HIS DESKTOP 1.0
FUNCTIONAL SPECIFICATIONS

June 2009

Table of Contents

1.   INTRODUCTION

This document provides functional specifications for a software product called HIS Desktop, which is part of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS). HIS includes extensive development of data storage and delivery tools and standards. Among these tools and standards are the Observations Data Model (ODM) for organizing and storing observational data, the WaterML language for communication of observational data, and the WaterOneFlow web services for publishing and providing access to observational data through the use of WaterML and a core set of query methods over the Internet. The CUAHIS HIS uses these tools within a service oriented architecture (SOA) to provide users with access to distributed data services that are published by a number of academic and government data sources.

To enable the use of WaterOneFlow web services, HIS includes custom applications for some of the most commonly used software in Hydrologic Science, which extend the software so that it can communicate with WaterOneFlow web services for data access. Examples include HydroSeek for a web browser, HydroGET for ESRI's ArcGIS, and HydroExcel for Microsoft Excel. However, no single application encompasses the functionality of all the others, and the choice of application often depends on what software licenses are available to the end user. As such, it can be confusing to potential HIS users as to what exactly "HIS" is or how to access data using HIS.

HIS Desktop will be free software that operates on a user's desktop computer, with the intent of serving as a common entry point into data resources published using the CUAHSI HIS WaterOneFlow web services. HIS Desktop will consolidate many key capabilities of existing HIS tools (e.g., data query, map-based visualization, data download, local data maintenance, data editing, data visualizing, and linkage with integrated modeling systems such as OpenMI).

1.1.   INTENDED USE

HIS Desktop is intended to solve the problem of how to obtain, organize, and manage hydrologic data on a user's computer to support analysis and modeling. HIS Desktop will be a platform for the integration of hydrologic data, which can be used in analysis applications such as R, MATLAB, and Excel, or in custom code developed by the end user. The HIS Desktop design paradigm includes the use of a plug-in architecture and data abstraction layer that will allow extension of the core functionality. HIS Desktop will provide local access to data obtained from distributed data services that are part of the Internet-based, SOA that the CUAHSI HIS project has developed for the sharing of hydrologic data.

1.2.   TARGET USERS FOR HIS DESKTOP

It is anticipated that HIS Desktop users will include university faculty, graduate and undergraduate students, K-12 students, engineering and scientific consultants, and others. HIS Desktop users may or may not have a technical scientific or computer science background. It is expected that these users will be primarily interested in discovering and retrieving observational data from the HIS system for use in software installed on their local computer.

2.   HIGH LEVEL DESIGN

The high level design for HIS Desktop is shown in Figure 1. HIS Desktop will serve as a common window into observational data published using WaterOneFlow web services. Data discovery will be accomplished through searches across a comprehensive metadata catalog maintained at HIS Central and/or by querying individual HIS Servers hosting WaterOneFlow web services. Searches against the central metadata catalog will be facilitated by a web service application programmer interface (API) that exposes the metadata catalog and the Hydrologic Ontology maintained at HIS Central. Searches on the central metadata catalog will result in a set of metadata that can then be used to further refine a search and specify the datasets that a user would like to download. Data downloads will be performed by making GetValues calls to the appropriate WaterOneFlow web services. Downloaded data will be stored in a desktop data repository database following a relational database schema. This database will be accessible either through an API or directly. Visualization and analysis tools that are part of HIS Desktop (e.g., Time Series Analyst) will be developed using the API data access method to maintain a level of data access consistency and integrity. Additionally, users will be able to access the data through third party data analysis applications that have the ability to read from a Relational Database Management System (RDMS). Such applications include but are not limited to R, MATLAB, and Excel. HIS Desktop will allow users to export and import data in a number of commonly used file formats as described in the data export section below.
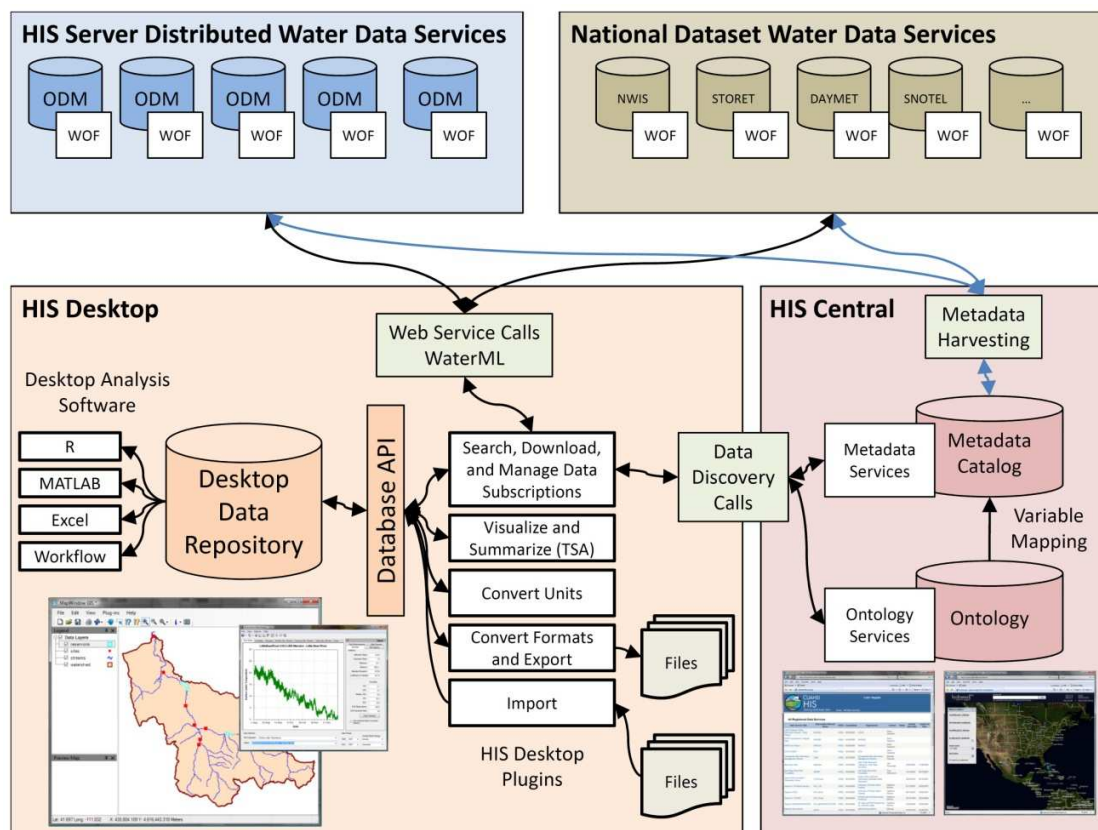

Figure 1. High level design for HIS Desktop.

3. FEATURES AND FUNCTIONAL REQUIREMENTS

The primary purpose of HIS Desktop is to facilitate discovery and access of hydrologic data. A secondary purpose is to provide support for data manipulation and synthesis. The user will primarily interact with HIS Desktop via a graphical user interface (GUI) with the functionality described below.

3.1. DATA DISCOVERY

HIS Desktop will support two different methods of data discovery: 1) ontology-based discovery across all WaterOneFlow web services that have been registered at HIS Central and for which metadata has been harvested and stored in the HIS Central metadata catalog; and 2) discovery of data within a single WaterOneFlow web service that has not been registered at HIS Central. The first type of data discovery will be supported by HIS Central metadata web services that expose the contents of the HIS Central metadata catalog. The second type of data discovery will involve making data discovery calls directly to the web service that has not been registered with HIS Central. This approach will facilitate both the use of datasets cataloged and documented at HIS Central, as well as use of datasets stored on individual or regional HIS Servers but not necessarily registered with HIS Central.

3.1.1. DATA DISCOVERY USING THE HIS CENTRAL METADATA CATALOG

HIS Central includes a catalog of metadata describing the observational data series served by WaterOneFlow web services that have been registered with HIS Central. The metadata catalog, which includes the mappings between variables and HIS Ontology concepts, is updated weekly and represents a comprehensive listing of data published using WaterOneFlow services and registered at HIS Central. The contents of the HIS Central metadata catalog will be exposed by a web service API that is currently under development. At a minimum, this metadata catalog web service API will provide methods for retrieving the following information:

- A metadata description (including the WSDL URI) for all WaterOneFlow web services registered at HIS Central.
- A listing of all searchable keywords/concepts from the HIS Ontology.
- A metadata description for all data series that meet certain spatial, temporal, and variable search criteria.

HIS Desktop will use the methods from the HIS Central metadata catalog API to provide search capability across the metadata catalog to determine relevant data series for specific users. HIS Desktop will present users with a data discovery form that enables them to input the following search criteria. All of these criteria are optional, but at least one must be specified.

- A latitude/longitude bounding box to serve as the spatial constraint on the query. The box will be input by typing in coordinates, by drawing a rectangle on the HIS Desktop map, or by selecting a polygon feature from one of the layers in the HIS Desktop map (e.g., a watershed boundary – the extent of the feature would be converted to a latitude/longitude box).
- A searchable concept from the HIS Ontology (to be input by the user with autocompletion to suggest terms based on what a user types)
- A begin date and end date to serve as the temporal constraint on the query. If any part of the period of record for a data series overlaps this temporal window, the description of the data series will be returned.

- A minimum number of observations for the entire data record (only data series that have more than this minimum number for the entire data record will be selected, regardless of time window specified.)
- A list of WaterOneFlow web services to include in the search. This will be a user-specified subset of the web service registered at HIS Central that constrains search results to only a selected set of WaterOneFlow web services.

The result of a data discovery query using the HIS Central metadata catalog web service API will be the full metadata description for a listing of all of the data series cataloged at HIS Central that meet the search criteria. Because it is anticipated that the central metadata catalog web service API will respond quickly to data discovery queries, the results of queries against the central metadata catalog will not be cached by HIS Desktop. Instead, HIS Desktop will call the central metadata catalog web service API each time data discovery is performed. Additionally, because it is anticipated that there may eventually be metadata catalogs implemented for specific regions (e.g., the Texas HIS metadata catalog), the URI for the metadata catalog web service API will be user-customizable. HIS Desktop will support data discovery across multiple metadata catalog services provided that they implement the standard API.

3.1.2. DATA DISCOVERY DIRECTLY FROM WATERONEFLOW WEB SERVICES

While it is preferable that all WaterOneFlow web services be registered with HIS Central, it is anticipated that some services will exist that are not registered. Hence, HIS Desktop will support data discovery for individual WaterOneFlow web services using the following approach. HIS Desktop will harvest the Site, Variable, and Data Series metadata for one or more web services by calling the WaterOneFlow GetSites and GetSiteInfo methods repeatedly until all of the metadata have been retrieved (effectively building a local copy of the metadata served by the WaterOneFlow web service). The result of harvesting the metadata for an individual web service by calling it directly is essentially the same as the result of data discovery queries against the HIS Central metadata catalog – i.e., the metadata description is assembled for the list of data series published by the WaterOneFlow web service by parsing the WaterML results of the web service calls. However, because it is anticipated that harvesting the entire metadata contents of a WaterOneFlow service will require many web service calls and will potentially result in a much larger amount of information, the harvested metadata will be cached locally. Figure 2 shows the tabular schema that will be used by HIS desktop to store the metadata harvested from individual WaterOneFlow web services. HIS Desktop will use the cached metadata to allow users to select a subset of the contents of a harvested WaterOneFlow web service for download (see section 3.1.4 below).

Because WaterOneFlow web services that are not registered with HIS Central do not have mappings between variables and concepts from the HIS Ontology, data discovery directly from WaterOneFlow web services will not be keyword/concept based. Rather, it will involve simply harvesting all of the Site, Variable, and Data Series metadata for the service, caching that metadata locally, and allowing the user to make specific data requests based on this metadata. Because of this, variable metadata harvested directly from WaterOneFlow web services will be stored in the local metadata cache without linkages to ontology concepts. Additionally, variable metadata for data series discovered directly from the WaterOneFlow service (i.e., NOT discovered through HIS Central) and downloaded directly from WaterOneFlow web services will be stored in the data repository database without linkages to ontology concepts.
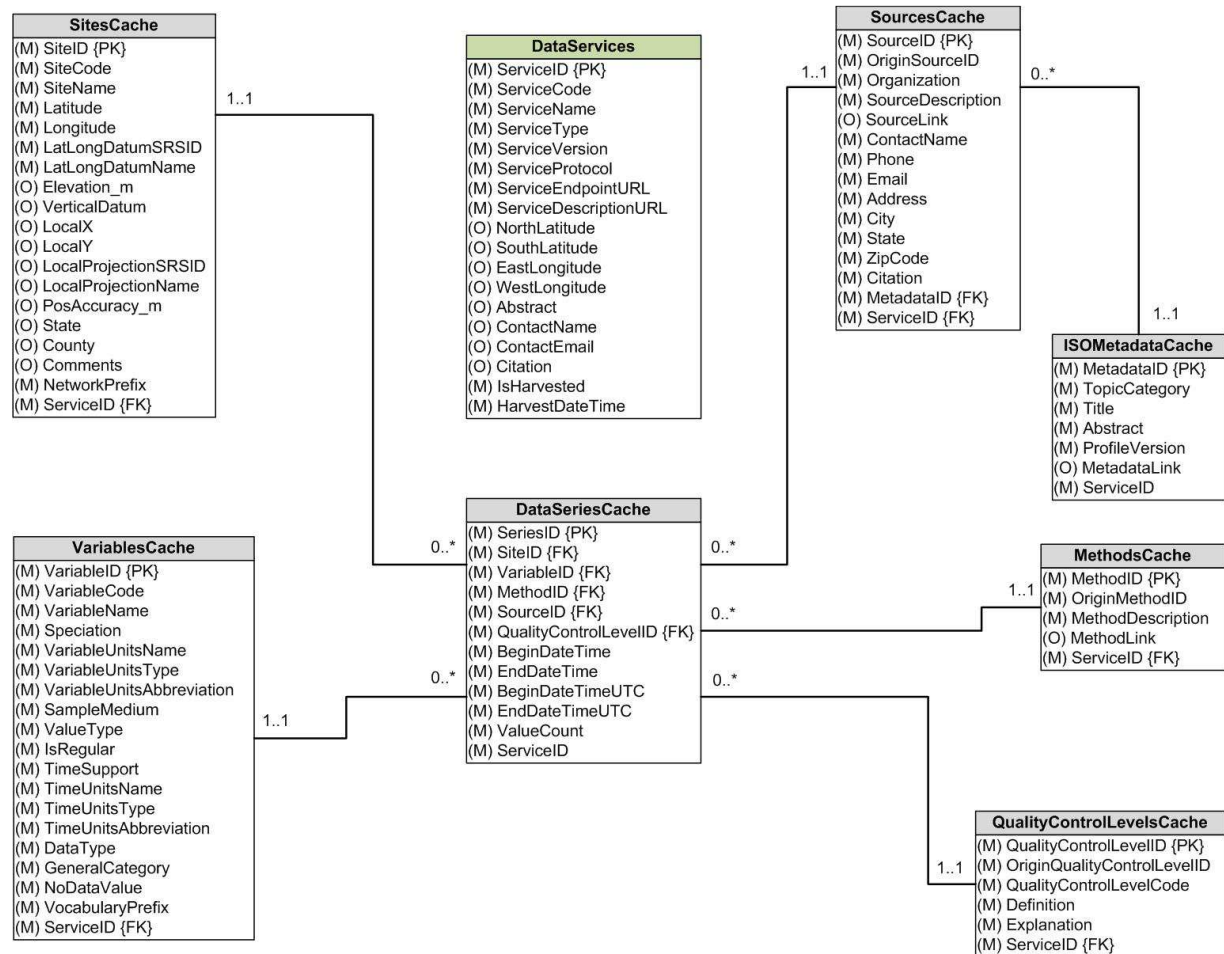
SitesCache
- (M) SiteID {PK}
- (M) SiteCode
- (M) SiteName
- (M) Latitude
- (M) Longitude
- (M) LatLongDatumSRSID
- (M) LatLongDatumName
- (O) Elevation_m
- (O) VerticalDatum
- (O) LocalX
- (O) LocalY
- (O) LocalProjectionSRSID
- (O) LocalProjectionName
- (O) PosAccuracy_m
- (O) State
- (O) County
- (O) Comments
- (M) NetworkPrefix
- (M) ServiceID {FK}

DataServices
- (M) ServiceID {PK}
- (M) ServiceCode
- (M) ServiceName
- (M) ServiceType
- (M) ServiceVersion
- (M) ServiceProtocol
- (M) ServiceEndpointURL
- (M) ServiceDescriptionURL
- (O) NorthLatitude
- (O) SouthLatitude
- (O) EastLongitude
- (O) WestLongitude
- (O) Abstract
- (O) ContactName
- (O) ContactEmail
- (O) Citation
- (M) IsHarvested
- (M) HarvestDateTime

SourcesCache
- (M) SourceID {PK}
- (M) OriginSourceID
- (M) Organization
- (M) SourceDescription
- (O) SourceLink
- (M) ContactName
- (M) Phone
- (M) Email
- (M) Address
- (M) City
- (M) State
- (M) ZipCode
- (M) Citation
- (M) MetadataID {FK}
- (M) ServiceID {FK}

ISOMetadataCache
- (M) MetadataID {PK}
- (M) TopicCategory
- (M) Title
- (M) Abstract
- (M) ProfileVersion
- (O) MetadataLink
- (M) ServiceID

DataSeriesCache
- (M) SeriesID {PK}
- (M) SiteID {FK}
- (M) VariableID {FK}
- (M) MethodID {FK}
- (M) SourceID {FK}
- (M) QualityControlLevelID {FK}
- (M) BeginDateTime
- (M) EndDateTime
- (M) BeginDateTimeUTC
- (M) EndDateTimeUTC
- (M) ValueCount
- (M) ServiceID

VariablesCache
- (M) VariableID {PK}
- (M) VariableCode
- (M) VariableName
- (M) Speciation
- (M) VariableUnitsName
- (M) VariableUnitsType
- (M) VariableUnitsAbbreviation
- (M) SampleMedium
- (M) ValueType
- (M) IsRegular
- (M) TimeSupport
- (M) TimeUnitsName
- (M) TimeUnitsType
- (M) TimeUnitsAbbreviation
- (M) DataType
- (M) GeneralCategory
- (M) NoDataValue
- (M) VocabularyPrefix
- (M) ServiceID {FK}

MethodsCache
- (M) MethodID {PK}
- (M) OriginMethodID
- (M) MethodDescription
- (O) MethodLink
- (M) ServiceID {FK}

QualityControlLevelsCache
- (M) QualityControlLevelID {PK}
- (M) OriginQualityControlLevelID
- (M) QualityControlLevelCode
- (M) Definition
- (M) Explanation
- (M) ServiceID {FK}

Figure 2. Tabular schema for the metadata cache resulting from harvests of individual WaterOneFlow web services.

WEB SERVICE WSDL URI LIST

A WaterOneFlow web service is currently implemented as a SOAP service where the WSDL (Web Service Definition Language) is exposed at a unique URI. HIS Desktop will maintain a user-customizable list of web service URIs that it can access for data discovery directly (i.e., without going through a central metadata catalog). Users will be able to select from this list of web services to specify services to be harvested for Site, Variable, and Data Series metadata. Web service URIs can be removed or added to the list manually or potentially from a centrally maintained listing of services.

In addition to the URI associated with each web service, the list of web services in HIS Desktop will also include the information shown for the DataServices table in Figure 2. When possible, HIS Desktop will populate the fields describing web services automatically. HIS Desktop will read and write this information to and from the DataServices table within the data repository database.

3.1.3.   DATA DISCOVERY FOR THEMATIC DATASETS

An individual or regional HIS Server can potentially host thematic time series data as feature datasets served using the OGC Web Feature Service (WFS) protocol. In the case that a user knows the direct URI to the WFS that defines

a thematic dataset, then this information can be used to retrieve all metadata associated with the observations contained therein. Once this metadata is retrieved, observation data may be downloaded for a subset of the theme datasets in the same way that data can be downloaded for any subset of metadata resulting from other search mechanisms.

3.1.4.   PROCESSING OF SEARCH RESULTS

Results of data discovery queries against the HIS Central metadata catalog and directly against a WaterOneFlow web service will be displayed in an interactive GUI from which the user can select a subset of specific data series to be retrieved from the server(s). Sub-setting/refining of data discovery query results will be supported by allowing users to do one or more of the following:

- Select a subset of sites returned by the data discovery query. This could be done from the HIS Desktop interactive map or from a tabular list of sites.
- Select a subset of variables from a list of all variables returned by the data discovery query.
- Select a subset of data series from a list of all data series returned by the data discovery query.

These screening/sub-setting options will be linked. For example, if a user selects a subset of sites, only the variables and data series associated with the selected sites will be shown for subsequent selection. To facilitate site selection using the interactive map, HIS Desktop will build and manage GIS layers of the sites returned by searches for display. Once a user has identified a set of data series for download, HIS Desktop will make GetValues calls against the WaterOneFlow web service(s) that host the data series to retrieve the data. Data values and all of their associated metadata will be downloaded and stored in the HIS Desktop data repository database. This mechanism has the potential to result in duplicate data series in the local data repository database. In the case that a user attempts to download a data series that already exists in their database, either in whole or in part, HIS Desktop will prompt the user to select an appropriate conflict resolution option. Details of the data download and conflict resolution functionality of HIS Desktop are described below.

3.1.5.   WORK FLOW FOR DATA DISCOVERY

DATA DISCOVERY USING THE HIS CENTRAL METADATA CATALOG

1. A user opens a search form within HIS Desktop and chooses an option to search the HIS Central Metadata Catalog.
2. The user populates the search criteria by doing one or more of the following:
    a. Using the HIS Desktop interactive map (see Data Visualization section below), the user zooms in to an area of interest.
    b. The user draws a box defining the spatial extent of the data search or selects a polygon feature from one of the layers in the map. Alternatively, the user types a bounding box into the search form.
    c. The user inputs/selects a searchable ontology concept from a predefined list of searchable concepts.
    d. The user inputs a time window for the search as a begin date and end date.
    e. The user selects a subset of the HIS Central WSDLs that are to be searched.
3. The user initiates the search.
4. HIS Desktop calls the HIS Central metadata catalog web service API and retrieves the metadata for sites, variables, and data series that meet the search criteria.

5. HIS Desktop builds GIS layer(s) that show the locations of sites from the search result.
6. The user can screen and subset the search results before downloading.
   a. The user clicks on sites on the map to see what variables/data series were returned by the search for each site.
   b. The user selects a subset of sites on a map for downloading data. Alternatively, the user selects from a list of returned sites.
   c. The user can narrow the search results by selecting a subset of the variables returned from a list. If a subset of sites has already been selected, only variables for selected sites will be displayed for selection.
   d. The user selects a subset of the data series that meet the site and variable filters for download by selecting series from a list.
7. The user specifies a local "theme" or shapefile/feature set with which the downloaded data should be associated, and into which the site/location data will be added. By default a new theme based on the ontological keyword(s) will be proposed as a new feature set for the data.
8. The user initiates the download of the selected data series into their local data repository.
9. HIS Desktop downloads the data into the local database and builds/appends the appropriate theme feature layers.

DATA DISCOVERY FROM INDIVIDUAL WATERONEFLOW SERVICES

1. A user opens a search form within HIS Desktop and chooses to search from one or more individual WaterOneFlow Services.
2. A user selects a one or more WaterOneFlow web services from the service WSDL URI list to be searched/harvested. If the URI does not exist in the list then the user can add it to the list.
3. The user initiates the search.
4. HIS Desktop checks to see if the WaterOneFlow web services have already been harvested. If they have, the user can choose to clear the metadata cache and do the harvest again, or to use the existing results.
5. HIS Desktop calls the WaterOneFlow web service(s) and retrieves and caches the metadata for sites, variables, and data series that are served by the selected WaterOneFlow web service(s).
6. HIS Desktop builds GIS layer(s) that show the locations of sites from the search result.
7. The user can screen and subset the search results before downloading.
   a. The user clicks on sites on a map to see what variables/data series were returned by the search for each site.
   b. The user selects a subset of sites on a map for downloading data. Alternatively, the user selects from a list of returned sites.
   c. The user can narrow the search results by selecting a subset of the variables returned from a list. If a subset of sites has already been selected, only variables for selected sites will be displayed for selection.
   d. The user selects a subset of the data series that meet the site and variable filters for download by selecting series from a list.
8. The user specifies a local "theme" or shapefile/feature set with which the downloaded data should be associated, and into which the site/location data will be added. Since there are no ontology concepts associated with variables in a WaterOneFlow web service, by default a new theme based on the name of the WaterOneFlow service will be proposed as a new feature set for the data.
9. The user initiates the download of the selected data series into their local data repository.

10. HIS Desktop downloads the data into the local database and builds/appends the appropriate theme feature layers.

3.2. DATA DOWNLOAD

The goal of the HIS Desktop data download functionality is to retrieve observational data series that have been identified for download using the data discovery tools described above and to create a local copy of the data in the desktop data repository database. HIS Desktop will also be capable of connecting to, displaying, and downloading GIS datasets published using OGC Web Feature Services (WFS) and Web Map Services (WMS). The architecture developed will support additional geosptatial data web services for future addition, such as Web Coverage Services (WCS) and OPeNDAP.

3.2.1. DOWNLOADING OBSERVATIONAL DATA

The result of data discovery (described above) is a set of metadata describing data series that have been identified by a user for download. Using this list, HIS Desktop will issue GetValues calls to retrieve each data series in WaterML format. HIS Desktop will save a copy of the result of each GetValues call as a WaterML formatted XML file on the user's hard drive in a user specified folder. XML files resulting from GetValues calls will be saved as part of a "Project Workspace" (see the section on Project Workspace that follows). Next, HIS Desktop will parse each of the WaterML results into the HIS Desktop data repository database. The purpose of saving the WaterML files is to preserve the data as they were retrieved from the web service when the GetValues call was made as part of data provenance. The purpose of loading the data into the data repository database is to facilitate and enable analysis and manipulation of the data.

The data repository database will have a relational structure, will be implemented within a relational database management system (RDBMS), and will serve as a local copy of the data that have been retrieved. The relational schema of the data repository database will be semantically similar to ODM with similar naming conventions and data types, but will be modified and extended to facilitate management of the data series that have been downloaded and storage of provenance information. Figure 3 shows the relational schema of the HIS Desktop data repository database.

The data repository database will be capable of storing the information encoded within WaterML files resulting from GetValues calls and will also support the storage of provenance information that includes the following list:

- From where was the data obtained (i.e., which web service)
- How was the data obtained (e.g., from a web service, from a local data import, or derived from an existing data series)
- The query that resulted in the data that was loaded (the definition of the GetValues call used to get the data)
- A pointer to the WaterML file from which the data originated (the file will be cached locally on disk)
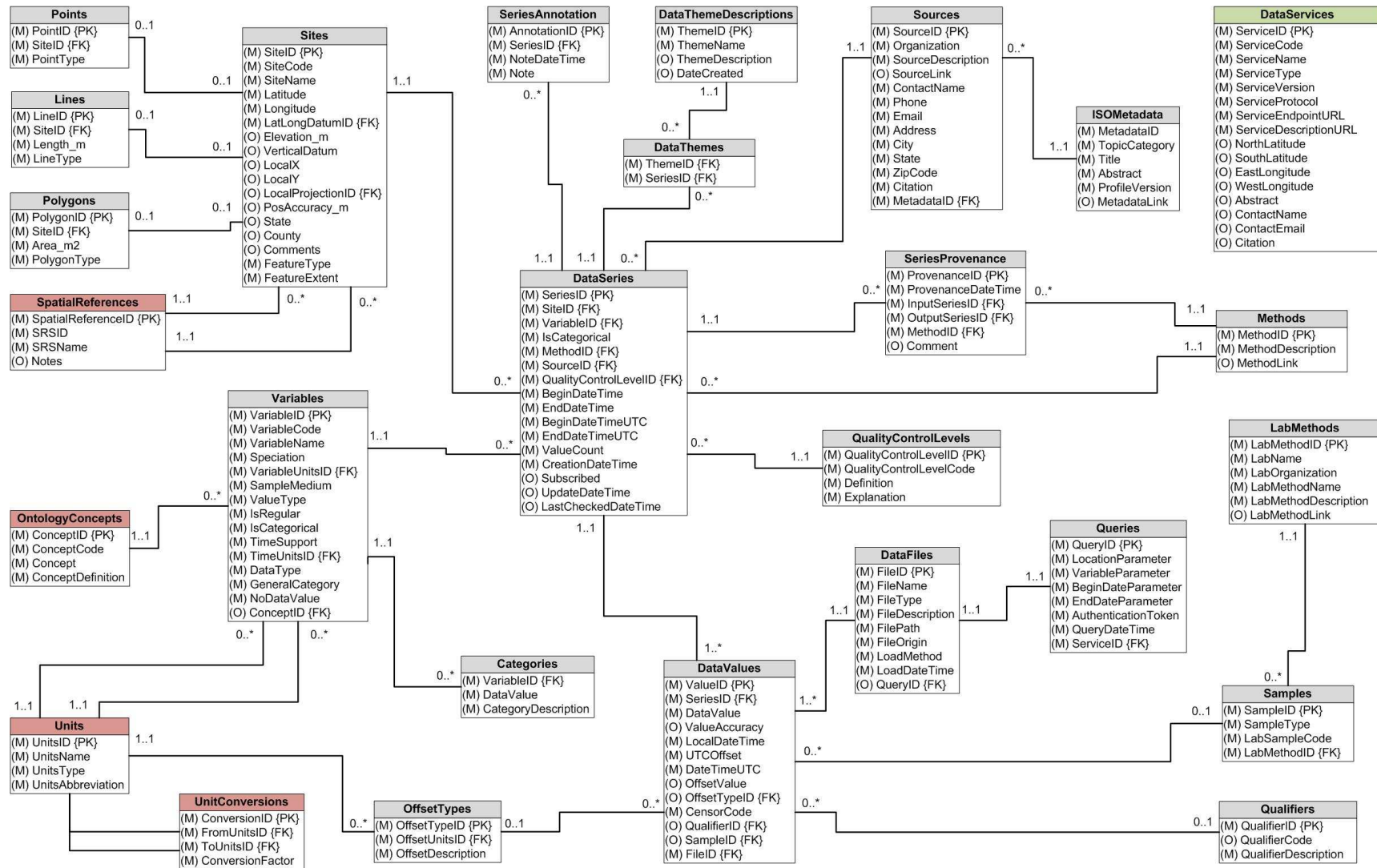- The date on which the data were loaded.

Points
- (M) PointID {PK}
- (M) SiteID {FK}
- (M) PointType

Lines
- (M) LineID {PK}
- (M) SiteID {FK}
- (M) Length_m
- (M) LineType

Polygons
- (M) PolygonID {PK}
- (M) SiteID {FK}
- (M) Area_m2
- (M) PolygonType

Sites
- (M) SiteID {PK}
- (M) SiteCode
- (M) SiteName
- (M) Latitude
- (M) Longitude
- (M) LatLongDatumID {FK}
- (O) Elevation_m
- (O) VerticalDatum
- (O) LocalX
- (O) LocalY
- (O) LocalProjectionID {FK}
- (O) PosAccuracy_m
- (O) State
- (O) County
- (O) Comments
- (M) FeatureType
- (M) FeatureExtent

SpatialReferences
- (M) SpatialReferenceID {PK}
- (M) SRSID
- (M) SRSName
- (O) Notes

Variables
- (M) VariableID {PK}
- (M) VariableCode
- (M) VariableName
- (M) Speciation
- (M) VariableUnitsID {FK}
- (M) SampleMedium
- (M) ValueType
- (M) IsRegular
- (M) IsCategorical
- (M) TimeSupport
- (M) TimeUnitsID {FK}
- (M) DataType
- (M) GeneralCategory
- (M) NoDataValue
- (O) ConceptID {FK}

OntologyConcepts
- (M) ConceptID {PK}
- (M) ConceptCode
- (M) Concept
- (M) ConceptDefinition

Units
- (M) UnitsID {PK}
- (M) UnitsName
- (M) UnitsType
- (M) UnitsAbbreviation

UnitConversions
- (M) ConversionID {PK}
- (M) FromUnitsID {FK}
- (M) ToUnitsID {FK}
- (M) ConversionFactor

SeriesAnnotation
- (M) AnnotationID {PK}
- (M) SeriesID {FK}
- (M) NoteDateTime
- (M) Note

DataThemeDescriptions
- (M) ThemeID {PK}
- (M) ThemeName
- (O) ThemeDescription
- (O) DateCreated

DataThemes
- (M) ThemeID {FK}
- (M) SeriesID {FK}

DataSeries
- (M) SeriesID {PK}
- (M) SiteID {FK}
- (M) VariableID {FK}
- (M) IsCategorical
- (M) MethodID {FK}
- (M) SourceID {FK}
- (M) QualityControlLevelID {FK}
- (M) BeginDateTime
- (M) EndDateTime
- (M) BeginDateTimeUTC
- (M) EndDateTimeUTC
- (M) ValueCount
- (M) CreationDateTime
- (O) Subscribed
- (O) UpdateDateTime
- (O) LastCheckedDateTime

Sources
- (M) SourceID {PK}
- (M) Organization
- (M) SourceDescription
- (O) SourceLink
- (M) ContactName
- (M) Phone
- (M) Email
- (M) Address
- (M) City
- (M) State
- (M) ZipCode
- (M) Citation
- (M) MetadataID {FK}

ISOMetadata
- (M) MetadataID
- (M) TopicCategory
- (M) Title
- (M) Abstract
- (M) ProfileVersion
- (O) MetadataLink

DataServices
- (M) ServiceID {PK}
- (M) ServiceCode
- (M) ServiceName
- (M) ServiceType
- (M) ServiceVersion
- (M) ServiceProtocol
- (M) ServiceEndpointURL
- (M) ServiceDescriptionURL
- (O) NorthLatitude
- (O) SouthLatitude
- (O) EastLongitude
- (O) WestLongitude
- (O) Abstract
- (O) ContactName
- (O) ContactEmail
- (O) Citation

SeriesProvenance
- (M) ProvenanceID {PK}
- (M) ProvenanceDateTime
- (M) InputSeriesID {FK}
- (M) OutputSeriesID {FK}
- (M) MethodID {FK}
- (O) Comment

Methods
- (M) MethodID {PK}
- (M) MethodDescription
- (O) MethodLink

QualityControlLevels
- (M) QualityControlLevelID {PK}
- (M) QualityControlLevelCode
- (M) Definition
- (M) Explanation

LabMethods
- (M) LabMethodID {PK}
- (M) LabName
- (M) LabOrganization
- (M) LabMethodName
- (M) LabMethodDescription
- (O) LabMethodLink

Categories
- (M) VariableID {FK}
- (M) DataValue
- (M) CategoryDescription

DataValues
- (M) ValueID {PK}
- (M) SeriesID {FK}
- (M) DataValue
- (O) ValueAccuracy
- (M) LocalDateTime
- (M) UTCOffset
- (M) DateTimeUTC
- (O) OffsetValue
- (O) OffsetTypeID {FK}
- (M) CensorCode
- (O) QualifierID {FK}
- (O) SampleID {FK}
- (M) FileID {FK}

DataFiles
- (M) FileID {PK}
- (M) FileName
- (M) FileType
- (M) FileDescription
- (M) FilePath
- (M) FileOrigin
- (M) LoadMethod
- (M) LoadDateTime
- (O) QueryID {FK}

Queries
- (M) QueryID {PK}
- (M) LocationParameter
- (M) VariableParameter
- (M) BeginDateParameter
- (M) EndDateParameter
- (M) AuthenticationToken
- (M) QueryDateTime
- (M) ServiceID {FK}

Samples
- (M) SampleID {PK}
- (M) SampleType
- (M) LabSampleCode
- (M) LabMethodID {FK}

OffsetTypes
- (M) OffsetTypeID {PK}
- (M) OffsetUnitsID {FK}
- (M) OffsetDescription

Qualifiers
- (M) QualifierID {PK}
- (O) QualifierCode
- (M) QualifierDescription

Figure 3. HIS Desktop data repository database schema.

The following "best practices" will be followed by HIS Desktop in managing the data repository database:

- If and when data series are deleted, the changes are cascaded to related tables so that the data repository database remains consistent (e.g., if a data series for a site is deleted and no other data series exist for that site, then the site is deleted as well).
- Edits and/or transformations of data series will result in new data series rather than overwriting or editing existing data series. Provenance for transformations will be tracked at the series level.

The data repository database will also be part of the "Project Workspace." HIS Desktop will be capable of saving/connecting to multiple "Project" related databases (one per project).

3.2.2.   DOWNLOADING GIS DATASETS

While there is currently no HIS developed/supported discovery mechanism for geospatial datasets, HIS Desktop will be capable of viewing and downloading GIS data that are published using OGC web feature services (WFS) and web mapping services (WMS). As WMS services are image services, they will only be viewable in HIS Desktop (i.e., data download is not possible). Specifically, HIS Desktop will enable users to: 1) enter a URI for a GIS data service and view the data published by that service in the HIS Desktop interactive map (see the Data Visualization section below), and 2) for WFS, allow users to download the data from the service and create a local copy that can be manipulated and used for analyses. GIS data published as WFS will be downloaded and stored within the Project Workspace as shapefiles.

3.2.3.   WORK FLOW FOR DATA DOWNLOAD

DOWNLOADING OBSERVATIONAL DATA

1. A user completes data discovery as described above and has identified a list of data series to be downloaded.
2. The user initiates the download.
3. For each data series, HIS Desktop checks to see if the data series already exists in the data repository database. If so, the user is presented with the option of: 1) replacing the existing data series entirely (e.g., deleting the existing series and adding the newly download series to the database), 2) appending new data to the existing data series (e.g., the existing series is left alone and only new data values are added to the database), or 3) adding the newly downloaded data series as a new data series to the database (e.g., creating an updated "copy" of the data series in the database).
4. For each data series, HIS Desktop performs the appropriate GetValues call to retrieve the data series.
5. For each data series, HIS Desktop saves the WaterML result to an XML file in the Project Workspace.
6. For each data series, HIS Desktop parses the WaterML result and associated provenance information into the desktop repository database.
   a. If the data series is new, it is created in the database along with its metadata and provenance information.
   b. If the data series is to be replaced, HIS Desktop first checks to see if any data series have been derived from the data series to be deleted. If so, the user is prompted to make sure that they want to delete the data series and break the provenance linkage. The data series is then deleted from the database and then re-created using the newly downloaded result. Provenance associated with the deleted data series is also deleted.

c. If the data series it to be updated, only new data values from the new GetValues call are added to the database and the existing data is left alone. Data value and data series provenance information is updated.

d. If the data series is to be added to the database without deleting or modifying the existing series, the data series is added to the database as a new data series with a new CreationDateTime and where the data values for the new data series are associated with the newly downloaded WaterML file.

DOWNLOADING GIS DATASETS

1. A user clicks on a tool to add an online data layer to the HIS Desktop interactive map.
2. The user inputs the URI of the spatial data server.
3. The user selects the layers/services from the spatial data server that they want to add to the map.
4. The user selects whether the layers should be downloaded and copied to local files on their hard disk.

3.3. DATA VISUALIZATION

HIS Desktop will support visualization of both geospatial and time series data. Geospatial data visualization will be enabled through an interactive GIS map. Visualization of observational data will be provided through a variety of plots and will be focused on exploratory data analysis for data series that are downloaded and stored in the HIS Desktop data repository.

3.3.1. VISUALIZATION AND ANALYSIS OF SPATIAL DATA

The HIS Desktop interactive map will be used for displaying and manipulating spatial datasets as well as for setting the context for data discovery. As described in the sections above, an area of interest is often used as a spatial filter for narrowing a search for data. Therefore, the HIS Desktop interactive map will be used to enable the user to set the geographic context for data discovery and access by enabling users to draw a bounding box or select a polygon feature from one of the GIS layers in the map (e.g., state boundaries, watershed boundaries, etc.) within which they would like to conduct their search.

HIS Desktop will use the customizable MapWindow GIS to provide users with the ability to visualize and manipulate spatial datasets. MapWindow supports a variety of vector, raster, and image GIS data types, and includes functionality for navigating the map as well as many other GIS tools and features. All of the functionality provided by MapWindow for the visualization and manipulation of GIS datasets will be available within HIS Desktop.

SPATIAL DATA LAYERS

HIS Desktop will provide access to the following spatial data types via MapWindow:

- Vector data as shapefiles
- Raster data as grids or images
- Online data as OGC WFS and WMS

HIS Desktop will include the following base map layers pre-loaded within the interactive map:

- World political boundaries and oceans

- USA state boundaries
- USA county boundaries
- Major cities
- Major roads
- Major rivers
- Major water bodies
- Boundaries for USGS 8-digit Hydrologic Unit Code (HUC) areas.

The datasets behind these layers will be stored locally, so that the basemap can be displayed even when the desktop computer is not connected to remote mapping resources. HIS Desktop will allow users to toggle layer visibility, change layer symbology, remove layers, and add layers from disk. This will permit users to add local GIS datasets that may further define a study area of interest.

GIS TOOLS

HIS Desktop will include the following tools for map navigation and GIS analysis:

- Pan
- Zoom in and out
- Zoom to the full extent of a given map layer
- Zoom to the full extent of all data in the map
- Identify a feature in the map (i.e., show attributes of the clicked feature)
- Select one or more features in the map (for layers supporting data retrieval from HIS)
- Indicate current position in latitude/longitude
- Draw a box to set latitude/longitude bounds for data discovery (see Data Discovery section)
- Zoom to previous extent / zoom to next extent (similar to forward and back buttons on a web browser)
- Measure the distance between two points
- View and query the attribute table for a GIS dataset
- Export copies or subsets of map layers

3.3.2. VISUALIZATION AND ANALYSIS OF OBSERVATIONAL DATA

Once time series of observational data have been retrieved and stored in the desktop data repository database, HIS Desktop will provide users with a tool for visualizing and analyzing the data. HIS Desktop will maintain one or more GIS data layers showing the locations of the sites for which data have been downloaded to the desktop data repository database. These layers will be dynamically built/updated from the data repository database each time data are downloaded and may be organized around the data themes that users have assigned to data series when they were downloaded. Users will be able to select a site on the interactive map and launch the time series visualization and analysis tool with data populated for the selected site. A variety of plot types will be available for visualizing time series data at a selected site. These include time series, histogram, box-and-whisker, and probability plots for a selected time series. The HIS Desktop time series visualization and analysis tool will also enable users to view a selected time series in a simple tabular view as well as calculating simple descriptive statistics (minimum, maximum, mean, percentile values, etc.) for the selected time series.

In addition to the HIS Desktop visualization and analysis tool, users will be able to connect to and extract data from the HIS Desktop data repository database from 3rd party analysis software such as R, MATLAB, or Microsoft Excel

(see Figure 1). This functionality will facilitate users doing analysis within the software program that they are most familiar with, after having used HIS Desktop to download and organize the data.

3.3.3.   WORK FLOW FOR VISUALIZING OBSERVATIONAL DATA

1. A user launches the time series data visualization tool by clicking on a button on the toolbar or by clicking on a monitoring site for which data have been downloaded using the interactive map.
2. The user selects a variable that they would like to visualize.
3. The user selects a data series that they would like to visualize.
4. The user selects a time range for the data that they would like to visualize.
5. The HIS Desktop visualization tool generates a set of plots and descriptive statistics for the selected data series.
6. The user can click on a button to see the selected data series in a tabular view.
7. The user can click a button to export the data in the current view.

3.4. DATA IMPORT AND EXPORT

HIS Desktop will enable the import of datasets from a user's local hard drive as well as export of data from HIS Desktop to a number of common data formats.

3.4.1.   IMPORTING AND EXPORTING SPATIAL DATASETS

HIS Desktop will use the GIS Data import and export tools available in MapWindow to support adding spatial datasets to the HIS Desktop interactive map as well as exporting copies or subsets of data layers to a user's hard drive. Users will be able to add GIS data stored on their local hard drive to the HIS Desktop map, provided that the data are stored in formats compatible with MapWindow. GIS data formats supported by MapWindow are described in the Data Visualization section above. Users will also be able to import GIS data from OGC spatial data services as described in the Data Download section above. Users will be able to select a subset of the feature and raster data from a GIS dataset and export it from HIS Desktop to their local hard drive as a separate GIS dataset.

3.4.2.   IMPORTING AND EXPORTING OBSERVATIONAL DATASETS

HIS Desktop will provide a tool for importing observational datasets from a user's hard drive into the data repository database. The observational data import tool will accept data "packages" in a predefined format for import. Users must manipulate their data into this format before it can be imported into the data repository. The data import tool will validate the data package created by the user and automatically parse it into the desktop data repository database.

Users will be able to export data series from the desktop data repository database to a number of predefined file formats using a data export tool. The data export tool will allow users to perform queries on the desktop data repository database using a number of different criteria to identify data series for export. Query criteria will include: Site, Variable, Source, QualityControlLevel, and Method. Once identified for export, data series will be written to one of several predefined export data formats.

Implementation Note: Some of the source code needed for these functions is available within ODM Tools, the ODM Data Loader, and the ODM Streaming Data Loader. We will, to the extent possible, make use of existing code for importing observation data as well as for selecting data from the local database for export.

3.4.3. WORK FLOW FOR DATA IMPORT AND EXPORT

IMPORTING SPATIAL DATASETS FROM A USER'S HARD DRIVE

1. A user clicks on an "Add Layer" button in HIS Desktop or selects "Add Layer" from a toolbar drop down menu.
2. The user navigates to the location of the GIS layer to be added on their hard drive.
3. The user selects the layer to be added.
4. HIS Desktop adds the layer to the HIS Desktop map.
5. The user can then create an appropriate legend for the layer by editing the way it is symbolized.

EXPORTING SPATIAL DATASETS TO A USER'S HARD DRIVE

1. A user clicks on a layer in the interactive map's legend to make it active.
2. The user clicks on a "Select Features" toolbar button.
3. The user selects a number of features in the active layer by clicking on them, shift-clicking for multiple features, or dragging a box for multiple feature selection.
4. The user clicks on an "Export Features" tool to export the selected features to a new GIS layer.
5. The user selects whether to export all or only selected features.
6. The user navigates to the location on disk where the new layer should be stored.
7. HIS Desktop writes the new GIS data layer to disk.

IMPORTING OBSERVATIONAL DATA FROM A USER'S HARD DRIVE

1. A user creates a "package" of data to be loaded into the data repository database according to a predefined data import template.
2. The user clicks on a "Data Import" toolbar button or selects "Import Observational Data" from a toolbar menu.
3. The user navigates to the location on disk where the data package is located.
4. The user selects the data package and clicks on an "Import" button.
5. HIS Desktop validates the data to be imported and reports any errors to the user.
6. HIS Desktop imports the data package into the data repository database.

EXPORTING OBSERVATIONAL DATA TO A USER'S HARD DRIVE

1. A user clicks on an "Export Data" toolbar button or selects "Export Observational Data" from a toolbar menu.
2. HIS Desktop presents the user with a query form on which the user can specify a number of different criteria for identifying data series for export.
3. The user inputs query criteria, executes the query, and then selects a subset of the resulting data series that he/she wants to export.
4. The user clicks on an "Export" button.
5. The user navigates to a location on disk to store the exported data file.
6. The user chooses one of the predefined export formats for the exported data.

7. HIS Desktop writes the selected data to the user's hard drive in the selected format.

3.5. PROJECT WORKSPACE

HIS Desktop will allow users to create and save any number of "Project Workspaces." A Project Workspace is a user-defined, logical spatial or thematic grouping of data (i.e., the "Little Bear River" project or the "environmental flows" project). A Project Workspace will be a folder on a user's computer that contains project and configuration files generated by HIS Desktop. These files will store the definition of the GIS layers that have been added to the interactive map, the symbology of the layers, the definition of which plug-ins are currently loaded, database connection information, and other important project settings. The desktop data repository database and cached WaterML files will also be stored within the Project Workspace. Users will also have the option of storing all GIS data that has been downloaded using HIS Desktop in the Project Workspace. However, geospatial data files that exist outside this workspace can also be added to the HIS Desktop project through a linkage defined in an HIS Desktop project file.

3.5.1. WORK FLOW FOR CREATING A PROJECT WORKSPACE

1. A user opens HIS Desktop.
2. The user selects "Create New Project" from a pull down menu or clicks on a "Create New Project" toolbar button.
3. The user navigates to the location on disk where the new Project Workspace should be created.
4. The user gives HIS Desktop a name for the Project Workspace.
5. HIS Desktop creates a folder and any necessary subfolders and files for storing the Project Workspace.
6. HIS Desktop creates a blank data repository database for the Project Workspace.
7. The user uses the search and download tools of HIS Desktop to populate the Project Workspace with data.
8. The user uses the data import tools to import data from their local hard drive to the Project Workspace.
9. The user clicks on a "Save Project" toolbar button or selects "Save Project" to save all of the settings for their project to the Project Workspace.

3.6. PLUG-IN INTERFACE

HIS Desktop will support a plug-in interface to ensure that it is extensible. The HIS Desktop plug-in interface will allow the HIS development team and other developers to extend the software for unique and particular purposes without participating in the core HIS Desktop code development. Some of the functionality described in the previous sections may be built as HIS Desktop plug-ins (e.g., time series visualization and analysis, data discovery, data import/export). Additional plug-ins *that are beyond the scope of HIS Desktop Version 1.0* but that could be built using this plug-in interface include:

- **Data Series Transformation Plug-in** – This plug-in would provide users with the ability to create new data series from existing data series stored in their data repository database. Derivation of new data series would include conversion of units and data aggregation (e.g., derive daily average values from more frequent observations). The ability to add new transforms could also be provided.
- **NetCDF/UNIDATA Data Provider Plug-in** – This plug-in would provide access to NetCDF and UNIDATA data sets.
- **OpenMI Modeling Plug-in** – This plug-in would integrate the OpenMI modeling interface with the data downloaded using the HIS Desktop data discovery and download tools. This plug-in could also make all of

the GIS capabilities of MapWindow available to model components running within the OpenMI modeling plug-in.

- **Workflow Plug-in** – A workflow processing plug-in would enable users to set-up and execute scientific work-flows that use data and functionality that are part of HIS Desktop.
- **Spatial Data Discovery and Download Plug-in** – A plug-in could be developed for HIS Desktop that would enable users to search for and download GIS datasets, including those from national data providers (e.g., NED, NLCD, NHD) as well as those published as OGC compliant spatial data services. The functionality of this plug-in would be similar to the discovery and download functionality described above for time series of observational data.
- **BASINS Data Downloader** – The USEPA BASINS project includes several plug-ins for geospatial data download that follow the MapWindow plug-in interface and could be made to work with HIS Desktop.
- **Scripting Plug-in –** MapWindow 4.0 includes a scripting interface that allows users to write custom scripts and plug-ins directly within the MapWindow GIS GUI. This capability could be ported to a plug-in for HIS Desktop such that users can script their workflows and analytical processes.

HIS Desktop will use the existing MapWindow plug-in interface as well as a specific HIS plug-in interface that defines objects, methods, and events that are unique to HIS Desktop and are not included in MapWindow. As such, plug-ins will be capable of interacting with the underlying HIS Desktop datasets (e.g., the HIS Desktop data repository database and the GIS data layers). This approach effectively provides a database abstraction layer shown as the "database API" in Figure 1 – specifically providing needed methods for the most common data manipulation tasks (e.g. table reading and writing). Plug-ins will also be capable of interacting with the core components of the MapWindow interactive GIS map interface. For example, plug-ins will be capable of adding menus and toolbar buttons to HIS Desktop that are tied to specific functionality. Plug-ins will be capable of manipulating the map, layers within the map, and the map legend. Finally, plug-ins will have access to all of the core GIS functionality provided by MapWindow (e.g., shapefile creation and editing, grid creation and editing, etc).

4. TECHNICAL REQUIREMENTS

The following sections detail specific technical requirements for HIS Desktop.

4.1. DEVELOPMENT ENVIRONMENT AND SOURCE CODE

HIS Desktop will be developed in the Microsoft Visual Studio 2008 Interactive Development Environment. The language of HIS Desktop and its associated plug-ins will be C# or Visual Basic. HIS Desktop will be developed as an open-source software product, and, as such, its source code will be made freely available according to the CUAHSI HIS software policy.

4.2. OPERATING SYSTEM SUPPORT

All of the functionality of HIS Desktop Version 1.0 will be supported on Windows (XP and Vista) and Mac OS X operating systems. HIS Desktop will be compiled under the Microsoft .Net framework 3.5 or higher for installation on Windows based computers. It will also be compiled under the Mono framework v 2.1 or higher for installation on Macintosh based computers. HIS Desktop will not require commercial software beyond the operating system.

4.3. DATABASE SUPPORT

HIS Desktop will be developed with native/default support for SQLite (www.sqlite.org). Additionally, the proposed plug-in interface will allow for the future addition of support for other RDBMSs such as Microsoft SQL Server, MySQL, and others. The HIS Desktop database will avoid the use of specific data types that are not widely supported and would preclude implementation of the database in a variety of RDBMSs.

4.4. THIRD PARTY SOFTWARE COMPONENTS

Rather than recreating specific functionality that can be obtained through free or inexpensive third party software development components, HIS Desktop will use existing components where possible. HIS Desktop will use only freely available or open source components so that developers who wish to edit or recompile the source code for HIS Desktop will not have to purchase licenses for any third party software components beyond the operating system.

4.5. USER INTERFACE REQUIREMENTS

HIS Desktop will be a graphical user interface (GUI) based application. It will consist of a set of forms that will provide users with a GUI to the observational and GIS data that they assemble using the HIS Desktop tools. HIS Desktop will be focused primarily around a map-based interface. HIS Desktop will be menu and toolbar button driven, and will not require any user programming to use its functionality.

4.6. INSTALLATION AND CONFIGURATION

HIS Desktop will be deployed via appropriate automated installation package(s) that can be distributed via compact disk or downloaded from the CUAHSI HIS/HIS Desktop website. Installation instructions will be included for both Windows and Mac OS X in the HIS Desktop user documentation. The software installation package(s) will automatically install all of the necessary components and files for the HIS Desktop application to work, and it will also install a user manual in Adobe PDF format that details the functionality of HIS Desktop.

5. INTERFACES AND DEPENDENCIES

The purpose of this section is to identify the interfaces between HIS Desktop, other components of the CUAHSI HIS, 3rd party software, and any dependencies that must be in place to facilitate the full functionality describe above. The following interfaces and dependencies have been identified:

1. *Map Window version 6.x* – This version of MapWindow is required for HIS Desktop specifically to support the ability to compile and run on Mac OS X.
2. *WaterOneFlow Web Service API* – HIS Desktop will communicate with and retrieve data and metadata from WaterOneFlow web services that are served from distributed HIS Servers and from HIS Central (in the case of national datasets like USGS NWIS and USEPA STORET). The WaterOneFlow web service API will be relied upon to deliver data values and all of their associated metadata from an observational data service.
3. *HIS Central Metadata Catalog Web Service API* – HIS Desktop will perform data discovery queries against the contents of the HIS Central metadata catalog using the HIS Central Metadata Catalog Web Service API.

This API will be relied upon to deliver a listing of all searchable ontology keywords/concepts, a metadata description for all WaterOneFlow web services registered at HIS Central, and a metadata description for all data series hosted by registered WaterOneFlow web services.

NOTE: This critical element of the project that is under development at the present time.

4. ***OGC WFS and WMS API*** – HIS Desktop will rely on the OGC Web Feature Service (WFS) API for consuming GIS datasets and Themed Datasets that are published using WFS. HIS Desktop will also rely on the OGC Web Map Service (WMS) API for consuming GIS datasets that are published using WMS.
5. ***3rd Party Analysis Software*** – *The* HIS Desktop data repository database will support connections from 3rd party software such as Visual Basic, MATLAB, R, or Excel.

6. COMMUNITY CODE DEVELOPMENT SYSTEM AND MANAGEMENT

In addition to the core software design and development tasks that will be required in the development of HIS Desktop, a key activity and goal will be the development of an online community of users/developers who will jointly design, code, bug-test, and deploy the HIS Desktop software. The following section describes the functionality of components that will support the activities of this online community.

6.1. HIS DESKTOP WEB SITE

A dedicated website will be developed to support community code development, bug tracking, and forum discussions related to HIS Desktop and related plug-ins, software, and tools. The HIS Desktop website will be integrated with the existing CUAHSI HIS website (http://his.cuahsi.org). The website will serve as a workspace for individuals involved in the development of HIS Desktop and will be used to encourage external project participants including programmers, end-users, testers, and others drawn largely from the academic hydrologic sciences community as well as commercial entities and government agencies. The HIS Desktop website will include links to the CUAHSI HIS website, but will be focused on the development and maintenance of the user community and software associated with HIS Desktop. The site will include clear statements on the licensing under with the software is developed and deployed (currently the BSD license according to the CUAHSI HIS software policy), as well as instructions on how to participate in the development of HIS Desktop. The following collaborative software development tools will be available on the HIS Desktop website:

- Integrated Discussion Forum(s) – for sharing insights, questions, comments, and general discussion information.
- Source Code Repository – Subversion based code management repository for sharing source code. This will be set up such that anyone can read the code, but only authorized personnel can commit code.
- Coordinated Bug Management and Feature Identification – This will likely use the Mantis bug tracking system and will allow any user of HIS Desktop to post bug notices and feature requests.
- Downloadable Software – This will include a list of downloadable installation packages and plug-ins developed by the HIS project team as well as the open source community.