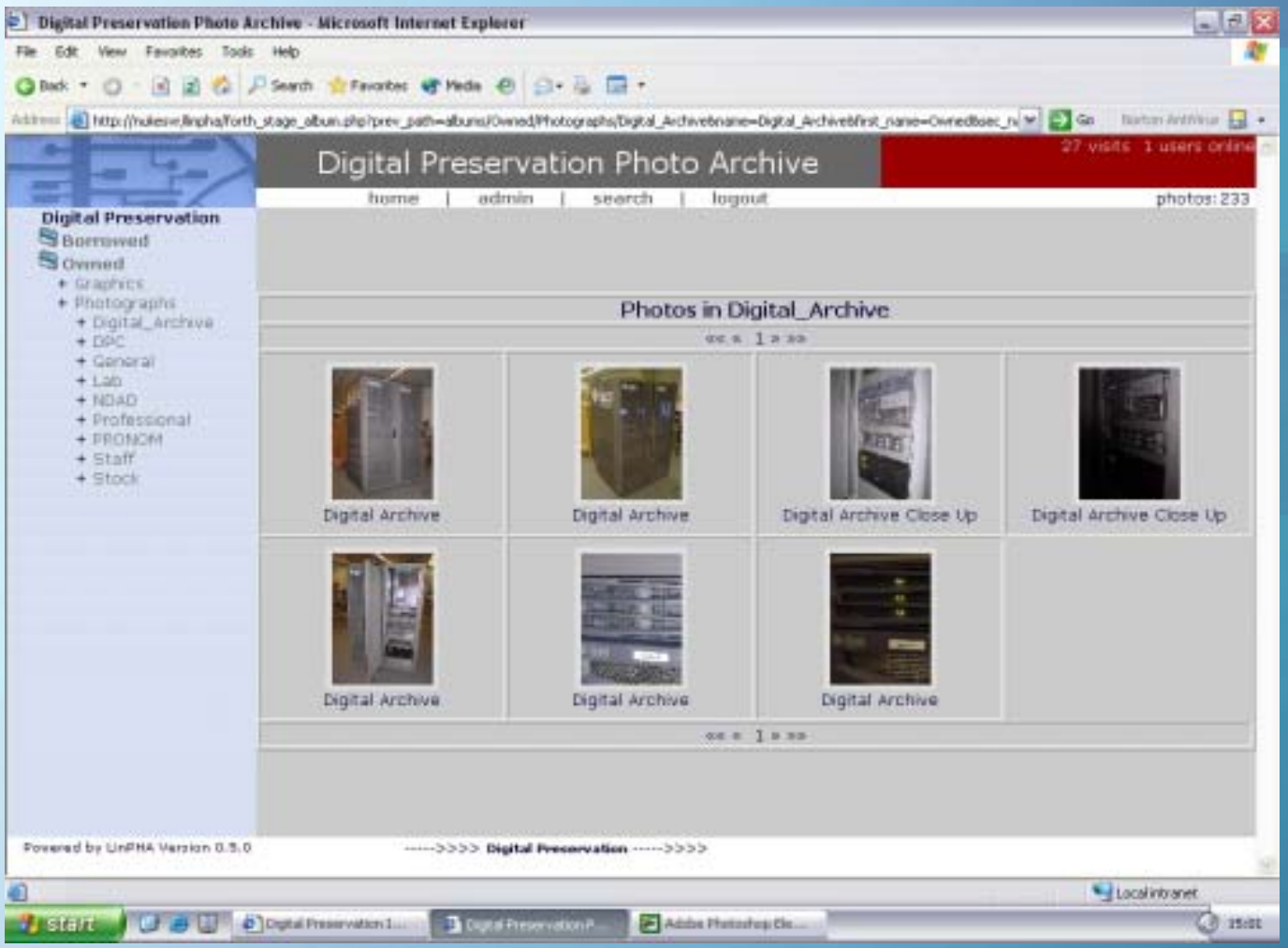


SuSE 8.2 Linux



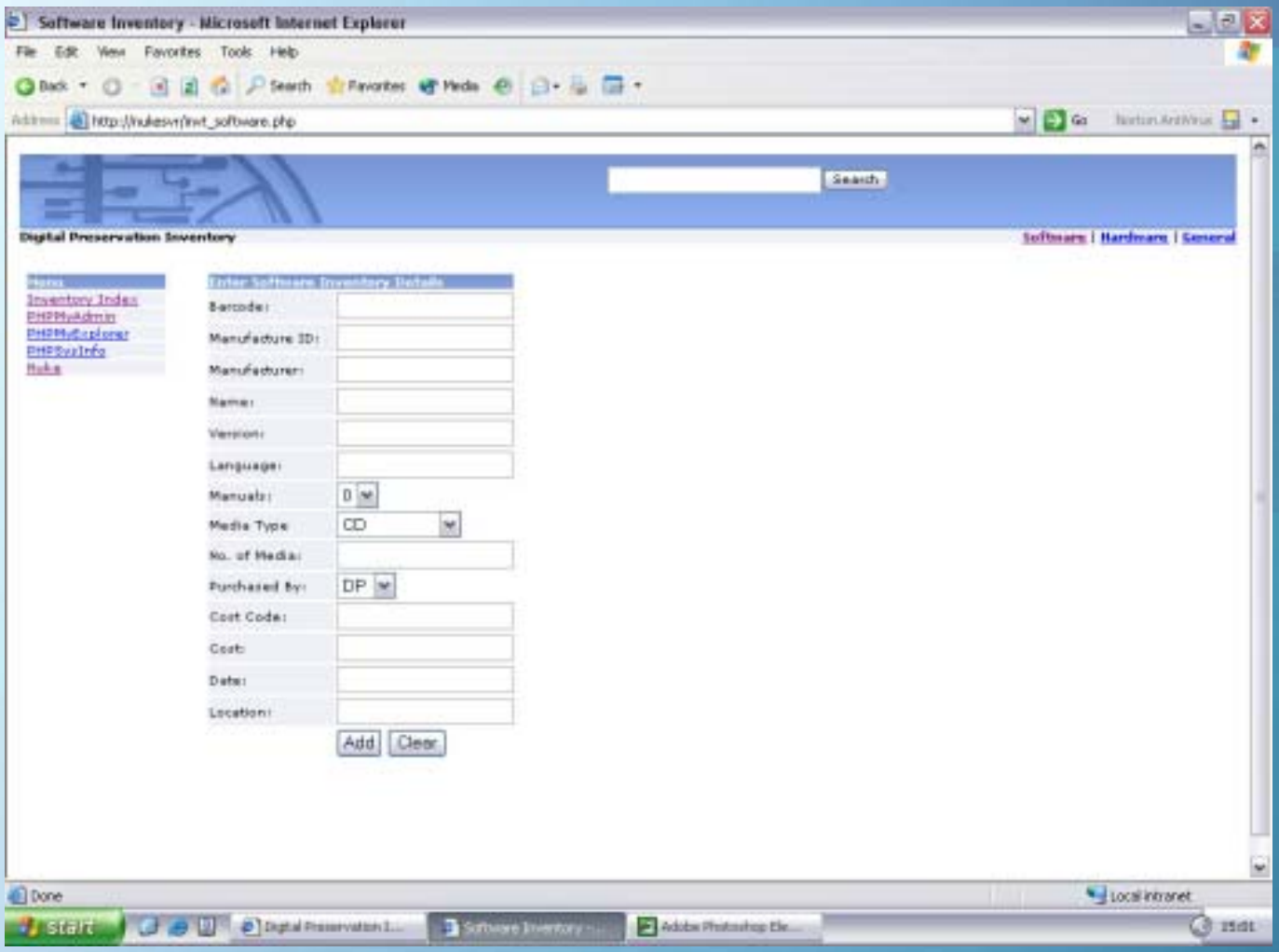
the national archives

Image Magick



the national archives

PHP MySQL Asset Tracking



the national archives

Upside

- Cheap/free
- Code is open
- No hidden spyware
- No strict EULAs

Downside

- Some immature releases
- Poor support
- Difficult to set-up
- Poor documentation

Thank you!