Workshop on Climate Data Rescue: Data and Methods

Nancy Westcott
Research Atmospheric Scientist
Midwestern Regional Climate Center
Illinois State Water Survey
Prairie Research Institute, University of Illinois
June 11-13, 2013 Amman, Jordan
Overview

Day 1
• Introduction
• What is Climate Data Rescue and Its Importance
• Climate Data in the Middle East
• **Climate Data Rescue Methods**
• Practicum

Day 2
• International Activities
• CDMP-Forts
• Metadata
• Practicum

Day 3
• Homogenization
• Practicum
• Climate Data Rescue in Jordan and Palestine
• Conclusion
Climate Data Rescue

- Preservation of current and past data into computer compatible form
- Data prior to 1950 is of particular importance

- To place current climate change trends in context of historical data, to study the natural variability of undisturbed climates
- To make current climate / agro-meteorological / hydrological / land-surface models more credible
Climate Data Rescue

• Preservation of Hard Copies
  o Sorting, and Storage of Hard Copies
  o Transfer to digital images

• Digitization
  o Translation into digital / tabular format
  o May not have to digitize all preserved data

• Quality Control
  o Insure keyed data represent observations
  o Transfer to common units
  o Insure time of observation is preserved

• Metadata and Documentation of Processing
What Data to Preserve?

- Data in danger of deterioration
- High quality datasets (historic or current) of importance - regionally, or globally that are in danger of being lost
- Data to fill in gaps of time or space;
- Data to help verify existing data
Examples of Data to Preserve

• Monthly/Annual data to help verify (sub-)daily
• Daily data or data that can be converted into monthly means
• Hourly, synoptic high resolution data
• Charts – precipitation / sunshine / pressure / temperature
• Metadata – description and definition of data
• Non-instrument data - visual weather observations; (Event descriptions; historical notes; observer notes on instruments or observation procedures)
• Phenological data (planting dates, first leaf date, first bloom date, first flower, pest emergence, pollen/spore counts)
Variable/Datasets

- Tables / Forms / Charts
- Temperature, precipitation, pressure are common variables
- Preference data prior to 1950
- Or data to fill in gaps in data (time and space)
- All associated metadata
- JMD:
  - Sorting, Imaging synoptic forms
  - Sorting, Imaging and digitizing charts
Preservation of Qualitative Resources

- Diaries
- Journals
- Pictures
- Maps
- Case Studies
- Reports
- Description of Networks
- Description of Observing Practices
- Descriptions of Instrumentation
- Description of Impacts (droughts, floods, dust storms)
Paper Notebook Archive

- Acid-free boxes
- Labeled with station name, years/months in box
- Store in room with some degree of climate control
- Index number for box

- Place box index number, station, year/month into a database for locate rapidly
Paper Chart Archive

• JMD Historical Data:
  o Sort by variable
  o Then by station, year, month
  o File drawer by file drawer

• JMD Recent and Current Data:
  o file charts with monthly notebook for easy verification of hourly, 3-hourly, 6 hourly data
Paper-copy to Digital Image
Paper-copy to Digital Image; Preservation

• For Keying, can manipulate image for easier reading
  o enlarge
  o change contrast
  o view by multiple persons

• More than one person can access at a time

• Answer questions about specific values (outliers, extremes; metadata, data not keyed, missing data)
To View Images:
Microsoft Office Picture Manager

- Microsoft Office 2007 and 2010 Tools –
  - Use to view, edit, magnify
  - High resolution jpeg images
    - 300-600 dot/inch
### Greenwich Mean Time

<table>
<thead>
<tr>
<th>Time</th>
<th>00</th>
<th>03</th>
<th>06</th>
<th>09</th>
<th>12</th>
<th>15</th>
<th>18</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attacked Thermometer aC</td>
<td>18.5</td>
<td>28.9</td>
<td>29.0</td>
<td>28.1</td>
<td>29.0</td>
<td>28.5</td>
<td>29.0</td>
<td>29.0</td>
</tr>
<tr>
<td>Barometer as read mb</td>
<td>756.5</td>
<td>736.8</td>
<td>735.7</td>
<td>734.8</td>
<td>733.8</td>
<td>733.1</td>
<td>732.8</td>
<td>732.4</td>
</tr>
<tr>
<td>Pressure at station level</td>
<td>1013.9</td>
<td>1014.7</td>
<td>1014.9</td>
<td>1014.7</td>
<td>1014.5</td>
<td>1014.3</td>
<td>1014.1</td>
<td>1014.0</td>
</tr>
<tr>
<td>Reduction to M.S.L.</td>
<td>219.7</td>
<td>218.9</td>
<td>218.2</td>
<td>217.5</td>
<td>216.9</td>
<td>216.2</td>
<td>215.5</td>
<td>215.0</td>
</tr>
<tr>
<td>Pressure at M.S.L.</td>
<td>994.2</td>
<td>993.9</td>
<td>993.7</td>
<td>993.5</td>
<td>993.3</td>
<td>993.1</td>
<td>992.9</td>
<td>992.8</td>
</tr>
<tr>
<td>Barometric tendency (mb)</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.5</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Minimum temperature aC</td>
<td>16.0</td>
<td>16.0</td>
<td>16.5</td>
<td>17.0</td>
<td>17.5</td>
<td>18.0</td>
<td>18.5</td>
<td>19.0</td>
</tr>
<tr>
<td>Maximum temperature aC</td>
<td>8.0</td>
<td>23.0</td>
<td>19.0</td>
<td>17.0</td>
<td>14.0</td>
<td>11.0</td>
<td>8.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Gross Minimum</td>
<td>4.0</td>
<td>8.0</td>
<td>7.0</td>
<td>6.0</td>
<td>5.0</td>
<td>4.0</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Earth temperature aC</td>
<td>2.0</td>
<td>3.0</td>
<td>4.0</td>
<td>5.0</td>
<td>6.0</td>
<td>7.0</td>
<td>8.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Dew Point</td>
<td>2.0</td>
<td>3.0</td>
<td>4.0</td>
<td>5.0</td>
<td>6.0</td>
<td>7.0</td>
<td>8.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>40.0</td>
<td>45.0</td>
<td>50.0</td>
<td>55.0</td>
<td>60.0</td>
<td>65.0</td>
<td>70.0</td>
<td>75.0</td>
</tr>
<tr>
<td>Evaporation</td>
<td>4.0</td>
<td>2.0</td>
<td>1.0</td>
<td>0.</td>
<td>0.</td>
<td>0.</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td>Rainfall</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Wind direction</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Wind speed</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Type of High Cloud</td>
<td>6.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Medium</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Low</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Height of base of low cloud Fe</td>
<td>1000</td>
<td>3000</td>
<td>5000</td>
<td>7000</td>
<td>9000</td>
<td>11000</td>
<td>13000</td>
<td>15000</td>
</tr>
<tr>
<td>Amount of low cloud Eighths</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total amount of cloud</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Visibility</td>
<td>5.0</td>
<td>4.0</td>
<td>3.0</td>
<td>2.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Present weather</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Past weather</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Initials of observer</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Greenwich Mean Time</td>
<td>00</td>
<td>03</td>
<td>06</td>
<td>09</td>
<td>12</td>
<td>15</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Attached Thermometer</td>
<td>-287.0</td>
<td>-292.0</td>
<td>290.0</td>
<td>251.0</td>
<td>290.4</td>
<td>290.0</td>
<td>292.6</td>
<td>290.0</td>
</tr>
<tr>
<td>Barometer as read</td>
<td>936.5</td>
<td>936.0</td>
<td>935.7</td>
<td>935.8</td>
<td>934.9</td>
<td>935.0</td>
<td>935.0</td>
<td>935.9</td>
</tr>
<tr>
<td>Reduction to station level</td>
<td>-1.6</td>
<td>-1.8</td>
<td>0.1</td>
<td>1.9</td>
<td>1.9</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Pressure at station level</td>
<td>834.9</td>
<td>834.2</td>
<td>833.8</td>
<td>833.1</td>
<td>833.9</td>
<td>833.2</td>
<td>833.8</td>
<td>833.1</td>
</tr>
<tr>
<td>Reduction to M.S.L.</td>
<td>888.7</td>
<td>888.0</td>
<td>887.8</td>
<td>887.6</td>
<td>887.3</td>
<td>887.2</td>
<td>887.4</td>
<td>887.1</td>
</tr>
<tr>
<td>Pressure at M.S.L.</td>
<td>1236</td>
<td>1232</td>
<td>1237</td>
<td>1237</td>
<td>1232</td>
<td>1230</td>
<td>1232</td>
<td>1232</td>
</tr>
<tr>
<td>Barometric tendency (999)</td>
<td>80.4</td>
<td>76.4</td>
<td>72.2</td>
<td>72.5</td>
<td>72.5</td>
<td>72.5</td>
<td>72.5</td>
<td>72.5</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>16.0</td>
<td>16.0</td>
<td>16.5</td>
<td>16.0</td>
<td>16.5</td>
<td>17.2</td>
<td>17.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Minimum</td>
<td>8.0</td>
<td>2.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Grass Minimum</td>
<td>4.5</td>
<td>4.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Earth temperature 0.8 m</td>
<td>8.5</td>
<td>8.5</td>
<td>8.5</td>
<td>8.5</td>
<td>8.5</td>
<td>8.5</td>
<td>8.5</td>
<td>8.5</td>
</tr>
<tr>
<td>Daily bulk temperature °C</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>Wet Bulk</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>Difference</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>Dew Point</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>50/50</td>
<td>50/50</td>
<td>50/50</td>
<td>50/50</td>
<td>50/50</td>
<td>50/50</td>
<td>50/50</td>
<td>50/50</td>
</tr>
<tr>
<td>Time</td>
<td>Data 1</td>
<td>Data 1</td>
<td>Data 1</td>
<td>Data 1</td>
<td>Data 1</td>
<td>Data 1</td>
<td>Data 1</td>
<td>Data 1</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Greenwich Mean Time</td>
<td>00</td>
<td>03</td>
<td>06</td>
<td>09</td>
<td>12</td>
<td>15</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Attached Thermometer °C</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>Barometer as read mb</td>
<td>935.0</td>
<td>935.0</td>
<td>935.0</td>
<td>935.0</td>
<td>935.0</td>
<td>935.0</td>
<td>935.0</td>
<td></td>
</tr>
<tr>
<td>Reduction to station level</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td></td>
</tr>
<tr>
<td>Pressure at station level</td>
<td>834.0</td>
<td>834.0</td>
<td>834.0</td>
<td>834.0</td>
<td>834.0</td>
<td>834.0</td>
<td>834.0</td>
<td></td>
</tr>
<tr>
<td>Reduction to M.S.L.</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td></td>
</tr>
<tr>
<td>Pressure at M.S.L.</td>
<td>888.0</td>
<td>888.0</td>
<td>888.0</td>
<td>888.0</td>
<td>888.0</td>
<td>888.0</td>
<td>888.0</td>
<td></td>
</tr>
<tr>
<td>Barometric tendency (°Sp)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Maximum temperature °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Minimum °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Average temperature °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Dry bulb temperature °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Wet bulb °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Difference °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Dew Point °C</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Evaporation mm</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Rainfall mm</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Wind direction</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Speed at m/s</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Type of High Cloud</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Height of base of low cloud Pt</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Amount of low cloud Eighths</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Total amount of cloud</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Visibility km</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Present weather</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Frost weather</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Initial of observer</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
</tbody>
</table>
Non-Digital Copies

- Books
- Loose pages / charts
- Microfilm

- If possible, keep “hard” copies as backup at least until papers are digitized. Permanently if possible – libraries may wish to keep copies.

- Copies of the data should be dispersed.

- Special rooms with temperature/humidity control where protected from fire/theft/insects/mold
Methods for Conversion to Image

- Books – camera / flatbed scanner
- Loose pages / charts – camera / feeding scanner
- Microfilm – (microfilm readers scan to pdf file)
- Microfiche - same
- (scanner technology improved so less expensive and faster)

- Camera
- Scanner
Microfilm to Image

• UIUC library “proscan”
  o Microfilm reader with scanning capabilities

• Convert to pdf (or tiff or jpeg)

• As microfilm is an antiquated technology/medium, only libraries with microfilm would likely have access to this technology
Cameras

- Fast
- Less handling than flatbed scanner
- Relatively Inexpensive
UIUC Camera Setup
Photographed image on screen
Camera

• Digital Single Lens Reflex (SLR)
• Camera Stand
• Lightning
• Software to view on computer
  o Comes with camera
  o Microsoft Office Tools
Camera

• Images
  o 300-600 dot/inch
  o Tiff or JPEG – UIUC library
• Review images on computer for quality
  o Lighting - external;
  o focus;
  o centering.
  o Document image name
• Inventory images by station, year and month
Camera Suggestions:

- Zoom lens.
- Macro close-up mode.
- Ability to adjust the film speed sensitivity that is also known as ISO.
- AC adapter to power the camera.
- Ability to cancel the flash.
- Ability to select the image quality.
- Ability to adjust the white-balance.
- Ability to connect to a tripod.

- Memory Card
- Self Timer
- Or
- Wireless Mobile Adaptor
- Wireless Remote Control for Shutter Release
- Many digital SLRs would be suitable.
One Camera – two lights
Two Cameras – Two Lights – Upside down
Scanners

• Much slower (both handling, scan-time)
• Easier to control quality

• Flatbed –
  o books with tight binding
  o Large maps

• Feeding scanner
  o charts
Scanners

- Plustek OpticBook A300 Flatbed Scanner - 600 DPI, 48-Bit Color, USB 2.0 – for books, hand turn page, slow
UIUC Library
Flatbed Scanner

SLOW

Glass to edge of scanner, so image to interior edge (gutter) of book.

Preserves delicate binding on books.

Software to view and store image to computer.
Old Scanning Method

Not recommended

CDMP Fragile, bound document scanning
UIUC Feeding Scanner
UIUC Feeding Scanner
Feeding Scanner (charts)

Feed in multiple charts at one time; each goes to a separate file.
Accounting / Progress Spreadsheets

- Design spreadsheet to record info about each image; accounting and file tracking important
- Count number of pages to be images
- Progress spreadsheets
Final Metadata, Image and Data Table Files

- Safeguard files
- Organize files so users can easily access data, via computer or web site
- Store on appropriate media (CD and hard-drive)
- OFF SITE STORAGE in case of disaster
Image Metadata

• Scanner/camera – type / resolution
• Document –
  o Publication Title, number of pages
  o Medium (loose pages, book, microfilm)
  o Image number (camera and scanner will provide)
• Create “Accounting/Progress” spread sheets
• Store Images in a directory by station and year
Digital Copies: Refresh Storage Medium

- Magnetic Tapes
- Floppy disks
- 8-mm tapes
- CD / DVD
- Hard drive – internal and external

- Migrate to newest technology, including software to record and store data
- Keep flat / ascii / text files too. - widely recognized
- Keep file naming simple
From Image to Data Tables - Digitizing

- Keying
- Digitizing hardware and software
- Optical Character Recognition (OCR) Software
- Crowd sourcing
Digitizing – all techniques need QC

- Keying
  - Double keying best
  - 1:1 keying (no coding of data)
  - Retain original images as it is not possible to key all data
  - Who Students / interns / interested parties (including retired climatologists)

- Chart digitization
  - Trace contours with hand-held mouse; use software to convert to specific time period
Digitizing Techniques – (tomorrow)

• Optical Conversion Software (OCR)
  o Best with “Clean” images
  o Best with typed, **NOT HANDWRITTEN** images
  o Software available for testing

• Crowdsourcing
  o Web-based keying by volunteers;
Keying Software

- CDMP – developed over years – contracted; software stayed with contractors when funding lost;
- IEDRO – developing keying software to take over keying task - through workstation or crowd sourcing (web) – on hold
- JCMD
- Australian (CliDe) – key and qc
Keying

- Often data keyed into an excel spread sheet.
- Data in columns, easier
- Data not in columns, developing a (web-based) template would be beneficial, where keyer would know where to look on a form to find element, and data would go into a database
Keying for CDMP-Forts

- Hand written monthly forms
- Many different types of forms
- Volunteer / non-meteorological observers
- Forms filled out in many ways
- Non-climatologist keyers

- Software and keying instructions developed over a number of years
- Web-based template for keying variables. (software remained with company when funding lost.)
IEDRO – International Environmental Data Rescue Organization

- Volunteer
- Data available to all
- Mainly African / South American counties
- Keyed with whatever resources were available
- Keying software development - On hold
Chart Digitization Software

- Commercial Software – need testing
  - [http://www.digitizeit.de/](http://www.digitizeit.de/)
  - [http://www.silkscientific.com/graph-digitizer.htm#Graph-Digitizer-b2](http://www.silkscientific.com/graph-digitizer.htm#Graph-Digitizer-b2)

- Software/database developed for Middle East
  - EXACT [http://exact-me.org/](http://exact-me.org/)

- ISWS – software; older programming language, needs to be adapted to a newer compiler
Digitizing Hardware

Tablet

Cursor

Summagraphics; new, refurbished
ISWS experience, #1

1) Digitized CCPN data for 20 years (25 gages)
2) Chicago – hourly data required
3) Laborious
   - Weekly charts - 1 revolution per day, typically 6-8 lines/chart – 1989-2001 (12 years x 52 weeks x 25 gages = 15,600 charts)
   - Monthly Charts - 4 revolutions per month, typically 4 lines/chart – 2001-2012 (12 years x 12 months x 25 charts = 300 charts)
ISWS experience, #1

• Prior to digitizing:
  o Id on/off times; id the order of 6-8 lines on the chart
• Digitizing
  o input specifics of chart
  o Trace one line at a time
  o Indicate missing data
  o Compute temporal offsets from speed of drive
• Software created sequential files for a single gage
• Created monthly tables at desired time steps
ISWS Experience #2

- NADP – National Acid Deposition Program
- Gages across US / and chemical samples
- Desire daily data – digitized by hand/eye
  - Observer notes on chart: on/off times, and daily rainfall amounts
  - Data QC’ed at ISWS: check daily rainfall amounts.
  - Scanning Charts, color inexpensive scanner, 150-300 dots per inch
  - Jpeg files / stored by station / year
Executive Action Team (EXACT) Multilateral Working Group on Water Resources Water Data Banks Project

- One aspect, a rain-intensity study
- rainfall-intensity database structure, rainfall-intensity software (RAINDIGITIZER, RAINPLOT, and RAINSHARE), and an online manual to document use of the software.
- custom Visual Basic programs that represent a complete solution for digitizing, storing, analyzing, reporting, mapping, and sharing rainfall-intensity data
EXACT

• Many participating Agencies in Jordan, Palestine, and Israel.
• Developed a regional database system to input, store, analyze, and report rainfall-intensity data, providing scientists with the capability to evaluate rainfall intensities in areas of Israeli, Jordanian, and Palestinian.
• Additional computer programming support was provided by the University of Central Florida (UCF).
• The project was established in 1999 and this aspect was completed in 2006 with publication of the final report.
RainDigitizer/Plot/Share

• The source code was developed by the University of Central Florida (UCF) Department of Engineering Technology (www.cecs.ucf.edu), project staff from the Core Parties, and the U.S. Geological Survey (USGS) (www.usgs.gov)

• It was maintained as an open source resource by the Department of Engineering Technology, UCF and the International Water Resources Branch, USGS (Reston).

• Demonstration here
• Demonstration coming

Questions ??????
EXACT

- Many participating Agencies in Jordan, Palestine, and Israel.
- Developed a regional database system to input, store, analyze, and report rainfall-intensity data, providing scientists with the capability to evaluate rainfall intensities in areas of Israeli, Jordanian, and Palestinian.
- Additional computer programming support was provided by the University of Central Florida (UCF).
- The project was established in 1999 and this aspect was completed in 2006 with publication of the final report "Application of Methods for Analysis of Rainfall Intensity in Areas of Israeli, Jordanian, and Palestinian Interest".
Optical Character Recognition (OCR) Software

- [http://captricity.com/](http://captricity.com/) Capture: translates the pages for your, $1.00 / 5 pages

- Other softwares, for free softwares, be careful not to download extra softwares

- Useless for Handwriting;
Intelligent Character Recognition (ICR) Software

- [http://www.a2ia.com/Addon_Site/Upload/Autres/WK/case_studies/A2iA_CaseStudy_Coutot-Roehrig.pdf](http://www.a2ia.com/Addon_Site/Upload/Autres/WK/case_studies/A2iA_CaseStudy_Coutot-Roehrig.pdf)

Features

- Film Speed Sensitivity or ISO: a key part of the camera operation in determining the shutter speed, aperture.

- For film speeds of 100 to 800 ISO:
  - With camera set to 800 ISO, less light is needed, using a higher shutter speed, and a smaller aperture.
  - This increases the depth of field which improves the sharpness of the text.
Features

• Zoom Lens – allow camera to stay at same distance from page, but can adjust focus

• Macro Close up Mode –
  o normal mode, the closest to an image is 1'8" and have the camera in focus.
  o Macro close-up mode, as close as 0.8". easier to properly compose the picture

• Florescent lighting, NOT recommended.