HIDE: A Hydrological Integrated Data Environment

Yao Liang, Nimmy Ravindran
Virginia Tech
Acknowledgement

- Xu Liang at UC Berkeley provides valuable insights and helps on our various questions regarding hydrological datasets, data sources, data retrieval and process.
Hydrological Integrated Data Environment (HIDE)

What is HIDE?

- A web-based data integration system for heterogeneous datasets from heterogeneous data sources/providers
- Data search and on-the-fly data integration based on user queries
Hydrological Integrated Data Environment (HIDE)

How does HIDE work?

- Datasets are published by the data providers through various technologies
- Definition of basic unit of information – DataNode (data sources or data sets)
- XML Schema based data integration
- Taxonomical organization of DataNodes while incorporating the domain and structural knowledge.
Hydrological Integrated Data Environment (HIDE)

- Application services
- Data Models
- Transformation Engine
- Access Engine

Distributed Data Sources
HIDE: Schema

- XML Schema based data integration.
- Three kinds of schema.
  1. Integration schema.
  2. Syntax and semantic schema of data sources.
  3. Data Model schema of the integrated data.
Integration schema

1. Describe the data source or dataset.
2. What kind of operations is permitted on the data source / dataset?
3. How to access the external data engines?
4. What kind of transformations should be applied to the data for an effective integration?
Syntax and semantic schema

1. What are the syntax and semantic details of the data from the data source? E.g.: formats, type of data files etc.
DataModel schema

- Describe the data model of the integrated data.
- Temporal, Spatial and Temporal-Spatial data model.
- Records the nature of the integrated data with more emphasis on temporal and spatial characteristics of the data.
- Suitable for data exchange.
DataModel schema contd.

1. A comprehensive description of all aggregated datasets.
2. Temporal and spatial coverage.
3. Syntactic details of the aggregated data.
4. Data representation.
An example of DataModel metadata.

```
<DataModel xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation='c:\IDAM\xml\xsd\dataModelSchema.xsd' type="temporal">
    <Datasets>
        <Dataset name="">
            <DataSource> …… </DataSource>
        </DataSource>
        <coverage>
            <temporal type="single"> …….. </temporal>
            <spatial> …….. </spatial>
        </coverage>
    </Dataset>
    ...
</Datasets>
<temporal datasetType="dataTable" column="1" dateFormat="YYYY-MM-DD hh:mm" timezone="EST">
    <syntax type="Ascii" delimiter=",” />
    <dataTable numRecords="unbounded">
        <dataFields>
            <dataField name="Precipitation" type="integer"/>
        </dataFields>
    </dataTable>
</temporal>
```
HIDE - Snapshots

Welcome to HIDE - Search engine

Search Results
1 results found

GPCC Global Precipitation climatology center - precipitation Data for the period 1986-1987
Precipitation data are main input to global hydrological cycles and climate models. The conventional
rain-gauge measurements are the only direct measure of rainfall. This dataset is comprised of monthly
gridded area-mean rainfall totals for the period January 1986 to March 1999 on a 1 by 1 degree global
grid.
daac.gsfc.nasa.gov/daac-bin/nph-ff/DODS/inter_disc/hydrology/precip/gpcp/gpcc/
HIDE - Snapshots

Please select the search conditions

Sites: List of sites for the state of Alabama
UCHEE CREEK NEAR FORT MITCHELL AL

Days: Number of Days in the range 1-31

Please select the output params

- 00060 Discharge (DD01)
- 00065 Gauge Height (DD02)
- 00045 Precipitation (DD04)

Submit
### HIDE - Snapshots

<table>
<thead>
<tr>
<th>Date</th>
<th>Agency code</th>
<th>Site no</th>
<th>Precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-09-25 00:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 01:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 02:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 03:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 04:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 05:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 06:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 07:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 08:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 09:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 10:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 11:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 12:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 13:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 14:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 15:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 16:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 17:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 18:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
<tr>
<td>2004-09-25 19:00</td>
<td>USGS</td>
<td>02361000</td>
<td>0.0</td>
</tr>
</tbody>
</table>
HIDE prototype

- Can be accessed,
  
  http://ari423x16.ari.vt.edu:8080/IDAM/home.html