EPA Continuous Monitoring Data Strategy: Synopsis

IOOS DMAC

29 May 2015

Interviews- New Jersey Department of Environmental Protection

TCEQ HIGHLIGHTS

Texas CEQ Highlights

- TDS early warning for irrigators
- Guiding and assessment of remediation activities; assessment of WQ attainment
- Emergency preparedness, recreational activities (flows)
- Assessment of sediment NPS loads
- Public engagement

Texas CEQ Highlights

- Moving from LEADS (air quality!) and MANVAL to AQUARIUS
- Provisional data published within 30-60 minutes, manually validated data with 150 days.
- Pre- and post-deployment QC are necessary, not sufficient.
- Sensors deployed by partners (GBRA, USGS)

Interviews- New Jersey Department of Environmental Protection

NJDEP HIGHLIGHTS

New Jersey DEP Highlights

- Buoys and airborne sensor to support shellfish management
- Glider deployed to improve marine assessments of attainment
- Continuous data in streams and rivers used for criteria and TMDL development, source water protection, long-term trends
- QAPPs required for assessment data

New Jersey DEP Highlights

- Some data available near-realtime with lower QC, all available in 3-9 months with full QC.
- Marine data currently available through web portal, freshwater data will also be added.
- Other:
  - "Continuous Monitoring could use EPA's continuous support.
  - Get USGS to put more emphasis on WQ.

Interviews – Integrated Ocean Observing System

IOOS HIGHLIGHTS

IOOS Highlights

- Coordinates ocean data from buoys, gliders, satellites, models for monitoring, assessment and management of HABs, acidification, coral reefs, beach closures, biodiversity
- Most data are realtime with automated QC, some climate-related data published on a delayed basis
  - Certain data published worldwide through GTS

IOOS Highlights

- Hybrid data management framework includes both centralized and distributed elements.
  - How much should Data Management and Communications reflect governance structure?
- IOOS is more operationally oriented, consistent with its mission.
  - Cost per sample is lower for sensor data
  - Currently implementing automated QC: Quality Assurance of Real-Time Ocean Data

Interviews – Integrated Ocean Observing System

USGS HIGHLIGHTS

USGS Highlights

- Nutrients
  - Great Lakes tributaries: nutrient surrogates
  - Iowa: source water protection
  - Mississippi River: nitrate photometers
- Fisheries: temperature monitoring to improve understanding of fish movements
- Other applications
  - Road salt, lake turnover, manure spreading

IOOS Highlights

- Generally homogenous hardware
  - Campbell Scientific CR1000, CR800, CR10
  - Some level loggers, OnSet Hobos
- Extensive QA, QC and management toolset
  - Continuous Record Plan (CRP) and Techniques and Methods (T&M) documents
  - ADAPS, GRSat, DCT, RDT, NWIS RA, QW, AQUARIUS, CHIMP, SITES, …

Small Thoughts

LATENCY AND QC

Latency, Quality and Use

	Lower QC	Higher QC
Low Latency Available within minutes or hours	**“Operational”** • Automated QC • Useful for trends • Exceedance of an action limit may not be a violation	***Unrealistic?***
High Latency Available weeks or months after	***Useless?***	**“Assessment”** • Manual QC • Can be compared to WQ standards

Small Thoughts

DESIRABLE ATTRIBUTES

Desirable Attributes

- **Discoverability**, **transparency** and **interoperability** are data management virtues.
  - Development and publication of metadata in accessible catalogs.
  - Use of appropriate detail and a controlled vocabulary in metadata and datasets.
  - Publication of data using open standards.
- **Provenance, scalability** and **sustainability** too.

Small Thoughts

DEFINING “CONTINUOUS”

What we've heard …

- Hourly or better; minimal human involvement/automated (TWDB)
- Demonstrates serial correlation; reflects/captures diurnal cycles (TCEQ)
- Usually no less than hourly frequency; deployed for at least 72 hours (NJDEP)
- Unattended/autonomous; metadata less variable than data; considers Nyquist limit (IOOS)
- In-situ time series in-situ, cost-effective (USGS)

Working definition?

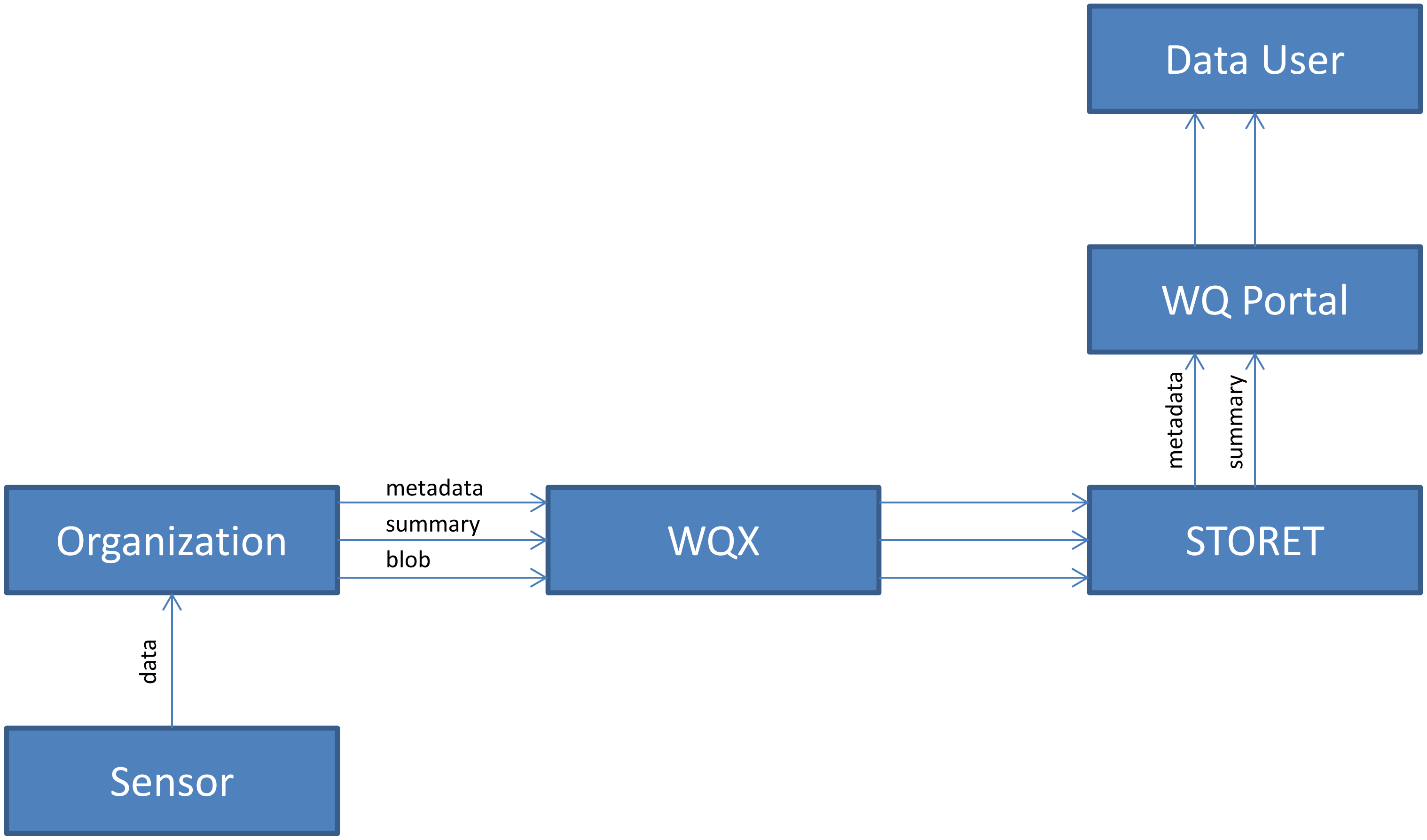
- A time series of observations taken at regular intervals with an automated sensor:
  - Sensor may be fixed or moving (glider/AUV)
  - Data may be delivered in real time or logged for later retrieval
  - Metadata are invariant for a deployment (?)
  - QC information may be observation-specific (?)
- What about automated samplers?

SOME EXISTING ALTERNATIVES

Some Existing Alternatives

- Status Quo: WQX, STORET, Water Quality Portal
- USGS/NWIS
- USEPA/AirNow
- CUAHSI/HIS
- NOAA/IOOS

Status Quo



Data User
WQ Portal
metadata
summary
Organization
metadata
summary
blob
WQX
STORET
data
Sensor

Status Quo

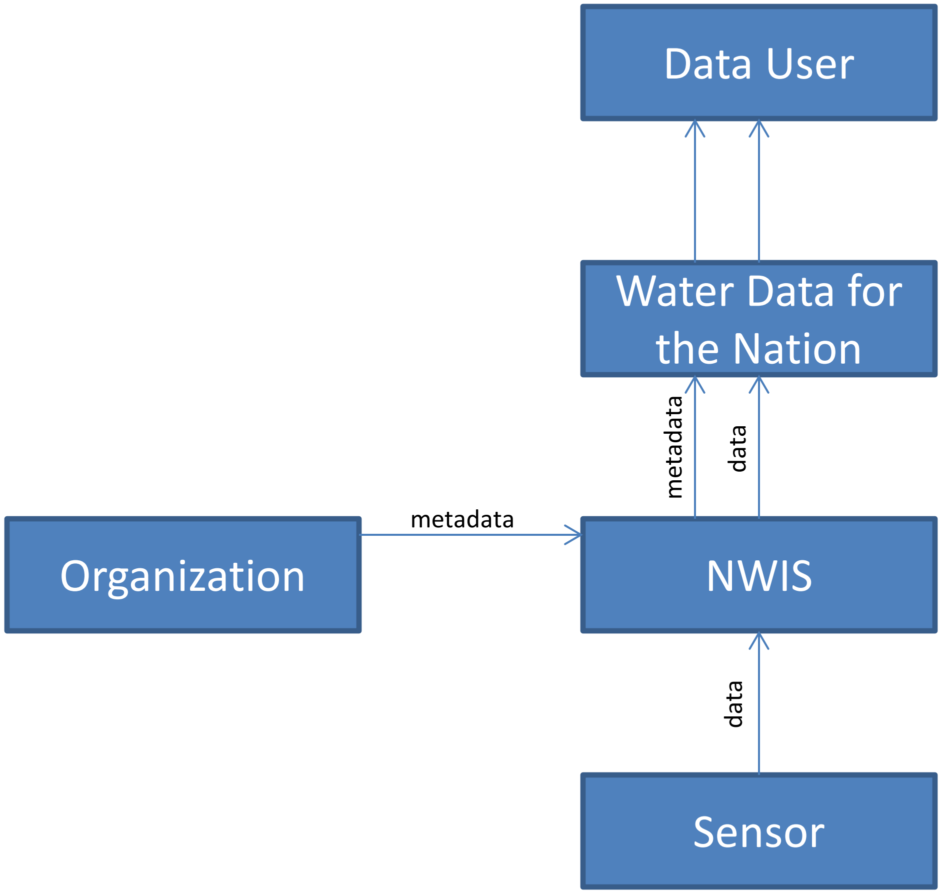
Pro

- Familiar
- Existing support and sustainability

Con

- Clumsy to enter
- Inefficient storage
- Difficult to discover and access time series data

USGS NWIS (AQUARIUS Buildout)



USGS NWIS

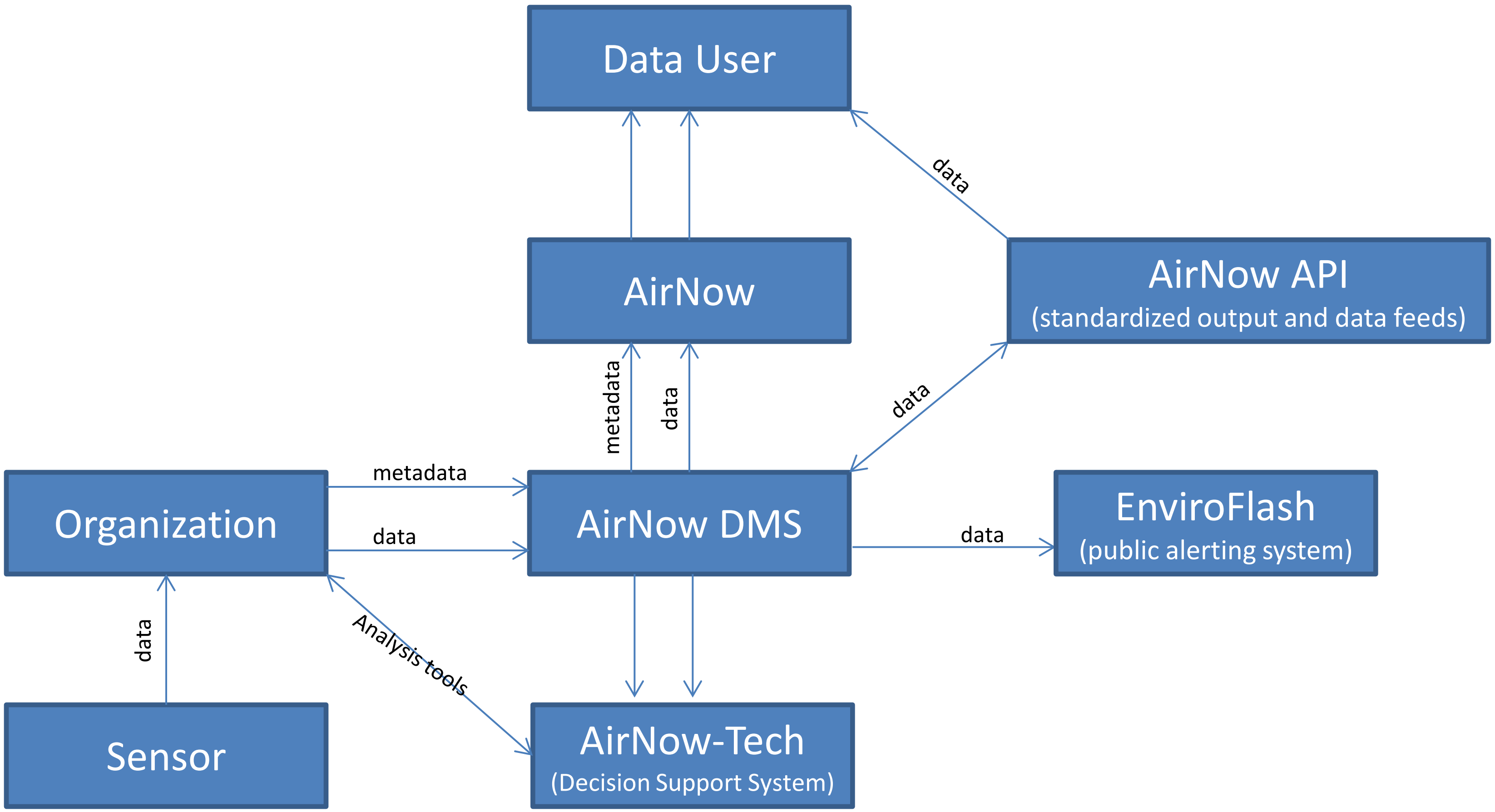
Pro

- Centralized
- Standardized
- Well-established

Con

- Rigorous QA and QC requirements preclude publication of “just anybody’s” data
- In the midst of transition

AirNow



Data User
data
AirNow
AirNow API
(standardized output and data feeds)
metadata
data
data
metadata
Organization
data
AirNow DMS
data
EnviroFlash
(public alerting system)
data
Analysis tools
Sensor
AirNow-Tech
(Decision Support System)

AirNow

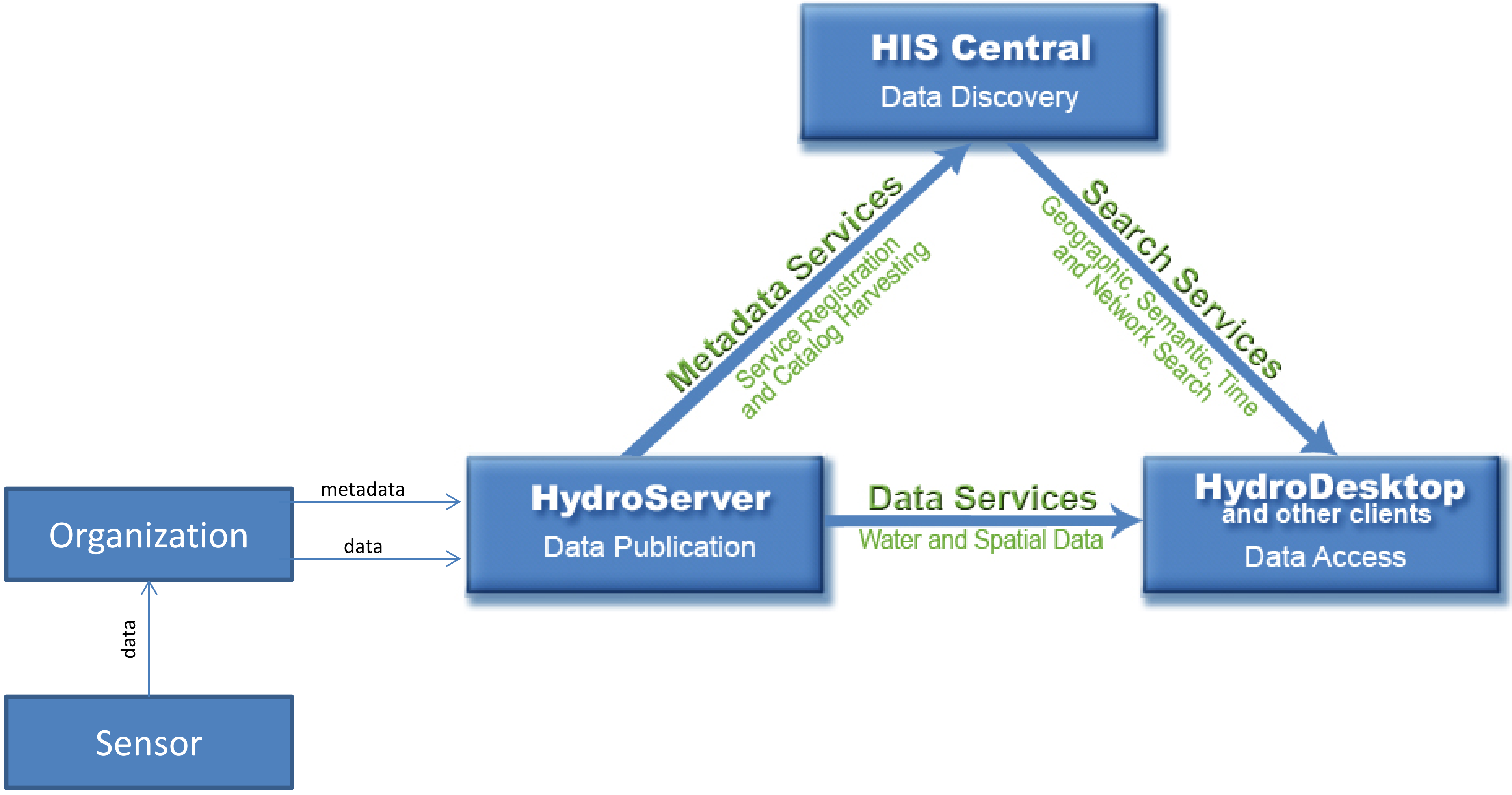
Pro

- Established system already integrating sensors across the US and the world.
- Designed to support extension to water, other media
- Adaptable to new parameters
- Handles real-time and bulk upload data processing
- Existing homogeneous and heterogeneous QA/QC routines
- Web-based

Con

- Does not currently support full range of desired OGC services (SOS/WaterML 2, CS-W)

CUAHSI-HIS



HIS Central
Data Discovery
Metadata Services
Service Registration
and Catalog Harvesting
Search Services
Geographic, Semantic, Time
and Network Search
HydroServer
Data Publication
Data Services
Water and Spatial Data
HydroDesktop
and other clients
Data Access
Organization
metadata
data
data
Sensor

CUAHSI-HIS

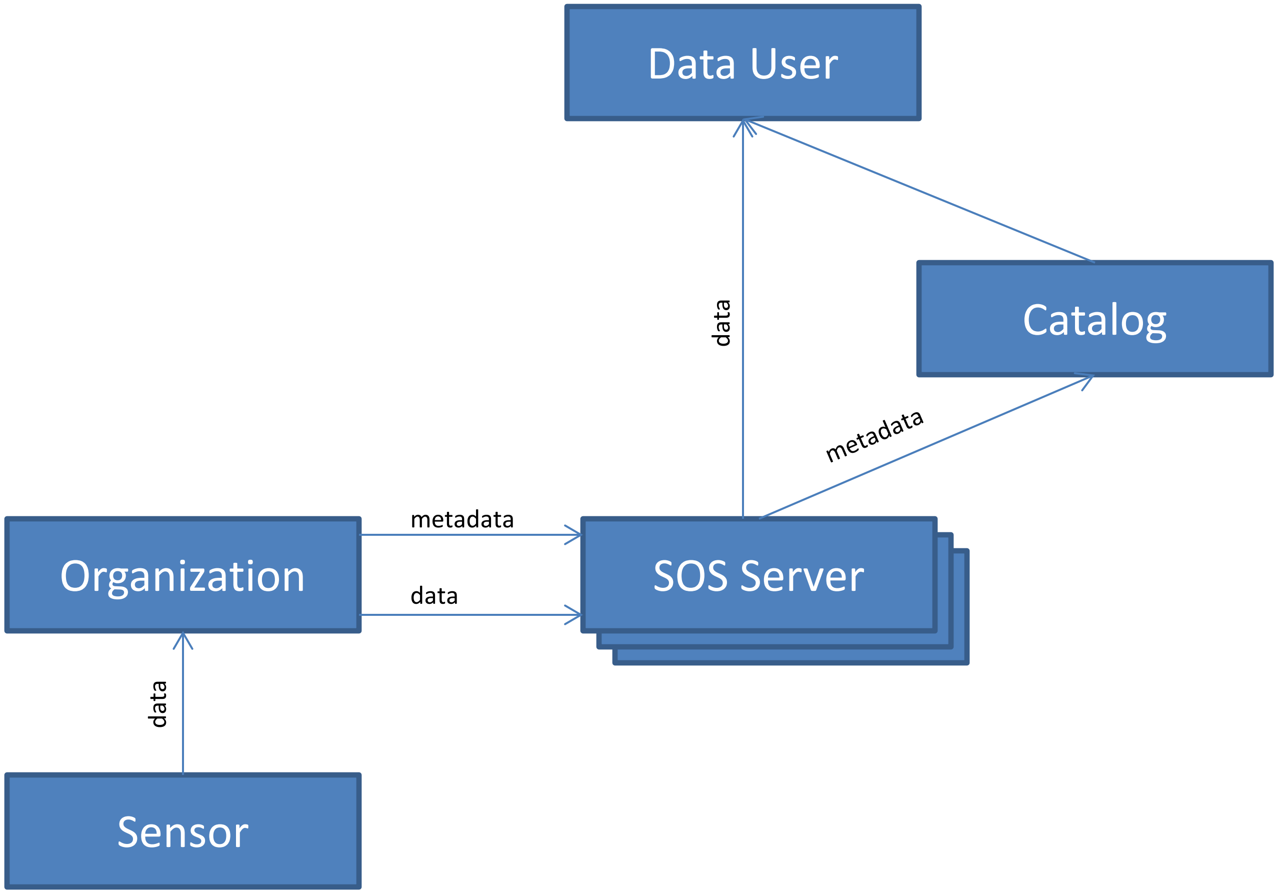
Pro

- Built on OGC standards to support discoverability and interoperability
- Working on inexpensive cloud-based appliances for serving data.
- Supports heterogeneous problem domain by standardizing data exchange

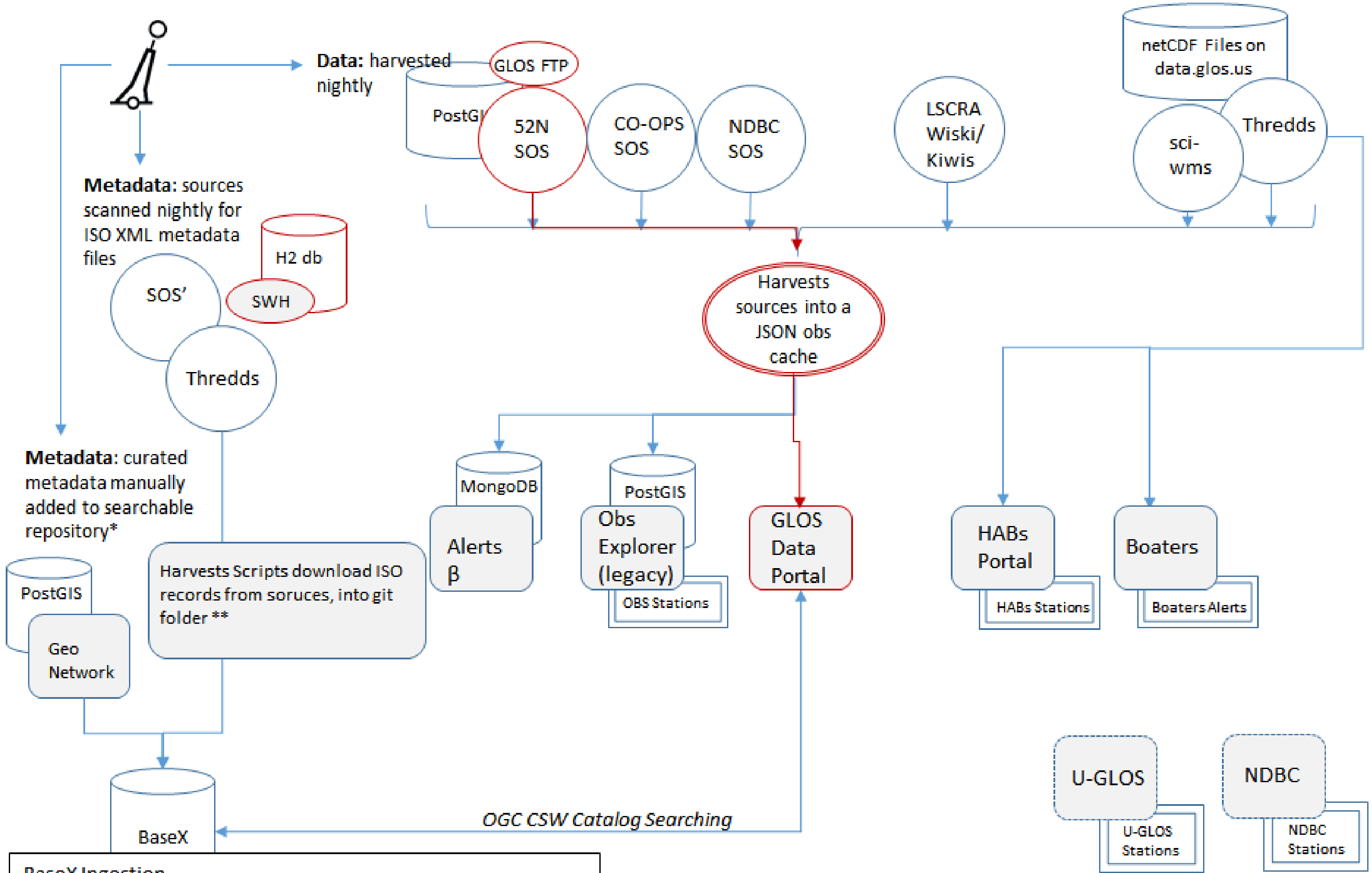
Con

- Scalability uncertain
- Current issue with connection between sensors and cloud appliances.

NOAA/IOOS


Data User
data
Catalog
metadata
metadata
Organization
data
SOS Server
data
Sensor

NOAA/IOOS: Detail for GLOS



BaseX Ingestion
Every other hour, an ingestion script reads the ISO Metadata Documents on db.glos.us and inserts them into the BaseX XML Data Store

*proposed future would be direct harvest of Thredds to GeoNetwork to avoid duplication

**At various intervals, depending on the source (usually by size of the dataset)

April 24, 2015

NOAA/IOOS

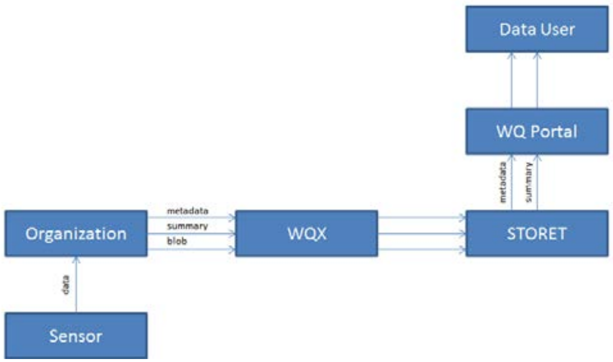
Pro

- Like CUAHSI, heavily OGC-compliant to promote interoperability and discoverability.
- Involved in development of standardized QC for real-time data.
- Tools like Scalability Experiment demonstrate value of approach.

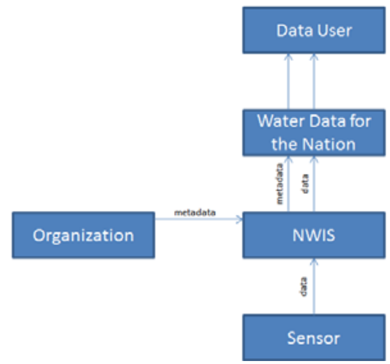
Con

- Data standards are well-defined, but many different technology stacks exist for implementation.
- Focus is on oceanographic variables, which include only some WQ parameters.

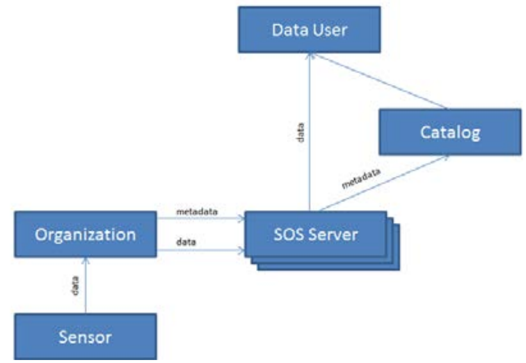
Status Quo



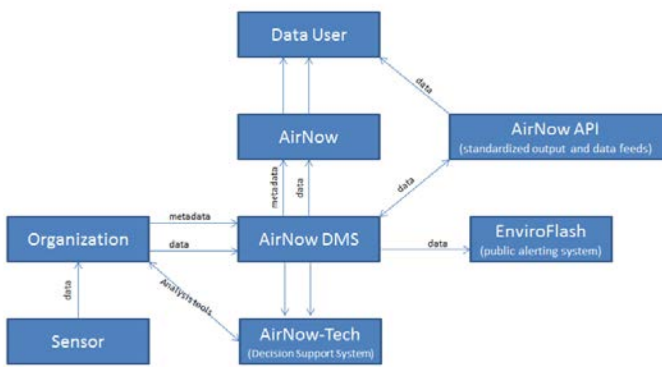
USGS NWIS (AQUARIUS Buildout)



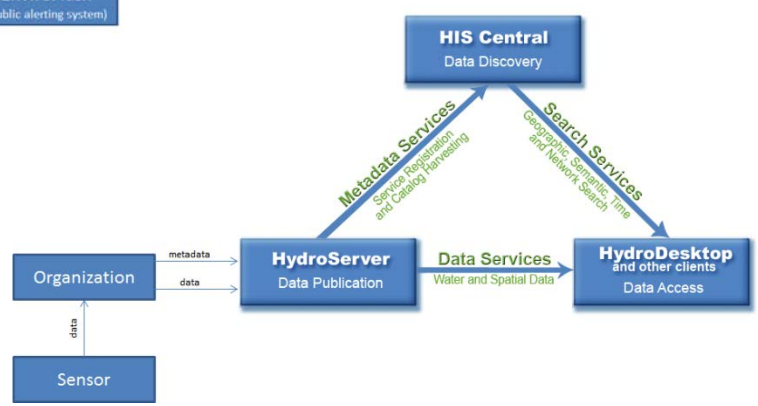
NOAA/IOOS



AirNow



CUAHSI-HIS



HYBRID APPROACH?

Hybrid Approach: A Trial Balloon

Principles:

- Data (observations) available from distributed servers.
- Metadata (site/sensor/deployment) harvested and made available from centralized server.
- Interoperable with WQX and Water Quality Portal

Hybrid Approach

User Activities

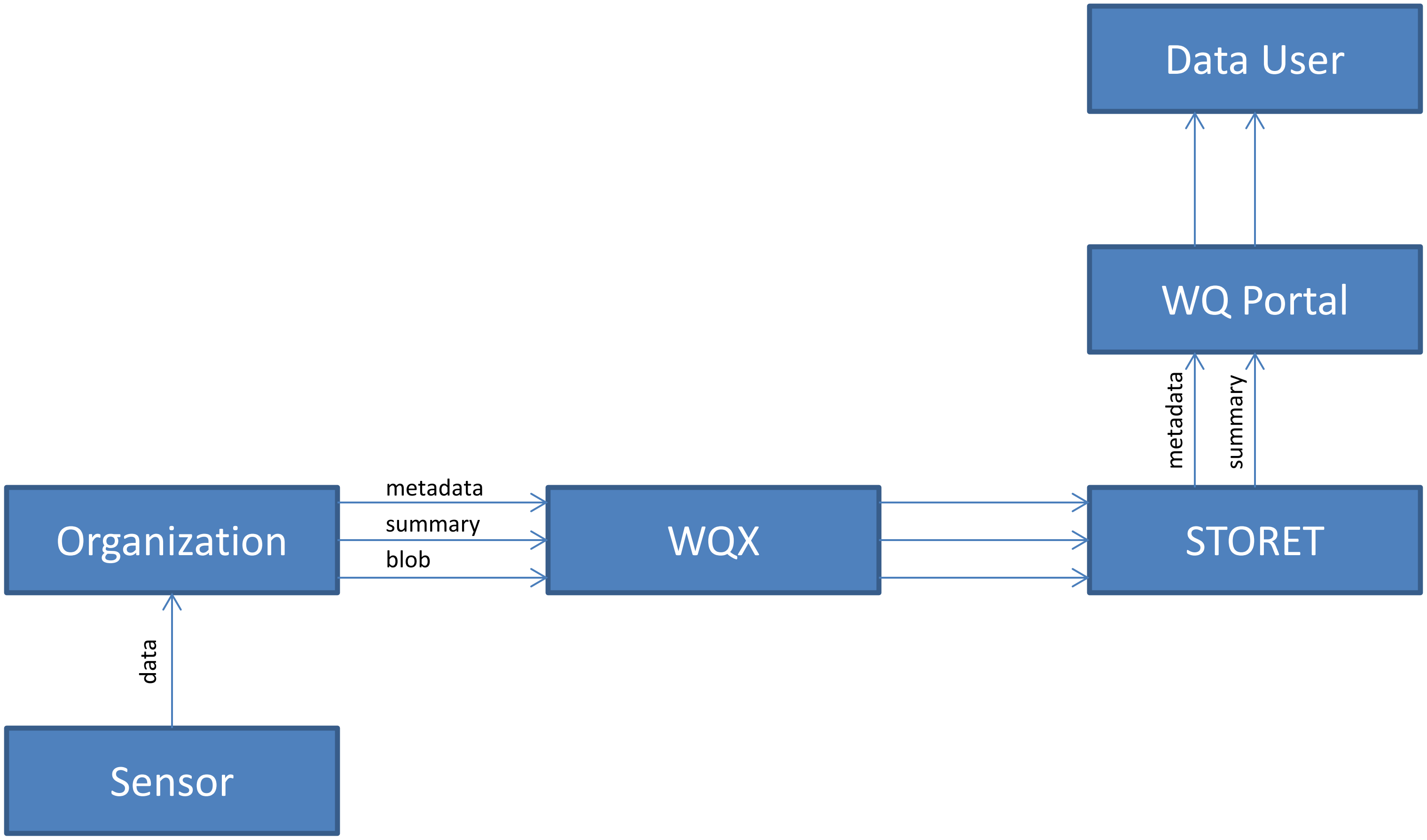
- Submit CWQS site to WQX
- Submit sensor/deployment QC metadata to ?
- Connect (near-realtime) or upload (batch) observations and QC data to appliance
- Search for sensors in discovery tool
- Download metadata and data through discovery tool

Hybrid Approach

Required Elements

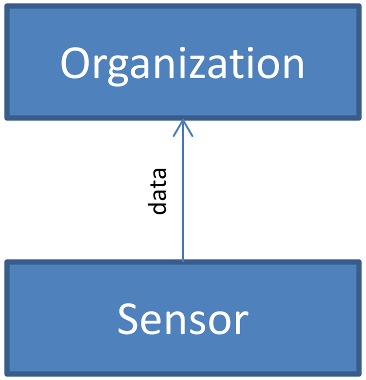
- Servers for data, ideally "owned" by submitter
  - Commercial software (AQUARIUS, KiWIS, …) or
  - Data appliance in cloud (CUAHSI HIS server?)
- Repository and server for metadata
  - Extend WQX with tagging of sites as continuous
- Discovery tool
  - WQ Portal or IOOS Scalability Experiment

Status Quo

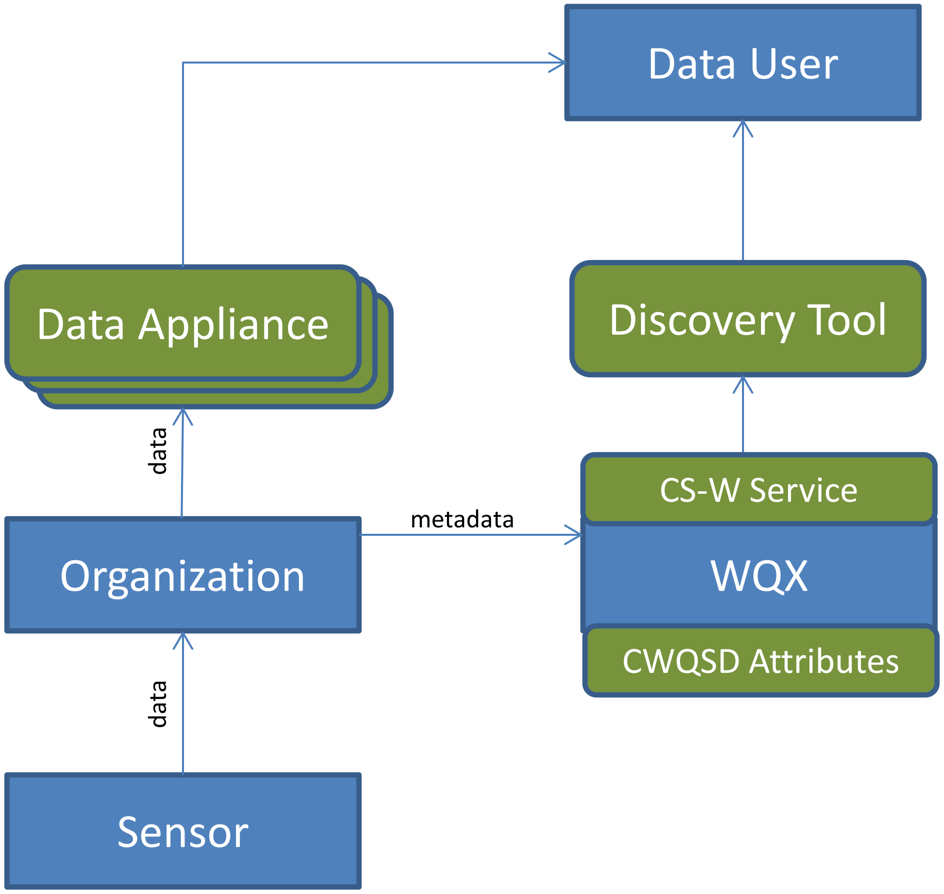


Data User
WQ Portal
metadata
summary
Organization
metadata
summary
blob
WQX
STORET
data
Sensor

Status Quo (worst case)



Hybrid Approach



Hybrid Approach

Concerns (so far)

- Handling of QC metadata
  - Serve with site (WQX) or with data (appliances)?
- Disambiguation/deduplication of data
  - Integrating NWIS, IOOS, other data streams
- Sustainability: archiving of data
  - Design to support optional transition to centralized data? System of record?

Hybrid Approach (2)