300 Area IFC Data Management

Roelof Versteeg

IFC Project Meeting, April 29-30,2008

Idaho National Laboratory

Outline

- Data management objectives
- General and technical approach
- Discussion topic: the borehole and sample database
- Getting data in (and out) of the database
  - Biggest challenge: accurately capturing experimental data

Data management Objectives

1. **Provide a central, well organized, accessible repository for historic and new data relating to the 300 Area IFC**

2. **Provide tools for efficient data access, data management and result generation**

Data inventory effort (organized by John Zachara) – resulting in spreadsheet summarizing existing data

- Some 60 different data sources (geochemical, hydrological, geophysical, numerical model results, lab data, field data) with varying degrees of organization and amount of data
- Significant variability in data organization, documentation, ownership and accessibility
  - Best: data organized in documented, accessible databases
  - Medium: Excel spreadsheets with different levels of organization and documentation
  - Worst: data without history
- Virtually all data needs massaging to make it useful

Criteria for usefulness for IFC effort (everyone should meet these!)

- Data is useful if it meets the following criteria
  - Well organized
  - Fully documented
  - Accessible
  - Auditable
- Significantly higher bar for IFC project than for internal, single PI use

IFC Cyber infrastructure approach and elements

- Heavy use of relational databases
- User interface with system and applications through rich web interface
- Use of open source elements and solutions where possible
- Expose existing desktop applications through web interfaces
- Build and implement re useable, modular components

IFC Cyberinfrastructure in a nutshell

- **Web accessible database with rich client interface**
- **Central server and application – distributed users**
- **Allows for access control, data uniformity and ease of maintenance**

Server: Data QA/QC, Data storage in relational database, data management & analysis tools

Users use standard browser to connect to server and access data

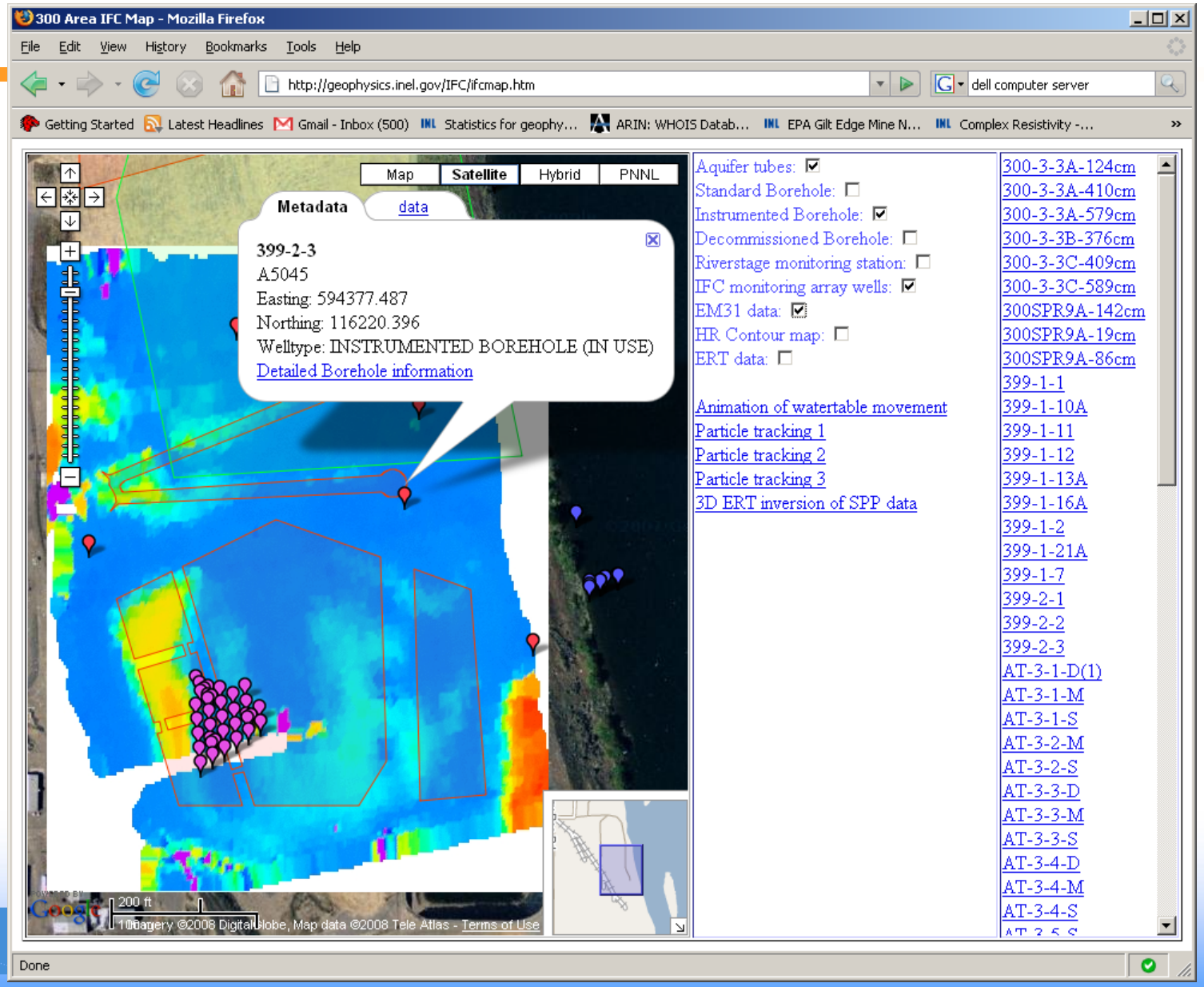
Technical details

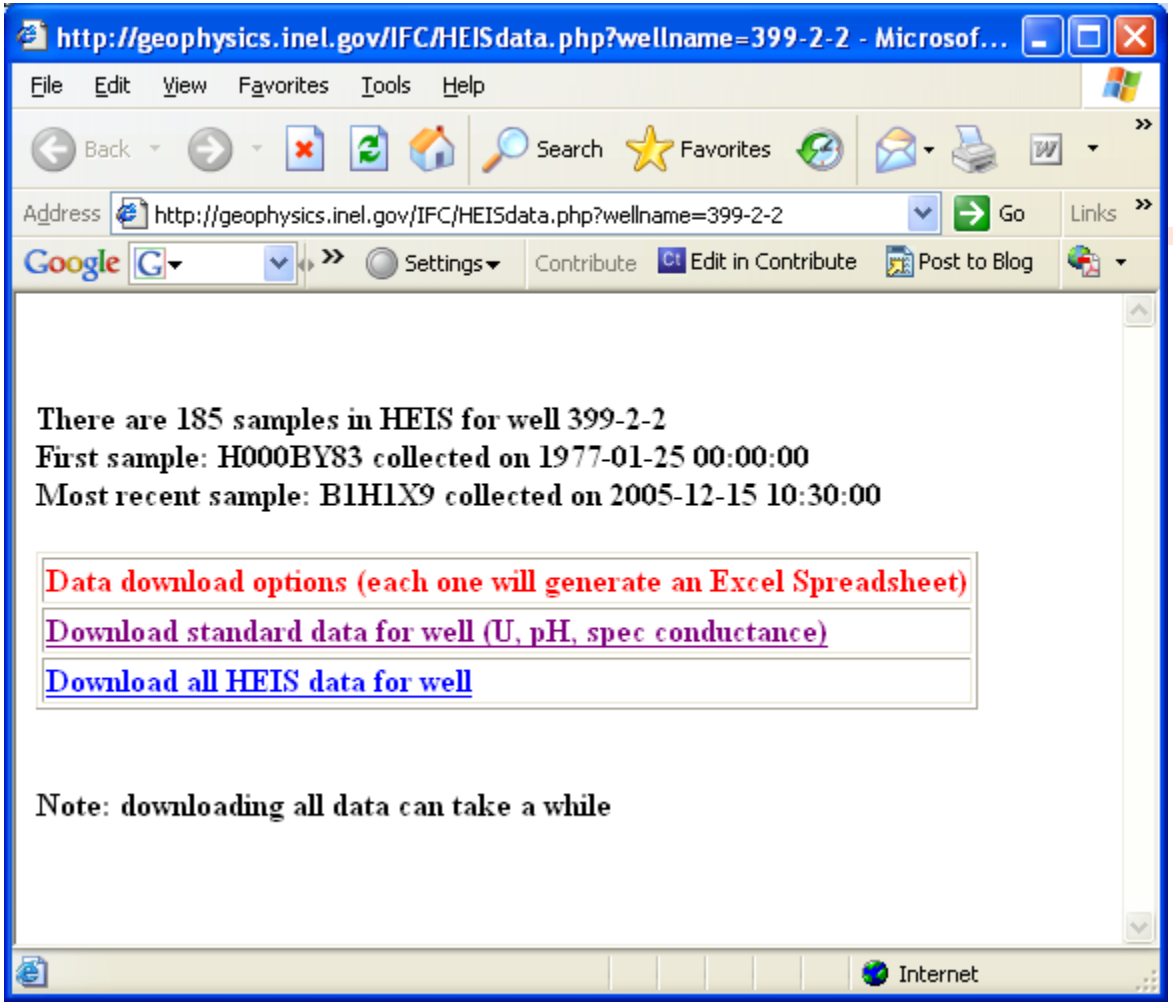
- Server side: Linux Server (Fedora Core)
- Standard Apache/MySQL/PHP stack
- Tomcat for webservices
- JFreeChart/OpenDX for local visualization
- Clientside: rich webinterfaces using Javascript/Cascading Style Sheets (CSS)
- Visual data access through Google Maps API

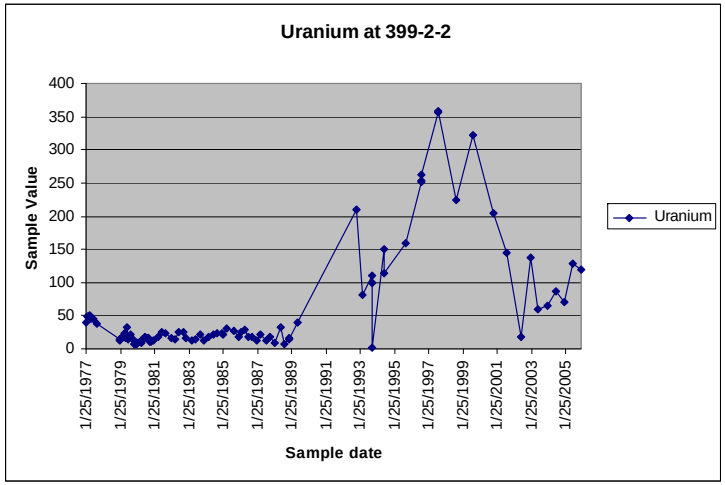
Example of available data through interface: HEIS Data

- HEIS – Hanford Environmental Information System (required by Tri Party Agreement)
- Official repository of data from soil, biota, atmospheric, miscellaneous material, surface water and groundwater samples at the Hanford Site.
- Currently maintained by Fluor
- Internally accessible through webpage frontend to Oracle backend. Externally as standalone Access Application
- Contains all environmental sampling results for all wells at Hanford (typically quarterly sampling) (200 K samples/2 Million results)
- No research data
- HEIS 300 Area sampling frequency currently 6 months (June/Dec)

Heis data access currently through Google Map interface







More download options (and graphing) can be added

Data from 399-2-2 generated From spreadsheet obtained Through page

Idaho National Laboratory

Ongoing effort: borehole and sample database

- Historic borehole and sample data management
- New approach (under way)

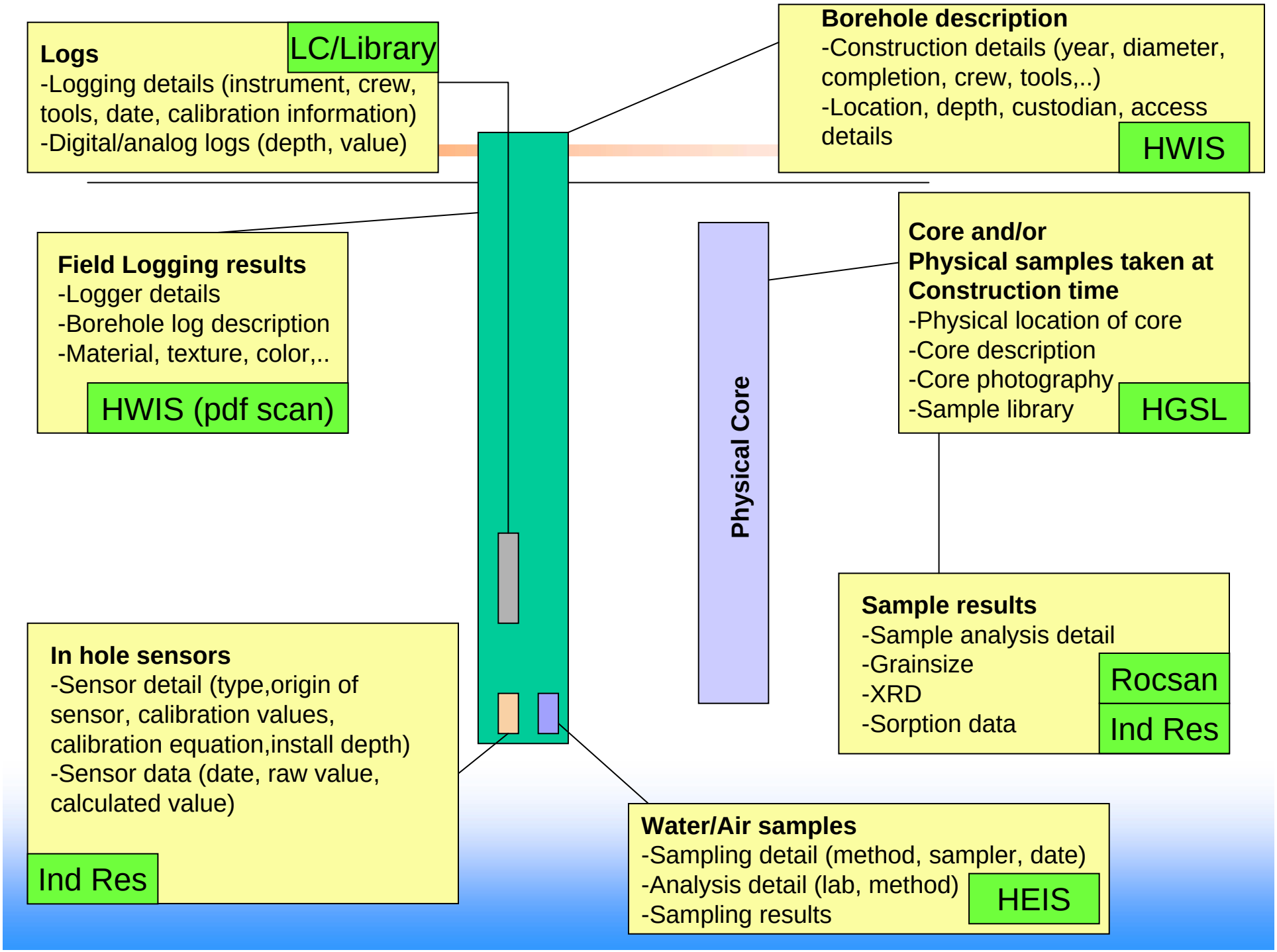
Importance of borehole data

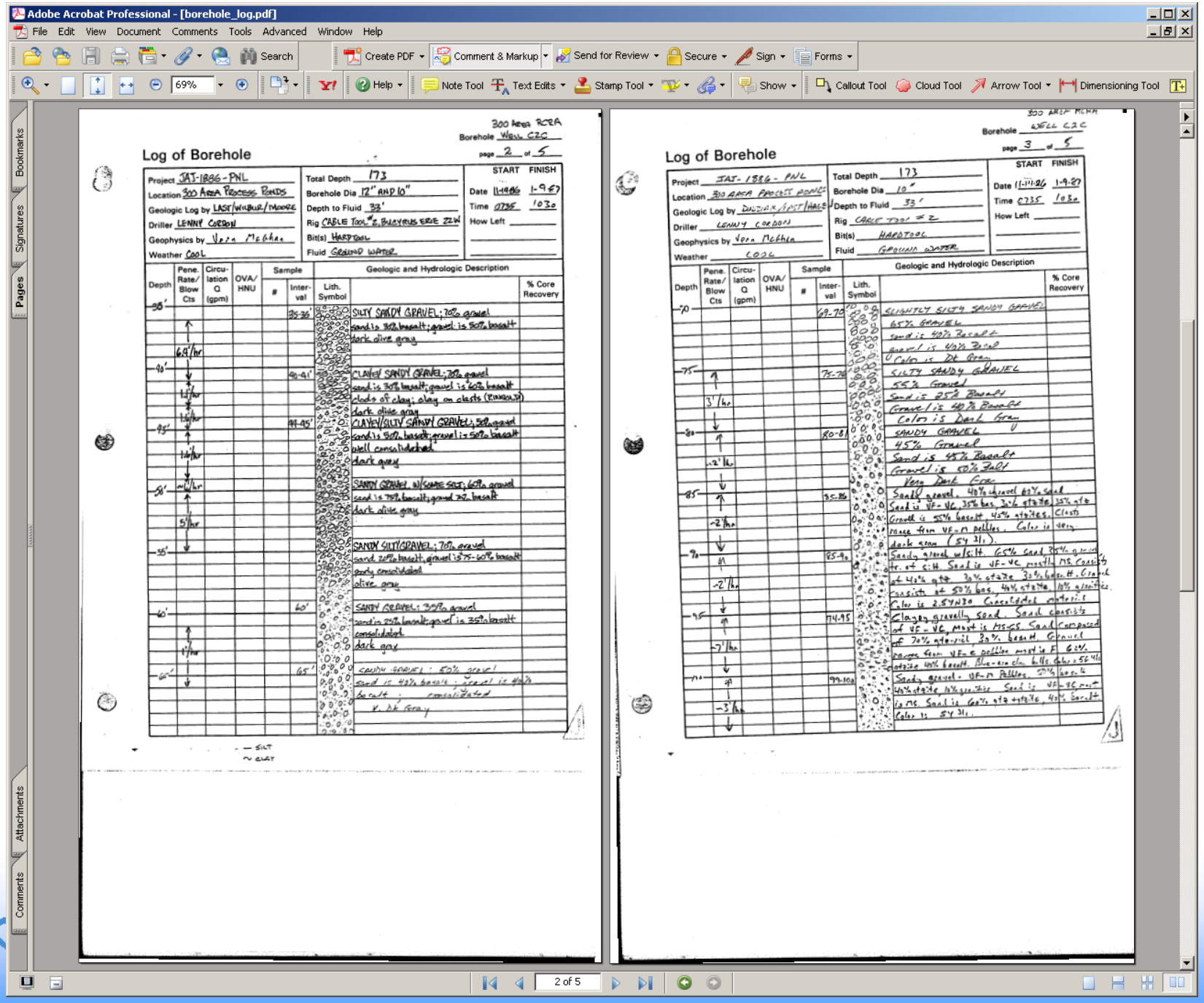
- Borehole data very important for several reasons
  - Core piece of hard data (lithology, logs,samples)
  - "Hook" for auxiliary data (samples, monitoring instrumentation, soft data (interpretations))
  - Start point for geologic mapping, parameter estimation, …
  - Used by multiple users with different objectives
  - Need to capture historic and novel borehole data is top priority

Data associated with boreholes (historically)

- **Construction and maintenance data.**
- **Geologic information.** Geologic/driller logs + core photos
- **Samples and cores.** Historically minimal management.
- **Core/sample analysis results.** Results from analysis are scattered.
- **Geophysical logs.** Older wells only hardcopy. For newer wells digital.
- **Groundwater sample analysis data.** Regulatory results in HEIS
- **Groundwater monitoring data.** Subset of wells in 300 area are instrumented. Data collected at hourly intervals

- **Note:** Data is housed and managed in different organizations and places


Logs
-Logging details (instrument, crew, tools, date, calibration information)
-Digital/analog logs (depth, value)
LC/Library
Borehole description
-Construction details (year, diameter, completion, crew, tools,..)
-Location, depth, custodian, access details
HWIS
Field Logging results
-Logger details
-Borehole log description
-Material, texture, color,..
HWIS (pdf scan)
Physical Core
Core and/or
Physical samples taken at Construction time
-Physical location of core
-Core description
-Core photography
-Sample library
HGSL
Sample results
-Sample analysis detail
-Grainsize
-XRD
-Sorption data
Rocsan
Ind Res
In hole sensors
-Sensor detail (type,origin of sensor, calibration values, calibration equation,install depth)
-Sensor data (date, raw value, calculated value)
Ind Res
Water/Air samples
-Sampling detail (method, sampler, date)
-Analysis detail (lab, method)
-Sampling results
HEIS


Adobe Acrobat Professional - [borehole_log.pdf]
Log of Borehole
Borehole WELL C2C
page 2 of 5
Project JAT-1886-PNL
Location 300 AREA PROCESS PONDS
Geologic Log by LAST/WILBUR/MOORE
Driller LENNY CORDON
Geophysics by Vern McGhan
Weather COOL
Total Depth 173
Borehole Dia 12" AND 10"
Depth to Fluid 33'
Rig CABLE TOOL "E. BUCYRUS ERIE 22W
Bit(s) HARDTOOL
Fluid GROUND WATER
START FINISH
Date 11-19-86 1-9-87
Time 0735 1030
How Left
Depth
Pene. Rate/ Blow Cts
Circu-lation Q (gpm)
OVA/ HNU
Sample # Inter-val
Geologic and Hydrologic Description
Lith. Symbol
% Core Recovery
SILTY SANDY GRAVEL; 70% gravel
sand is 30% basalt; gravel is 50% basalt
dark olive gray
CLAYEY SANDY GRAVEL; 30% gravel
sand is 30% basalt; gravel is 60% basalt
clods of clay; clay on clasts (RINDS ?)
dark olive gray
CLAYEY/SILTY SANDY GRAVEL; 50% gravel
sand is 50% basalt; gravel is 50% basalt
well consolidated
dark gray
SANDY GRAVEL w/SOME SILT; 60% gravel
sand is 70% basalt; gravel 70% basalt
dark olive gray
SANDY SILTY GRAVEL; 70% gravel
sand 20% basalt, gravel is 50-60% basalt
poorly consolidated
olive gray
SANDY GRAVEL; 35% gravel
sand is 25% basalt; gravel is 35% basalt
consolidated
dark gray
CLAYEY GRAVEL; 50% gravel
sand is 40% basalt; gravel is 40% basalt
consolidated
V. dk Gray
Log of Borehole
Borehole WELL C2C
page 3 of 5
Project JAT-1886-PNL
Location 300 AREA PROCESS PONDS
Geologic Log by DUZIAK/LAST/HALE
Driller LENNY CORDON
Geophysics by Vern McGhan
Weather COLD
Total Depth 173
Borehole Dia 10"
Depth to Fluid 33'
Rig CABLE TOOL #2
Bit(s) HARDTOOL
Fluid GROUND WATER
START FINISH
Date 11-11-86 1-9-87
Time 0735 1030
How Left
SLIGHTLY SILTY SANDY GRAVEL
65% GRAVEL
Sand is 40% Basalt
Gravel is 40% Basalt
Color is Dk Gray
SILTY SANDY GRAVEL
55% Gravel
Sand is 25% Basalt
Gravel is 40% Basalt
Color is Dark Gray
SANDY GRAVEL
45% Gravel
Sand is 45% Basalt
Gravel is 50% Felsic
Very Dark Gray
Sandy gravel, 40% gravel 60% sand
Sand is VF-VC, 35% bas, 30% qtzite,
Gravel is 55% basalt, 45% qtzites,
range from VF-M pebbles. Color is very
dark gray (5Y 3/1).
Sandy gravel w/silt 65% sand
fr. of c-M Sand is VF-VC mostly
at 40% qtz 30% qtzite 30% bas. M. Gravel
consists of 50% bas, 40% qtzite,
Color is 2.5YN3/0 Consolidated material
Clayey gravelly Sand. Sand consists
of VF-VC, most is MS-CS. Sand composed
of 70% qtz-felsic, 30% basalt. Gravel
ranges from VF-C pebbles most is F
qtzite 40% basalt. Fine-med cl. 60%, Color = 5Y 4/1
Sandy gravel. VF-M Pebbles, 70% bas-L
40% qtzite, 10% granitic Sand is VF-VC mst
is MS. Sand is 60% qtz+qtzite, 40% basalt
Color is 5Y 3/1.
2 of 5
Typical borehole log from 1986

Previous Borehole organization efforts

- HBGIS **(Hanford Borehole GIS).** Effort led by George Last. Involved several PNNL IFC researchers (Bruce, Mark, Vince). Currently focused mainly on 100 Area data. Some similarity with current efforts, but not linked to samples. Not externally accessible
- HWIS **(Hanford Well Information System).** Maintained by Fluor. Not externally accessible. Central repository of all borehole data. Metadata directly in database. Boreholes in database as pdf (so data not directly useable)
- **HEIS (Hanford Environmental Information System)**
- **Others**
  - http://www.orp.doe.gov/cp/gpp/modeling/databases.cfm
  - http://www.orp.doe.gov/cp/gpp/data/vzcp/vzcp.cfm **(borehole geophysical data of 300 FF5 area)**
  - http://www.osti.gov/bridge/servlets/purl/6602349-ngHns8/6602349.PDF **(recent geophysical borehole data)**

Some limitations with current software tools

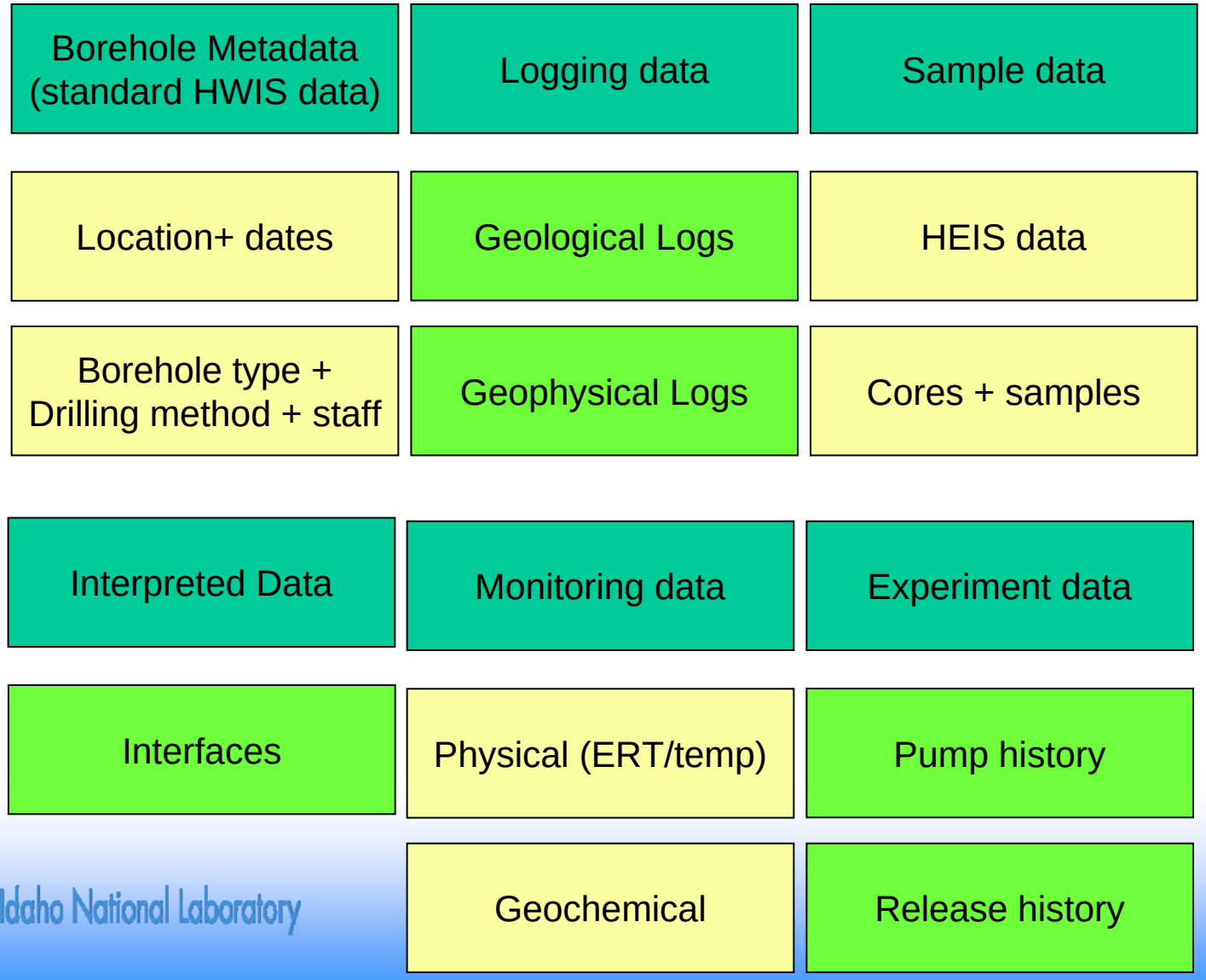
- Current approach separates information on boreholes and samples
- Current databases do not easily integrate with models
- Current databases do not store geophysical or geological results (nor soft/ interpreted data such as location of interfaces)
- Current approach limited in tracking and dealing with borehole samples and timelapse data
- Current approaches not easily integrated with data associated with experimental campaigns
- Access to existing tools generally PNNL internal only

- Conclusion: need for new tool

New data model

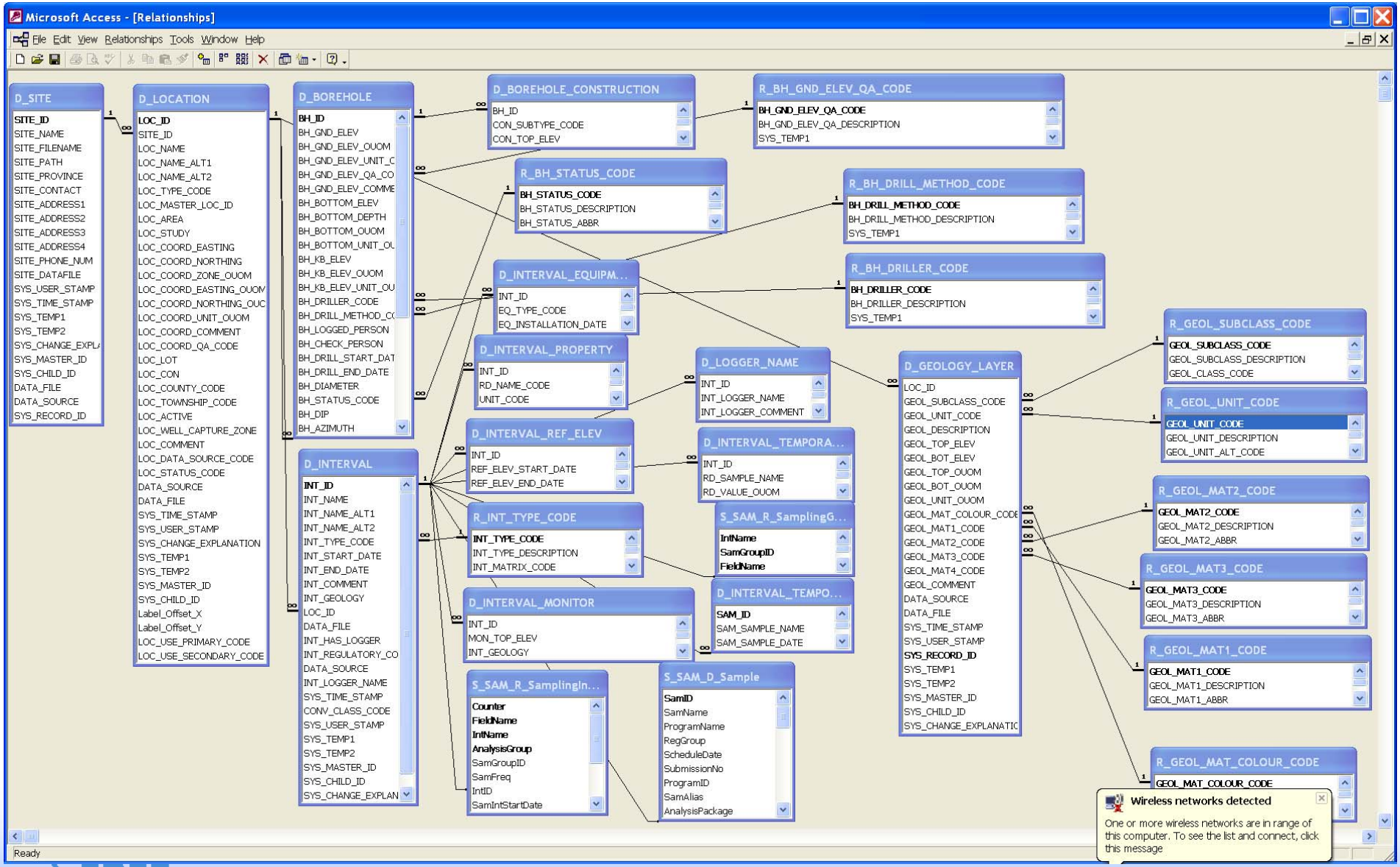
- Evaluated existing borehole data models (and data formats)
- Considered cradle to grave IFC needs
- Looked at models developed by
  - Geological surveys (e.g. Canada (LAS format), UK, Netherlands, Germany)
  - Geotechnical Organizations (e.g. UK AGS format)
  - Oil industry
  - Environmental industry

Elements which well database model should be able to accommodate



Model Choice

- Modified EarthFX Data model (commercial system)
- Can accommodate all data easily, except for timelapse temperature and resistivity data
  - resistivity data stored in dedicated resistivity database
  - data model will be adapted for temperature data
- MSAccess native, converted to MySQL



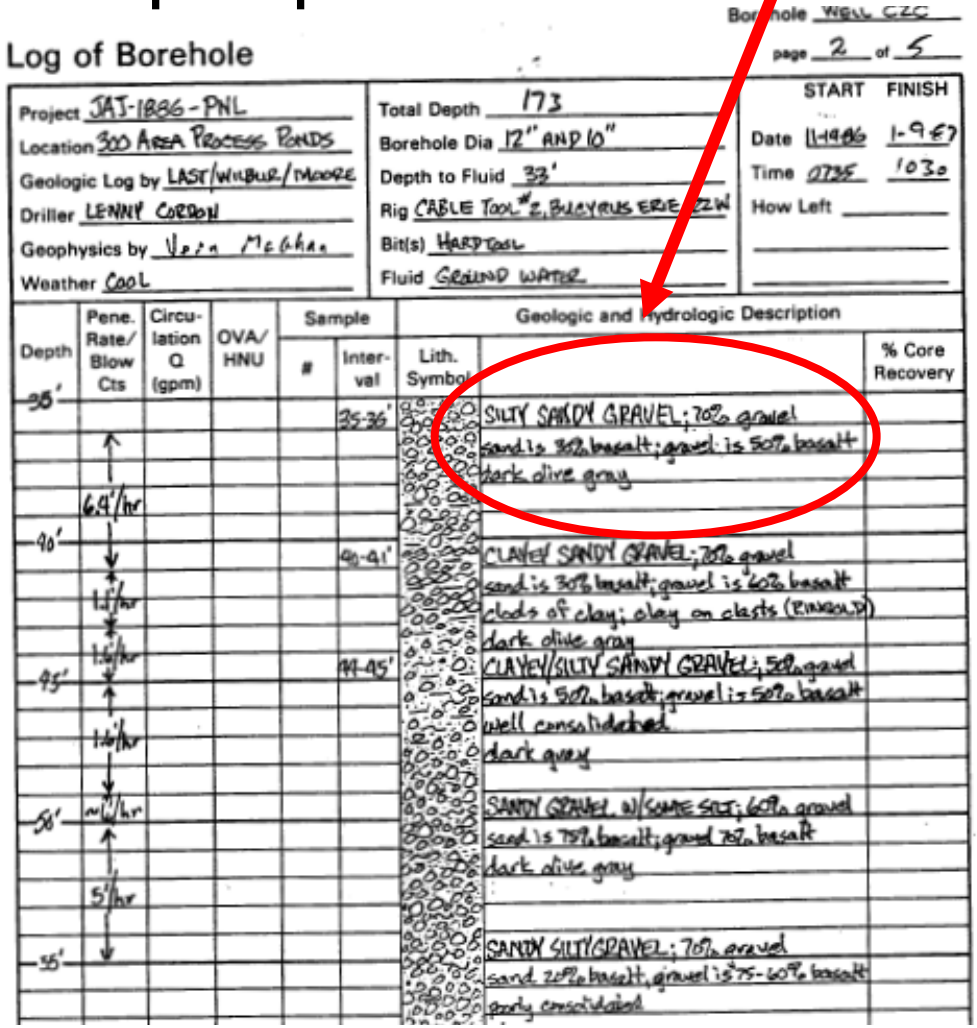
Idaho National Laboratory

Borehole Database status

1. Retrieve metadata from HWIS - done
2. Capture metadata in Excel - done
3. Capture borehole pdf data - done
4. Upload metadata to MySql database – done
5. Establish borehole conversion scheme – ongoing (for discussion today)
6. Import borehole data – coming month

Important open issue

- EarthFX Modified can take (without modification) any of the current metadata associated with wells
- We can capture free form description, but not easy to use
- Problem in classification - lump or split?
- Or, do we do
  - Main (Gravel)
  - Second (Sandy)
  - Third (Silty)
- Geologists to decide
- Will solicit input



Log of Borehole

Borehole WELL C2C

page 2 of 5

Project JAJ-1886-PNL
Location 300 AREA PROCESS PONDS
Geologic Log by LAST/WILBUR/MOORE
Driller LENNY CORDON
Geophysics by Vern McGhan
Weather COOL

Total Depth 173
Borehole Dia 12" AND 10"
Depth to Fluid 33'
Rig CABLE TOOL #2, BUCYRUS ERIE 22W
Bit(s) HARDTOOL
Fluid GROUND WATER

	START	FINISH
Date	11-19-86	1-9-87
Time	0735	1030
How Left		

Depth	Pene. Rate/ Blow Cts	Circu-lation Q (gpm)	OVA/ HNU	Sample #	Sample Inter-val	Lith. Symbol	Geologic and Hydrologic Description	% Core Recovery
35'					35-36'		SILTY SANDY GRAVEL; 70% gravel sand is 30% basalt; gravel is 50% basalt dark olive gray	
	6.9'/hr							
40'					40-41'		CLAYEY SANDY GRAVEL; 70% gravel sand is 30% basalt; gravel is 60% basalt clods of clay; clay on clasts (RINDS?) dark olive gray	
	1.5'/hr							
45'	1.5'/hr				44-45'		CLAYEY/SILTY SANDY GRAVEL; 50% gravel sand is 50% basalt; gravel is 50% basalt well consolidated dark gray	
	1.6'/hr							
50'	~1'/hr						SANDY GRAVEL w/SOME SILT; 60% gravel sand is 75% basalt; gravel 75% basalt dark olive gray	
	5'/hr							
55'							SANDY SILTY GRAVEL; 70% gravel sand 20% basalt, gravel is 55-60% basalt poorly consolidated	

Sample model proposal (for discussion)

- One person responsible sample information entry
- Sample metadata (date, personnel, method, container, weight, moisture content) recorded at entry
- Gatekeeper for sample handout
- Subsample procedure
  - Gatekeeper logs date, quantity and *sample analyzer* into the system.
  - System assigns subsample tracking number.
  - User can assign subsample name
  - *sample analyzer* does result upload
- Allows for rapid assessment of existing results and remaining sample material

Getting data in (and out) of the database

- Which data can you get today
  - HEIS data
  - Borehole metadata
  - Geophysical Imagery
  - Location of infiltration ponds
  - Aerial imagery of site
- Which data will you be able to get in 1-2 months from now
  - Historic borehole data (including interface picks)
  - Sample information for new wells
  - Monitoring hydrological data from instrumented wells and stream gages
  - Historic and current weather data
- Which data can you get by September 2008
  - “all” historical data + new borehole/sample data

Anticipated data in next year (please point out missing elements)

- Borehole data (metadata, borehole logs, geophysical logs)
- Sample data
  - Sample metadata
  - Sample analysis results
- Field Experiment related results
  - Experimental metadata (injection)
- Monitoring data
  - Geochemical monitoring
  - Geophysical cross well and surface electrical timelapse data
  - Temperature monitoring data
- Results from physical and biogeochemical experiments
- Modeling results

Experimental data challenge

- Most complex data to capture will be lab experiment related data
- As yet undefined what we want/should store
- Effort to comprehensively capture experiments and results will be iterative

- Example: Uranium adsorption results

Experiment documentation [proposal]

- Experiment classified primary and secondary type
  - Uranium adsorption experiment
  - Grainsize analysis
- Experiment objective (narrative)
- Experiment sample methods
- Sample preparation (Sieving, Acidifying) and procedures
- Environmental conditions

- Challenge: compare results from similar experiments by different groups (e.g. U adsorption/desorption)

Getting numerical model data in and out of the database (proposal)

- Document numerical model and version
- Upload page allows users to upload model metadata and descriptions
- Numerical result grids: use NetCDF4/HDF as standard storage format for IFC modeling results
- Ideally leave data files locally. Use web services for result extraction. Possible storage on the INL HPC cluster if needed
- Browser based tools for client side result extraction and visualization (e.g. profiling) requests

Database development

- Migration to merge with PNNL site over next two weeks
- Username/Password will be mailed out to all IFC participants
- Specific people will be approached to provide input/data descriptions/discussions on
  - Experimental results
  - Geologic classification
  - Modeling results

Final comment

- Need to make sure that we measure what we need
  - Measure U and other relevant properties in monitoring wells on monthly basis over project duration instead of every 6 months so we capture annual variability better?
  - Ensure consistency between experimental efforts (circulate experimental plan before experiments?)
- Need to formalize procedures for deriving model parameters from field and lab measurements


INL
Idaho National Laboratory

Talk objectives:

1. inform

2. obtain feedback on effort direction and input on data needs

More detailed …

- **Remind audience on high level objectives and deliverables of data management effort (*"what does Roelof do – what will it get me?"*)**
- **Provide a high level overview of the technical approach and motivation for doing this, and give an example implementation ("*how does this all work?*")**
- **Provide status snapshot and timeline (*"what can we get now (and how do we get it) and what can we expect to get in the future (and when)?"*)**
- **Obtain input on data and tool needs from audience**

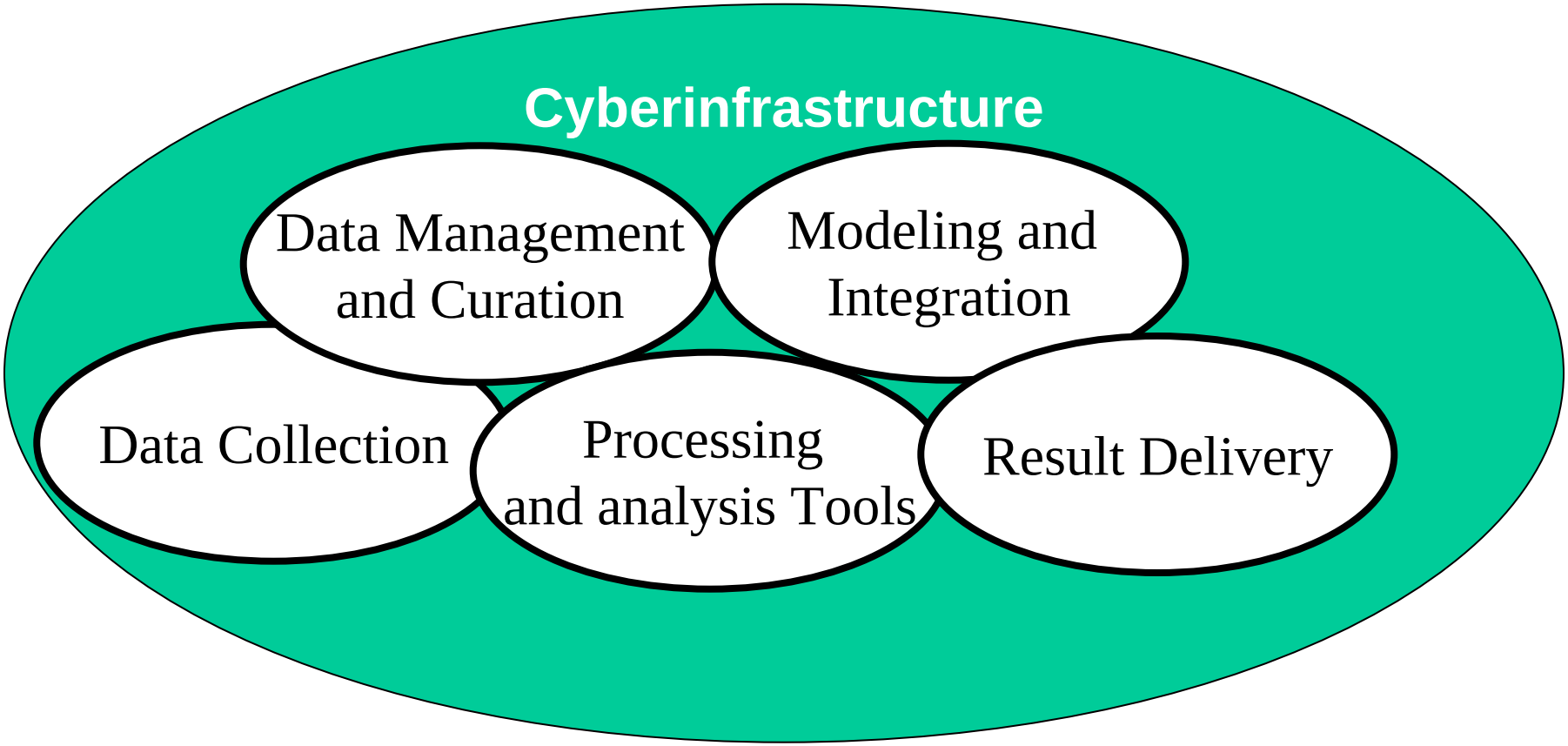
Example deliverables

- Access to geochemistry and hydrology monitoring data
- Historic and new borehole data access system
- Sample management and tracking system
- Visual and text based interfaces to data
- Tools for data import/export to models and desktop tools
- Model data storage and access system
- Access to data and results from infiltration experiments
- Graphing and contouring tools
- Secure environment for result exchange
- .....

General approach so far

1. Find out what data there is and who has it
2. Find out what form and state the data is in (did somebody do (part of) the work for us already?)
3. Decide on overall and database specific implementation architecture
4. Collect historic data
5. Start organizing historic data and make it available
6. Agree on rules for capturing new data and develop and provide tools

Cyberinfrastructure: the IT glue between the different components involved in the IFC Datamanagement effort



Cyberinfrastructure definition

- **Cyberinfrastructure : The comprehensive infrastructure needed to capitalize on dramatic advances in information technology.**
- **Cyberinfrastructure integrates hardware for computing, data and networks, digitally-enabled sensors, observatories and experimental facilities, and an interoperable suite of software and middleware services and tools (*from* NSF's Cyberinfrastructure Vision for 21st Century Discovery, www.nsf.gov/od/oci/ci-v7.pdf)**

One way to think about cyberinfrastructure

- Programming in the 80's: Cobol, Fortran, C – scientists could (and did) write their own code
- Relatively easy to understand – everything was self contained and local, but limited in what it could do
- Programming in the 21st century: Alphabet soup: SOA,OOP, XML,, .NET,IDE, Java, Javascript, RDMBS, Ant, Ajax, Web 2.0, CSS, workflows, Webservices, Kepler, Taverna, juddi – easy to lose sight of what is what, and what the underlying objectives are
- Things are more powerful – but not as intuitive or simple
- Most important concept to retain: Cyber infrastructure provides tools and middleware such that for data, models and tools we have
  - Self documenting access
  - Communication and mediation mechanisms
  - Effective possibilities for distributed utilization

HEIS Data

- Majority of the data in HEIS is associated with samples collected in support of the cleanup requirements, groundwater monitoring, environmental impact statements, waste site remediation and characterization, and biological/ecological monitoring.
- Note: HEIS does not have research data in it
- HEIS does currently not have high temporal resolution groundwater elevation monitoring data
- 189130 total samples in HEIS version that we use (more recent one should be available soon). Covers all of Hanford (300 Area maybe 5 %)
- About 2 million results in HEIS total
- 1200 possible constituents listed in constituent table

Quick review of borehole efforts at PNNL

- Many boreholes drilled at the Hanford area (~5000)
- Range of different purposes
  - Geological characterization
  - Contaminant delineation
  - Monitoring wells
  - Research and regulatory efforts
- In 300 Area IFC ~ 100 boreholes
- Boreholes include
  - Traditional boreholes
  - Aquifer tubes (shallow (about 10-12 feet) sampling tubes @ river bank)
  - River tubes (aquifer tubes in the streambed)
- Well known by well ID (A5024) and Well ID (e.g. 399-1-1)

Well details

- **Well ID.** Unique string (consisting of a letter followed by four numbers). Assigned by a central Hanford drilling coordinator prior to the well being drilled. Typical well ids would be A5024, B2831, C3074.
- **Well name.** Unique alphanumeric string. Assigned to wells planned to be completed as permanent or semi-permanent wells. Typically denotes something about where the well is located (e.g. 399-1-1 would in the 300 FF5 area). Optionally about the type of well: AT-3-7-D indicates an aquifer tube
- Well ID and Well name are both unique
- Wells such as 399-1-16A, 399-1-16B, 399-1-16C, 399-1-16D are clusters of wells, typically within meters from each other

Example: PNNL 3504-UC11

- A catalog of borehole geophysics on the Hanford site – 1958 to 1980
- Shows logging in 1963 in 300 Area, but data not readily available

6602349.pdf - Adobe Reader

TABLE 5. Geophysical Logs Available for Wells in the 300 Area

Well Name	Date	Logging Company	Gamma Gamma	Neutron Epithermal Neutron	Natural Gamma
1-1	6/63	Battelle			0-76
1-2	6/63	Battelle			0-100
1-3	6/63	Battelle			0-100
1-4	6/63	Battelle			0-100
3-1	6/63	Battelle			0-100
3-3	5/77	Battelle			0-85
4-1	6/63	Battelle			0-96
4-1	6/63	Battelle			0-140
	5/77	Battelle			0-130
5-1	6/63	Battelle			0-100
5-2	6/63	Battelle			0-408
	4/47	WSU	0-298	0-298	0-298
	4/74	WSU	0-410	0-410	0-410
6-1	6/63	Battelle			0-64
8-1	6/63	Battelle			0-100
	6/76	Battelle	0-100		
8-2	5/63	Battelle			0-116
8-3	6/63	Battelle			0-68

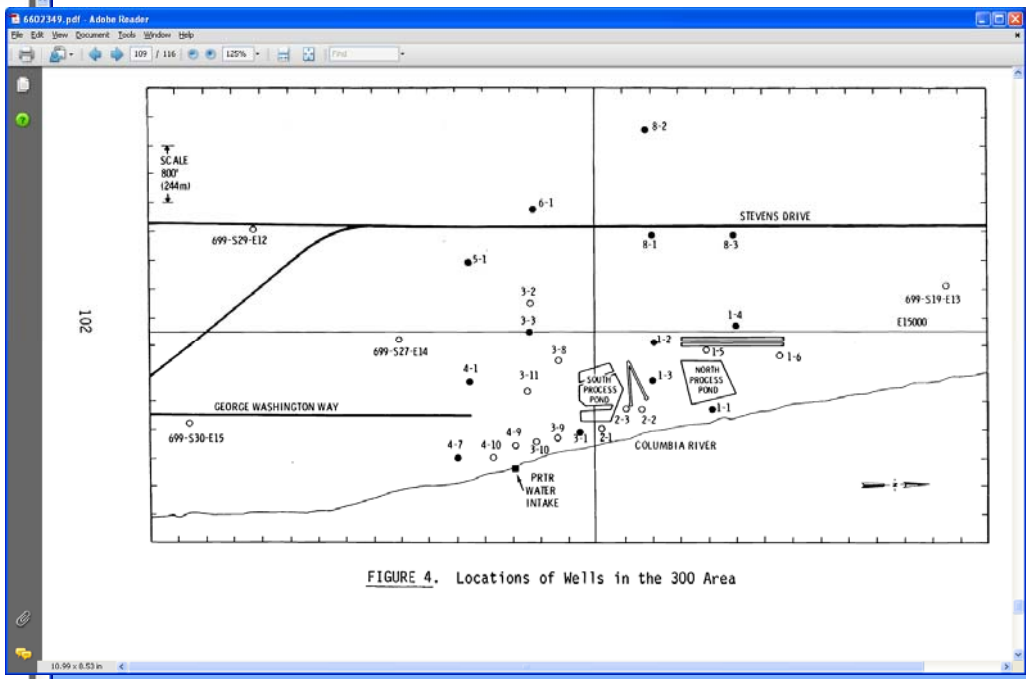


FIGURE 4. Locations of Wells in the 300 Area