

Building and Disseminating Databases at the Kentucky Geological Survey

2017 Data Preservation Workshop, Salt Lake City, UT

Doug Curl & Elizabeth Adams

doug@uky.edu / eladam2@uky.edu

KENTUCKY GEOLOGICAL SURVEY
Quarry Report No. 2

County: Anderson Property Owner: Ky. Stone Co. Operator: Ky. Stone Co. Date Sampled: 9/2/07
Location: Sampled By: Analyzed By: Date Received: Date Reported:

CHEMICAL ANALYSES										Sample		Ledge Thick-		DESCRIPTION		STRATI-	
CaCO ₃	MgCO ₃	SiO ₂	Iron	Alumina	Total	Oxide	Level	No.	ness	(Feet)	(Feet)	(Feet)	(Feet)	LITHOLOGY	GRAPHY		
55.7	37.0	2.80	0.44	0.55	96.5	19-20								Limestone, magnesian,			
70.3	23.2	2.92	0.27	0.54	97.2	18-19								gray to buff, locally			
70.7	23.0	2.70	0.28	0.39	97.1	17-18								and black colored			
58.5	33.4	3.28	0.37	0.94	96.5	16-17								and white, magnesian			
21.4	32.3	4.67	0.17	1.48	97.1	15-16	3	13						Two inch shale seam			
63.9	29.8	4.79	0.44	1.03	99.2	14-15								at bedding			
58.4	31.2	4.04	0.47	0.97	99.4	13-14											
58.2	32.8	4.00	0.33	0.76	99.4	12-13											
37.1	37.4	3.65	0.37	1.05	99.6	11-12											
36.4	38.8	3.46	0.40	0.66	99.6	10-11											
39.3	34.8	4.05	0.39	1.19	99.8	9-10											
61.8	32.6	3.52	0.38	0.92	99.2	8-9											
62.5	32.4	3.79	0.23	0.50	99.1	7-8								Limestone, highly bedded			
78.5	15.6	3.80	0.30	0.92	99.1	6-7								to magnesian, gray			
76.8	16.9	4.00	0.33	1.03	99.2	5-6	2	6						occasional thin shale			
87.1	8.6	3.10	0.27	0.73	99.8	4-5								gray, magnesian			
92.5	5.0	2.31	0.15	0.13	99.1	3-4								white, magnesian			
90.7	5.2	2.23	0.20	0.45	99.8	2-3								and			
83.7	5.2	5.78	0.47	1.89	99.0	1-2	1	2						Limestone, highly magnesian			
85.3	4.2	5.76	0.24	1.22	99.0	0-1								very mottled with thin and			
														dark gray, thin magnesian			
														lenses, containing several			
														thin, magnesian			
														layers.			

KGS Limestone Results.xlsx - Excel

Site	Suite_ID	Lab_ID	From	To	CaCO ₃	MgCO ₃	SiO ₂	Al ₂ O ₃	Fe ₂ O ₃
1393 BTQ	BTQ2	3137	243	244	90.5	1.77	1.65	5.41	0.33
1394 BTQ	BTQ2	3138	244	245	94	1.7	1.27	2.03	0.28
1395 BTQ	BTQ2	3139	245	246	93	3.3	1.54	1.57	0.46
1396 BTQ	BTQ2	3140	246	247	93.8	3.4	1.54	0.83	0.31
1397 BTQ	BTQ2	3141	247	248	90	4.8	2.75	1.68	0.6
1398 BTQ	BTQ2	3142	248	249	86	5.8	4.49	2.76	0.53
1399 BTQ	BTQ2	3143	249	250	84.7	6.2	4.7	3.64	0.65
1400 BTQ	BTQ2	3144	250	251	93.4	4.3	1.32	0.52	0.31
1401 BTQ	BTQ2	3145	251	252	91.8	3.9	1.86	1.89	0.48
1402 BTQ	BTQ2	3146	252	253	93.5	3.7	1.02	1.08	0.4
1403 BTQ	BTQ2	3147	253	254	89.5	6.6	1.93	1.31	0.52
1404 BTQ	BTQ2	3148	254	255	88.5	7.6	2.15	1.15	0.52
1405 BUR	BUR	4084	0	1	77.7	15.6	5.19	1.09	0.33
1406 BUR	BUR	4085	1	2	81.1	1.94	11.29	2.69	0.61
1407 BUR	BUR	4086	2	3	81.4	1.81	11.92	2.37	0.62
1408 BUR	BUR	4087	3	4	76.4	0.85	20.35	0.83	0.62
1409 BUR	BUR	4088	4	5	91	1.36	5.66	0.69	0.28
1410 BUR	BUR	4089	5	6	93.5	1.47	3.46	0.87	0.33



KY Limestone Resources

University of Kentucky
Kentucky Geological Survey

Limestone and Dolomite Resources of Kentucky

Find site, place, or address

WPA Quarry limestone quarry

- County: Powell
- Quadrangle: Stanton
- Site ID: 359
- Lat, Lon: 37.83066, 83.840752

Site was sampled: 57 samples (open report)
Site has 57 chemical analyses (view samples report)
Related Publication(s) (display info)
View Field Notes (PDF)

Site photo:

Site report for:
WPA Quarry limestone quarry

- County: Powell
- Quadrangle: Stanton
- Site ID: 359

Site Events

Ledge Chemistry

Increment Sample Average
Site sampled from the bottom.

export to CSV

Section	Ledge	Unit	Bottom (ft)	Top (ft)	No. of Samples	CaCO ₃	MgCO ₃	Al ₂ O ₃	Fe ₂ O ₃	SiO ₂	P	S	Lithology Descript
1	11	Point Creek Limestone	66	68	2	97.46	0.92	0	0.22	1.46	0	0	
1	10	Point Creek Limestone	45	66	21	88.76	5.53	0.55	0.22	4.37	0	0	
1	9	Renaute Limestone	40	45	5	97.87	0.45	0.02	0.28	1.15	0	0	
1	8	Renaute Limestone	30	40	10	90.78	1.45	0.53	0.4	6.1	0	0	
1	7	Renaute Limestone	23	30	7	98.37	0.44	0.04	0.26	0.73	0	0	

KGS IT / Data Infrastructure:



- Database Records:

Over 16 million records – populated since ~1970's

Relational Database (SQL)

- Publications
- Oil and gas
- Coal
- Groundwater
- Minerals
- Geotechnical
- Geologic mapping
- Research project data

- Document Scanning and Archiving:

- Oil and gas documents: 770,000+ scanned (page-sized and elogs)
- Publications: 7000+ scanned documents
- Past research documents (field notes, data sheets)
- Water well and spring documents*: ~800,000 documents

*from KY Division of Water – no KGS scanning

KGS Website:



- Via KGS website: static and dynamic web pages and services
- Since 1996: KGS has had a website – one of most popular at UK
- Since 2001: started development of dynamic and internet map services
 - Access to KGS database records and GIS data
 - Access to scanned documents
 - All data and map services provided at no cost to users

Geologic information:

Rocks and minerals ... Fossils ... General KY Geology ... Geoscience Education Resources ... Geologic Mapping ... Coal Research ... Oil and Natural Gas Research ... Carbon Sequestration ... Karst Information ... Groundwater Research ... Earthquakes ... Landslides ... Foundation Engineering

Data services:

Publications and maps catalog ... Oil and gas well records ... Water wells and springs records ... Groundwater quality ... Coal borehole ... Coal thickness and quality measurements ... Well sample and core holdings ... KYTC geotechnical data ... KGS photos and images ... Geologic descriptions ... Coordinate conversion ... GIS map services and downloads

Map Services:

Geologic Map Service (Oil and gas wells...Water wells and springs...Coal data...Sinkholes...Landslides...Non-coal minerals and quarries...Core holdings...Outcrops) ... Water Wells and Springs ... Groundwater Quality ... Coal Resource Information ... Minerals Information ... Kentucky Arches ... Oil and Gas Permits ... Class I and Class II Wells ... Kentucky Energy Infrastructure ... Landslide Information ... Story Map Tours

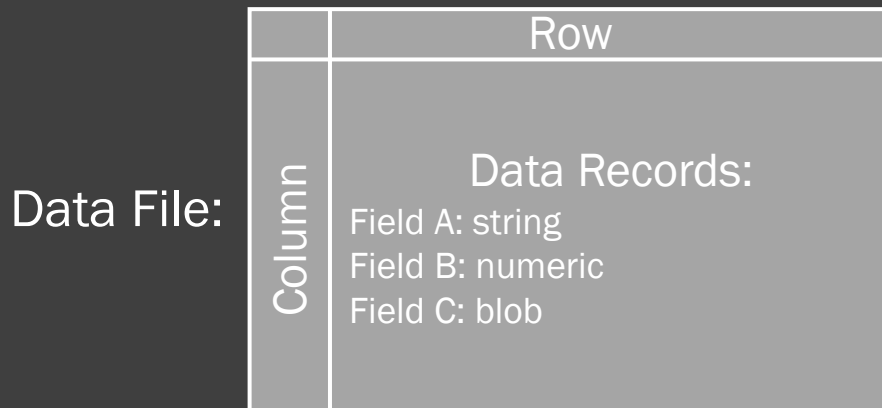
Data Structure



2 basic types – both are valid as databases

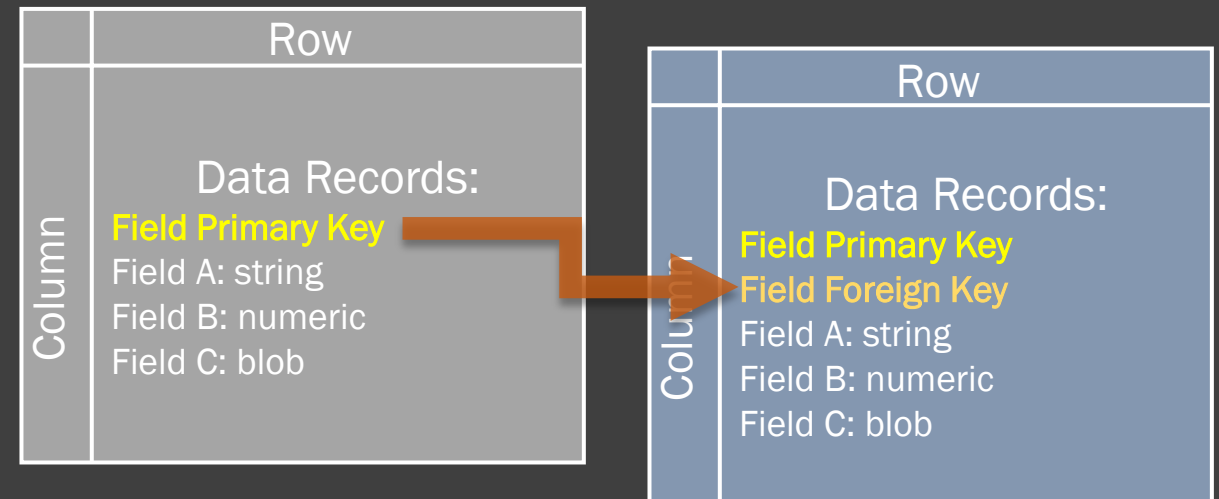
Flat File Database

- Single table of data
- Contains files, records, fields



Relational Database

- Multiple tables
- Records cross-referenced between tables (related)



Data Structure

Flat File vs Relational Database

	A	B	C	D	E	F	G	H	I	J
1	SiteID	SiteName	Sampled_from	Fieldnotes	north_latitude	west_longitude	quadrangle_name	county_name	EventDate	EventType
2	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	26-Jul-49	increment sampling
3	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	12-Apr-46	physical testing
4	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	24-Jun-49	physical testing
5	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	25-Sep-53	ledge description
6	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	08-Oct-46	physical testing
7	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	04-May-49	physical testing
8	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	15-Nov-50	physical testing
9	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	26-Aug-53	increment sampling
10	115	Valley Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter	16-May-49	physical testing
11	115	Valley Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter	09-Aug-54	increment sampling



Data Structure

Flat File vs Relational Database

Benefits:

- Simple design
- Simple format (no interpreter software required)
- Easy to populate
- Easy to transport (email, etc)
- Easy to read
- Can be useful for very large unstructured datasets (lots of records)

Considerations:

- Duplicate data
- Many fields
- Prone to data inconsistencies
- Merging data difficult
- Searching / analysis can be difficult

Benefits:					Considerations:					
1	SiteID	SiteName	Sampled_from	Fieldnotes	north_latitude	west_longitude	quadrangle_name	county_name	EventDate	EventType
2	105 Lawton Quarry/mine	B	-1		38.262356	-83.220174	Olive Hill	Carter	26-Jul-49	increment sampling
3	105 Lawton Quarry/mine	B	-1		38.262356	-83.220174	Olive Hill	Carter	12-Apr-46	physical testing
4	105 Lawton Quarry/mine	B	-1		38.262356	-83.220174	Olive Hill	Carter	24-Jun-49	physical testing
5	105 Lawton Quarry/mine	B	-1		38.262356	-83.220174	Olive Hill	Carter	25-Sep-53	ledge description
6	113 Carter Quarry	B	-1		38.439686	-83.168844	Wesleyville	Carter	08-Oct-46	physical testing
7	113 Carter Quarry	B	-1		38.439686	-83.168844	Wesleyville	Carter	04-May-49	physical testing
8	113 Carter Quarry	B	-1		38.439686	-83.168844	Wesleyville	Carter	15-Nov-50	physical testing
9	113 Carter Quarry	B	-1		38.439686	-83.168844	Wesleyville	Carter	26-Aug-53	increment sampling
10	115 Wallisville Mine No. 1	E	-1		38.316405	-83.13586	Olive Hill	Carter	16-May-49	physical testing
11	115 Wallisville Mine No. 1	E	-1		38.316405	-83.13586	Olive Hill	Carter	09-Aug-54	increment sampling

- Simple design
- Simple format (no interpreter software required)
- Easy to populate
- Easy to transport (email, etc)
- Easy to read
- Can be useful for very large unstructured datasets (lots of records)

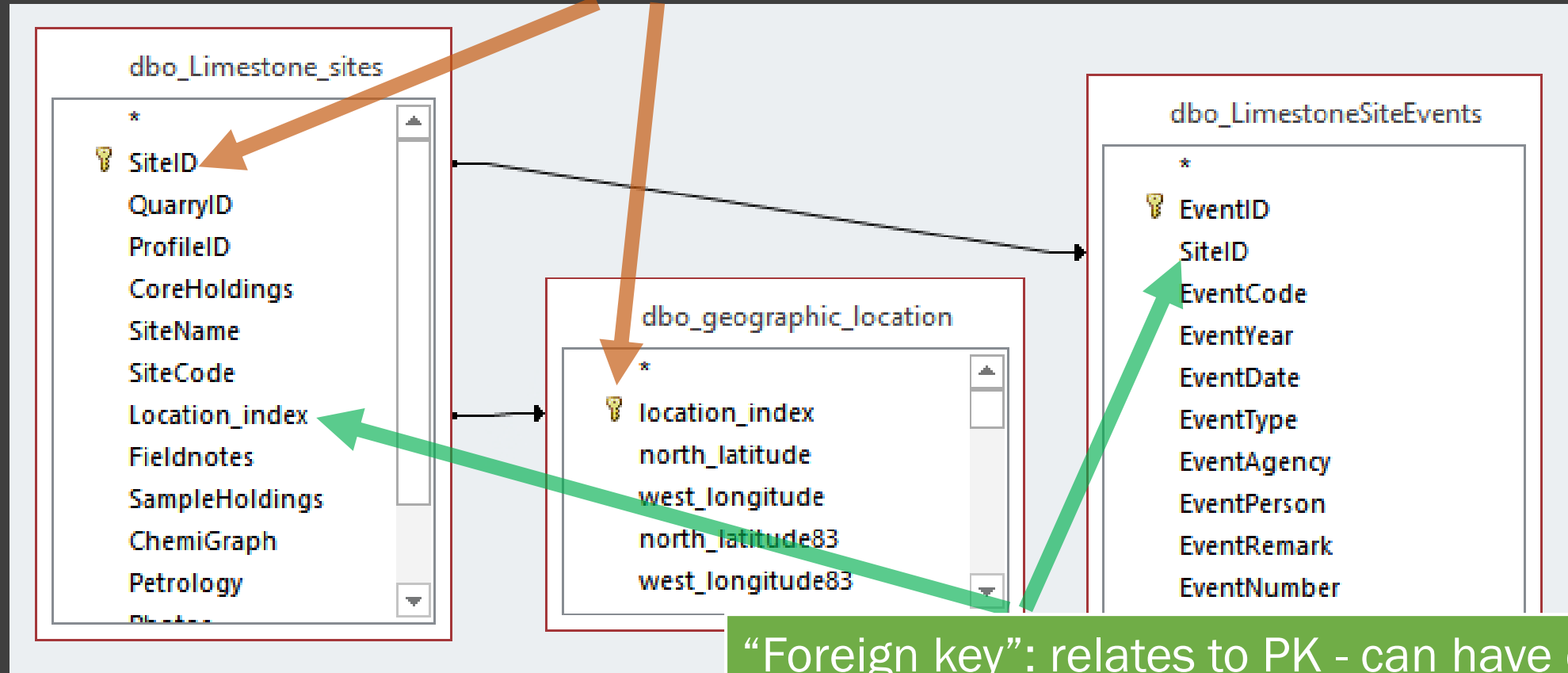
- Duplicate data
- Many fields
- Prone to data inconsistencies
- Merging data difficult
- Searching / analysis can be difficult



Data Structure

Flat File vs Relational Database

“Primary key”: unique identifiers





Data Structure

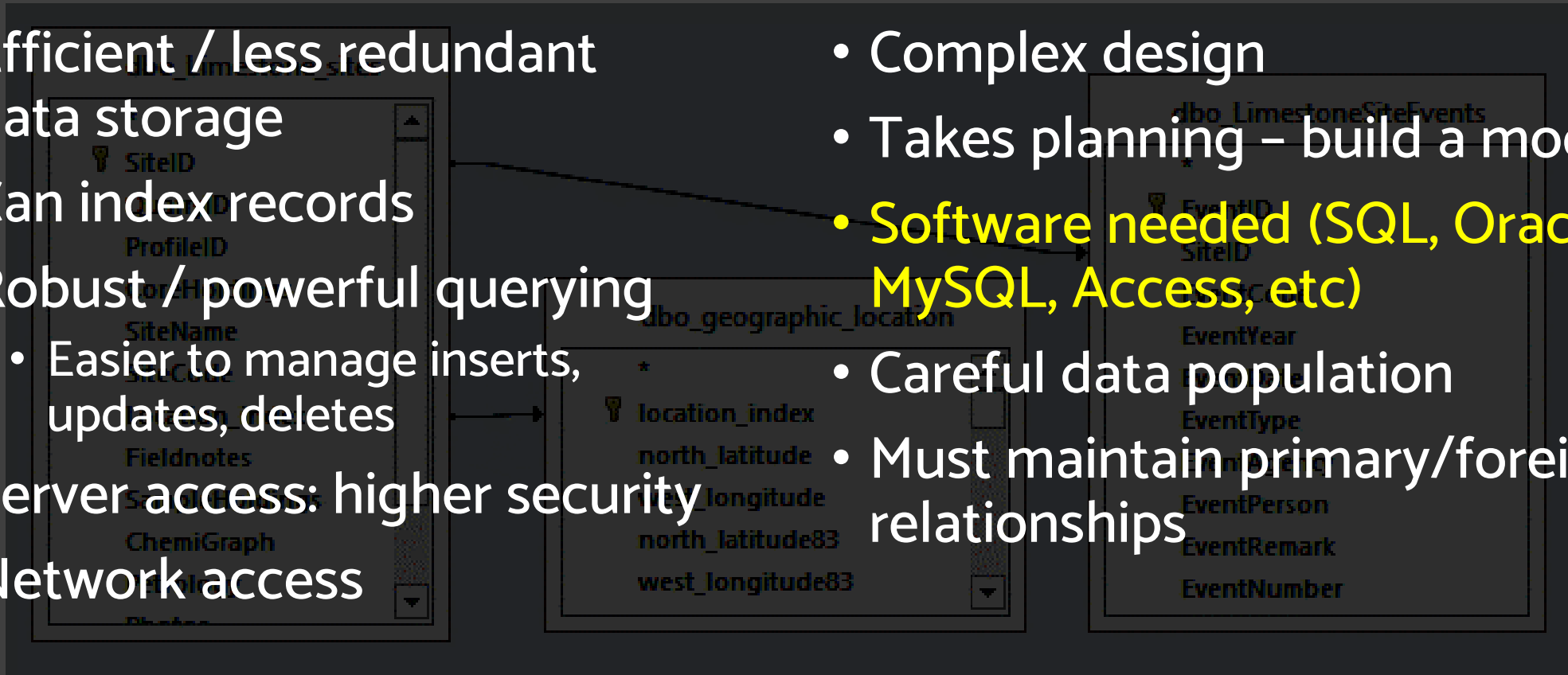
Flat File vs Relational Database

Benefits:

- Efficient / less redundant data storage
- Can index records
- Robust / powerful querying
 - Easier to manage inserts, updates, deletes
- Server access: higher security
- Network access

Considerations:

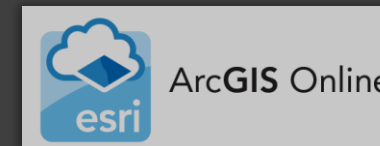
- Complex design
- Takes planning – build a model
- **Software needed (SQL, Oracle, MySQL, Access, etc)**
- Careful data population
- Must maintain primary/foreign key relationships



Design and build a database structure (schema/model)...



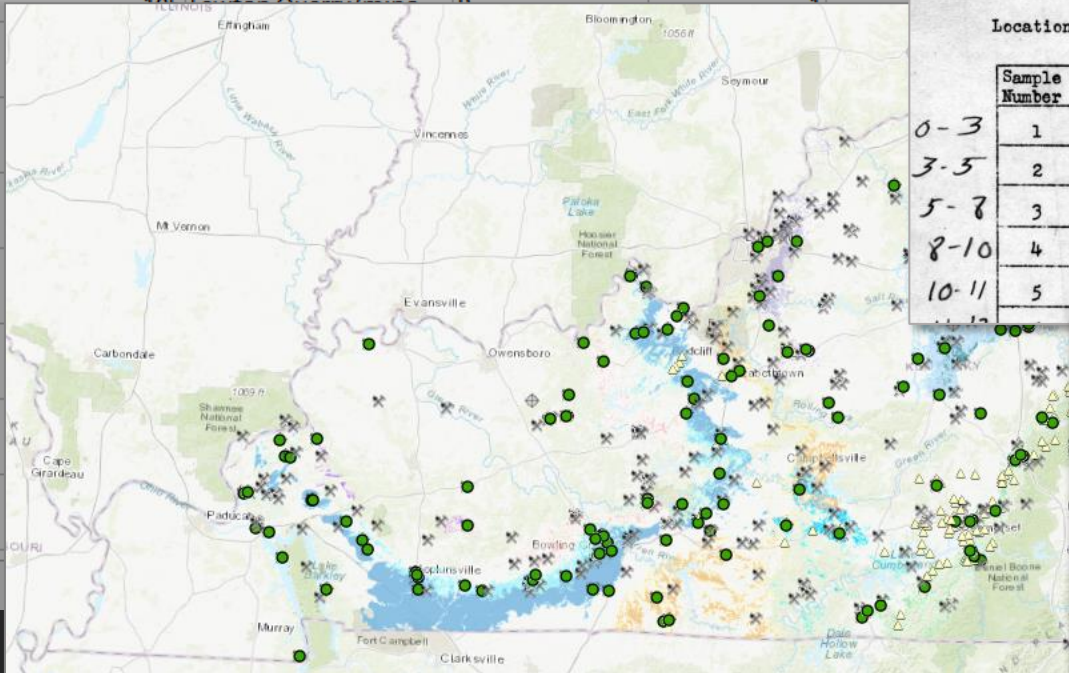
- Need database software/service – pick a flavor



Building a data structure: flat data to relational database: example with KGS limestone sites database



	A	B	C	D	E	F	G	H	I	J
1	SiteID	SiteName	Sampled_from	Fieldnotes	north_latitude	west_longitude	quadrangle_name	county_name	EventDate	EventType
2	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	26-Jul-49	increment sampling
3	105	Lawton Quarry/mine	B	-1					12-Apr-46	physical testing
4									24-Jun-49	physical testing
5									25-Sep-53	ledge description
6									08-Oct-46	physical testing
7									04-May-49	physical testing
8									15-Nov-50	physical testing
9									g-53	increment sampling
10									y-49	physical testing
11									g-54	increment sampling



ANALYSIS OF LIMESTONE SAMPLES

Location Menifee Date Aug 4, 1953

Sample Number	SiO ₂ %	Fe ₂ O ₃ %	Al ₂ O ₃ %	Ca ₃ P ₂ O ₈ %	CaCO ₃ %	MgCO ₃ %	Total %
0-3							
1	5.28	.32	.50		91.4	1.19	98.7
3-5							
2	2.71	.29	.11		95.0	.83	98.9
5-8							
3	4.14	.26	.79		92.4	1.22	98.8
8-10							
4	1.52	.18	.47		95.4	.91	98.5
10-11							
5	3.57	.56	.40		93.4	1.24	99.2

2581



Flat data to relational database... a journey



	A	B	C	D	E	F	G	H	I	J
1	SiteID	SiteName	Sampled_from	Fieldnotes	north_latitude	west_longitude	quadrangle_name	county_name	EventDate	EventType
2	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	26-Jul-49	increment sampling
3	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	12-Apr-46	physical testing
4	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	24-Jun-49	physical testing
5	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	25-Sep-53	ledge description
6	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	08-Oct-46	physical testing
7	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	04-May-49	physical testing
8	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	15-Nov-50	physical testing
9	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	26-Aug-53	increment sampling
10	115	Valley Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter	16-May-49	physical testing
11	115	Valley Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter	09-Aug-54	increment sampling



Identify “data segments” in flat data (or unstructured data)

Site Data

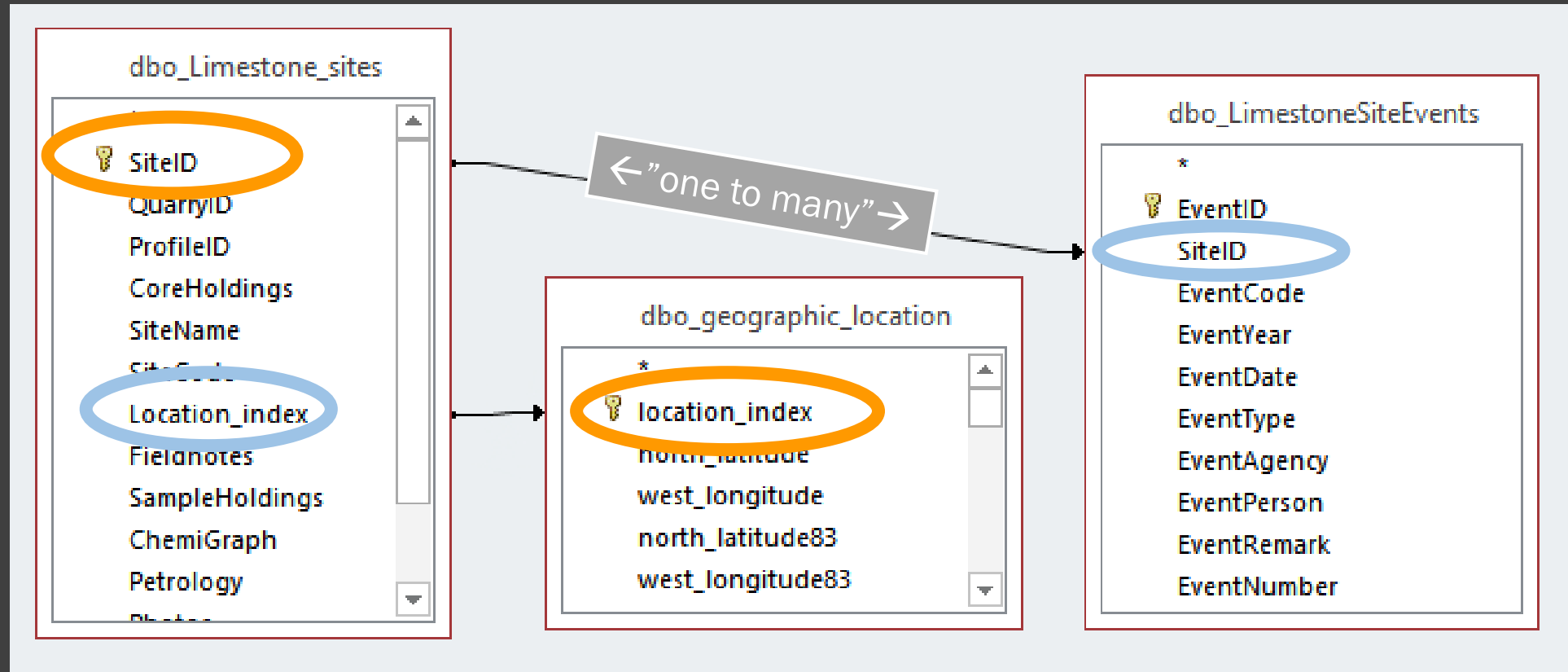
Location

Events

	A	B	C	D	E	F	G	H	I	J
	SiteID	SiteName	Sampled_from	Fieldnotes	north_latitude	west_longitude	quadrangle_name	county_name	EventDate	EventType
1										
2	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	26-Jul-49	increment sampling
3	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	12-Apr-46	physical testing
4	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	24-Jun-49	physical testing
5	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	25-Sep-53	increment sampling
6	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	08-Oct-50	physical testing
7	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	16-May-49	physical testing
8	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	26-Aug-53	increment sampling
9	113	Carter Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter	16-May-49	physical testing
10	115	Valley Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter	09-Aug-54	increment sampling
11	115	Valley Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter		

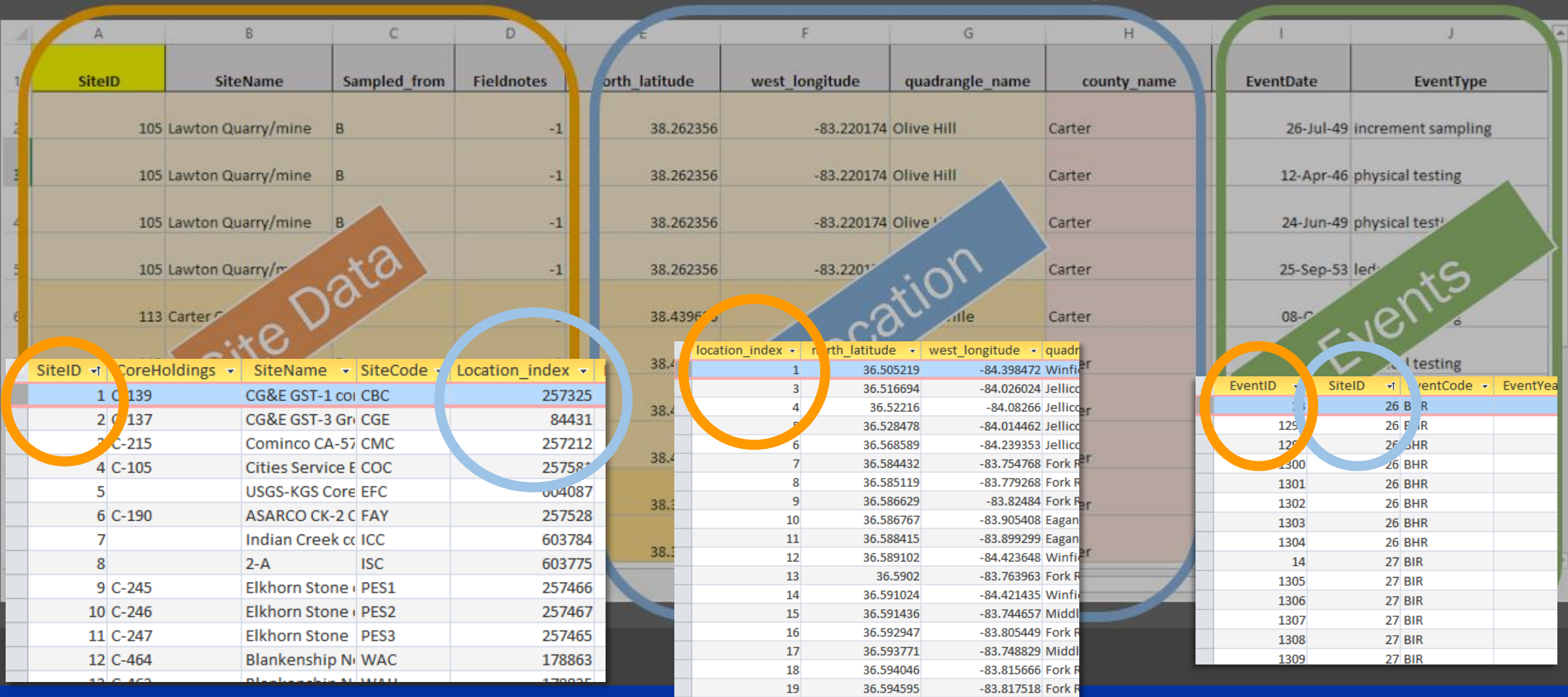
Design and build a database structure (schema/model)...

- Need database software – pick a flavor
- Design and build table schemas with **primary** and **foreign** keys



Migrate data from flat tables / populate tables in the relational database 

****maintain relationships with **primary** - to - foreign keys!**



SiteID	SiteName	Sampled_from	Fieldnotes	north_latitude	west_longitude	quadrangle_name	county_name	EventDate	EventType
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	26-Jul-49	increment sampling
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	12-Apr-46	physical testing
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	24-Jun-49	physical test
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter	25-Sep-53	led
113	Carter C			38.4396		file	Carter	08-C	

SiteID	CoreHoldings	SiteName	SiteCode	Location_index
1	C-139	CG&E GST-1 col	CBC	257325
2	C-137	CG&E GST-3 Gr	CGE	84431
3	C-215	Cominco CA-57	CMC	257212
4	C-105	Cities Service E	COC	257581
5		USGS-KGS Core	EFC	604087
6	C-190	ASARCO CK-2 C	FAY	257528
7		Indian Creek c	ICC	603784
8		2-A	ISC	603775
9	C-245	Elkhorn Stone	PES1	257466
10	C-246	Elkhorn Stone	PES2	257467
11	C-247	Elkhorn Stone	PES3	257465
12	C-464	Blankenship N	WAC	178863
13	C-463	Blankenship N	WAC	178865

location_index	north_latitude	west_longitude	quadrangle_name
1	36.505219	-84.398472	Winfi
3	36.516694	-84.026024	Jellico
4	36.52216	-84.08266	Jellico
5	36.528478	-84.014462	Jellico
6	36.568589	-84.239353	Jellico
7	36.584432	-83.754768	Fork R
8	36.585119	-83.779268	Fork R
9	36.586629	-83.82484	Fork R
10	36.586767	-83.905408	Eagan
11	36.588415	-83.899299	Eagan
12	36.589102	-84.423648	Winfi
13	36.5902	-83.763963	Fork R
14	36.591024	-84.421435	Winfi
15	36.591436	-83.744657	Middl
16	36.592947	-83.805449	Fork R
17	36.593771	-83.748829	Middl
18	36.594046	-83.815666	Fork R
19	36.594595	-83.817518	Fork R

EventID	SiteID	EventCode	EventYear
125	26	BHR	
126	26	BHR	
1300	26	BHR	
1301	26	BHR	
1302	26	BHR	
1303	26	BHR	
1304	26	BHR	
14	27	BIR	
1305	27	BIR	
1306	27	BIR	
1307	27	BIR	
1308	27	BIR	
1309	27	BIR	

Methods for populating a relational database

KENTUCKY GEOLOGICAL SURVEY
Quarry Report No. 2

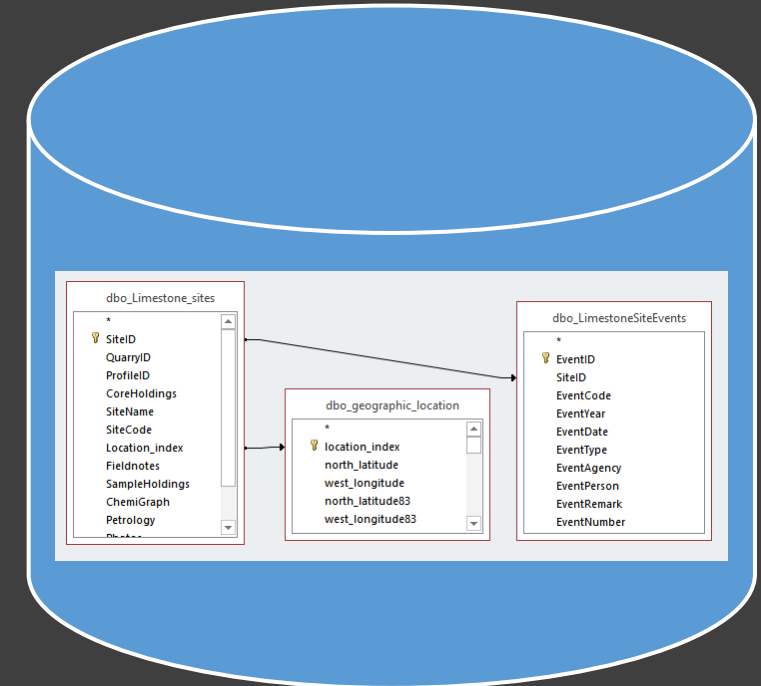
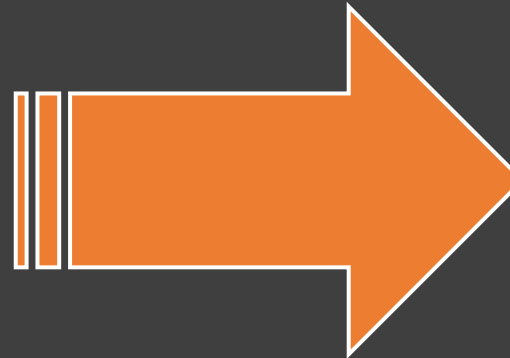
County: Anderson Property Owner: Ky. Stone Co. Operator: Ky. Stone Co. Date Sampled: _____
Location: _____ Date Received: _____
Sampled By: _____ Analyzed By: _____ Date Reported: _____

CHEMICAL ANALYSES										DESCRIPTION	
% CaCO ₃	% MgCO ₃	% SiO ₂	% Iron Oxide	% Alumina Total	% Sample Level	Ledge No.	Thick-ness (Feet)	LITHOLOGY	STRATI-GRAPHY		
55.7	37.0	2.80	0.44	0.55	96.5	19-20		Limestone, craggy, gray to buff, locally has dark nodules			
70.3	23.2	2.92	0.27	0.54	97.2	18-19		Fireclay, dense, calcareous, shaly			
70.7	23.0	2.70	0.28	0.39	97.1	17-18		Limestone, light gray to tan, shaly			
58.5	33.4	3.28	0.37	0.94	96.5	16-17	3	Limestone, light gray to tan, shaly			
61.4	32.3	4.67	0.17	1.48	97.7	15-16	13	Limestone, light gray to tan, shaly			
63.9	29.8	4.47	0.44	1.03	97.7	14-15		Limestone, light gray to tan, shaly			
62.4	31.5	4.04	0.47	0.97	97.4	13-14		Limestone, light gray to tan, shaly			
59.2	35.2	4.00	0.33	0.76	97.5	12-13		Limestone, light gray to tan, shaly			
57.1	37.4	3.65	0.37	1.05	97.6	11-12		Limestone, light gray to tan, shaly			
56.4	38.8	3.46	0.40	0.66	99.7	10-11		Limestone, light gray to tan, shaly			
59.3	34.8	4.00	0.50	1.19	99.8	9-10		Limestone, light gray to tan, shaly			
61.8	38.6	3.52	0.38	0.92	99.2	8-9		Limestone, light gray to tan, shaly			
82.5	12.4	3.97	0.23	0.50	99.1	7-8		Limestone, light gray to tan, shaly			
78.5	15.6	3.80	0.30	0.92	99.1	6-7		Limestone, light gray to tan, shaly			
76.8	16.9	4.00	0.33	1.03	99.3	5-6	2	Limestone, light gray to tan, shaly			
81.1	8.6	3.10	0.27	0.73	99.8	4-5		Limestone, light gray to tan, shaly			
92.5	5.0	2.31	0.15	0.13	100.1	3-4		Limestone, light gray to tan, shaly			
90.7	5.2	3.23	0.20	0.45	99.8	2-3		Limestone, light gray to tan, shaly			
85.7	5.2	5.78	0.47	1.89	99.0	1-2	1	Limestone, light gray to tan, shaly			
85.3	4.5	5.36	0.42	1.86	99.5	0-1	2	Limestone, light gray to tan, shaly			

BOTTOM OF QUARRY

KGS Limestone Results.xlsx - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Site	Suite_ID	Lab_ID	From	To	CaCO3	MgCO3	SiO2	Al2O3	Fe2O3			
1393	BTQ	BTQ2	3137	243	244	90.5	1.77	1.65	5.41	0.33			
1394	BTQ	BTQ2	3138	244	245	94	1.7	1.27	2.03	0.28			
1395	BTQ	BTQ2	3139	245	246	93	3.3	1.54	1.57	0.46			
1396	BTQ	BTQ2	3140	246	247	93.8	3.4	1.54	0.83	0.31			
1397	BTQ	BTQ2	3141	247	248	90	4.8	2.75	1.68	0.6			
1398	BTQ	BTQ2	3142	248	249	86	5.8	4.49	2.76	0.53			
1399	BTQ	BTQ2	3143	249	250	84.7	6.2	4.7	3.64	0.65			
1400	BTQ	BTQ2	3144	250	251	93.4	4.3	1.32	0.52	0.31			
1401	BTQ	BTQ2	3145	251	252	91.8	3.9	1.86	1.89	0.48			
1402	BTQ	BTQ2	3146	252	253	93.5	3.7	1.02	1.08	0.4			
1403	BTQ	BTQ2	3147	253	254	89.5	6.6	1.93	1.31	0.52			
1404	BTQ	BTQ2	3148	254	255	88.5	7.6	2.15	1.15	0.52			
1405	BUR	BUR	4084	0	1	77.7	15.6	5.19	1.09	0.33		BUR	
1406	BUR	BUR	4085	1	2	81.1	1.94	11.29	2.69	0.61			
1407	BUR	BUR	4086	2	3	81.4	1.81	11.92	2.37	0.62			
1408	BUR	BUR	4087	3	4	76.4	0.85	20.35	0.83	0.62			
1409	BUR	BUR	4088	4	5	91	1.36	5.66	0.69	0.28			
1410	BUR	BUR	4089	5	6	93.5	1.47	3.46	0.87	0.33			



Populating a relational database



Spreadsheet conversion: somewhat “brute” force

- deal with existing data
- one-time processing

The screenshot shows the Microsoft Excel interface with the 'DATA' tab selected. The spreadsheet contains data in columns A through H. A 'Remove Duplicates' dialog box is open, displaying the message: 'To delete duplicate values, select one or more columns that contain duplicates.' The dialog box has a question mark icon and an 'X' button.

A	B	C	D	E	F	G	H
SiteID	SiteName	Sampled	Fieldnote	north_latitude	west_longitude	quadrangle_name	county_name
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
105	Lawton Quarry/mine	B	-1	38.26			
113	Carter Quarry	B	-1	38.43			
113	Carter Quarry	B	-1	38.43			
113	Carter Quarry	B					
113	Carter Quarry	B					
115	Valley Stone No.1	B					
115	Valley Stone No.1	B					
119	Casey Stone	B					
119	Casey Stone	B					
119	Casey Stone	B					
119	Casey Stone	B					
119	Casey Stone	B					
119	Casey Stone	B					

The screenshot shows the Microsoft Excel interface with the 'DATA' tab selected. The spreadsheet contains data in columns A through E. The data is as follows:

A	B	C	D	E
SiteID	EventDate	EventType		
105	18105	increment sampling		
105	16904	physical testing		
105	18073	physical testing		
105	19627	ledge description		
113	17083	physical testing		
113	18022	physical testing		
113	18582	physical testing		
113	19597	increment sampling		
115	18034	physical testing		
115	19945	increment sampling		
119	20568	increment sampling		
119	31212	ledge sampling		
119	17790	physical testing		
119	28173	lithology sampling		
119	30946	ledge description		
119	30958	field check		

- query table to build new tables
- deal with existing data
- one-time processing

Populating a relational database



Spreadsheet conversion: write a program (Python, VB, etc) to parse table

- requires programming
- can re-use
- could be efficient means for population

	A	B	C	D	E	F	G	H
1	SiteID	SiteName	Sampled from	Fieldnotes	north latitude	west longitude	quadrangle name	county name
2	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
3	105	Lawton Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
		Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
		Quarry/mine	B	-1	38.262356	-83.220174	Olive Hill	Carter
		Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter
		Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter
		Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter
		Quarry	B	-1	38.439686	-83.168844	Wesleyville	Carter
		Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter
		Stone No.1	B	-1	38.316405	-83.13586	Olive Hill	Carter



```
GeologicMappingProfilesReplica.py - N:\arcgisserver\ArcGIS_Projects_10pt1\CollectorProjects\GeologicMappingPr...
File Edit Format Run Options Window Help

    pull_to_local(attachment_file, attachment['id'],
        '', 'jpg')
    except urllib2.HTTPError:
        raise urllib2.HTTPError('httperror')

    group_photos(root_file, "ALL")

def pull_attachments(self, query, field):
    query['token'] = self.token
    os.chdir(self.destination)
    layers = get_service_info(self.fs_url, self.token)['layers']
    root_name = self.get_root_name() + "_Photos"
    if os.path.exists(root_name):
        shutil.rmtree(root_name)
    service_file = create_and_set_dir(root_name)
    attachments = get_service_info(self.layer_url,
        self.token)['hasAttachments']
    if self.layer_id and attachments:
        self.find_attachments(query, layers[int(self.layer_id)], field)
    else:
        for layer in layers:
            if get_service_info(add_path(self.fs_url, layer['id']),
                self.token)['hasAttachments']:
                os.chdir(service_file)
                query['objectIds'] = ''
                self.find_attachments(query, layer, field)

def replicate(self, query):
    query['token'] = self.token
    replica_url = add_path(self.fs_url, 'createReplica')
    #MUST ADD THE TOKEN TO THE URL:
    arcpy.AddMessage(replica_url)
    zip_url = get_response2(replica_url, query)['responseUrl']+"?token="+self.token
    arcpy.AddMessage(zip_url)
    zip_file = get_response(zip_url, get_json=False)
    zip_name = pull_to_local(zip_file, self.get_root_name(), self.destination, 'zip')
    return unzip_to_local(zip_name, self.destination)

def pull_replica(self, query):
    query['token'] = self.token
    layers = get_service_info(self.fs_url, self.token)['layers']
    if self.layer_id:
        query['layers'] = self.layer_id
        return self.replicate(query)
    else:
        query['layers'] = [layer['id'] for layer in layers]
        return self.replicate(query)
```

Populating a relational database



Direct data entry:

- can plan model and enter into datasheet views
- fast setup for data entry
- prone to data inconsistencies / errors & may not be very flexible for deployment

The screenshot displays the Microsoft Access interface for a database named 'LimestoneData'. On the left, the 'Navigation Pane' shows three tables: 'dbo_Limestone_sites', 'dbo_geographic_location', and 'dbo_LimestoneSiteEvents'. The 'dbo_Limestone_sites' table has fields: SiteID (primary key), QuarryID, ProfileID, CoreHoldings, SiteName, SiteCode, Location_index, Fieldnotes, SampleHoldings, ChemiGraph, and Petrology. The 'dbo_geographic_location' table has fields: location_index (primary key), north_latitude, west_longitude, north_latitude83, and west_longitude83. The 'dbo_LimestoneSiteEvents' table has fields: EventID (primary key), SiteID, and EventCode. Arrows indicate relationships: SiteID in 'dbo_Limestone_sites' is linked to SiteID in 'dbo_LimestoneSiteEvents', and Location_index in 'dbo_Limestone_sites' is linked to location_index in 'dbo_geographic_location'. On the right, the 'Table Tools' ribbon is active, showing the 'TABLE' tab. Below the ribbon, the 'Datasheet View' of the 'dbo_LimestoneSiteEvents' table is shown. The table has columns: ID, SiteID, SiteName, Sampled_frc, Fieldnotes, north_latitu, west_longit, quadrangle, county_nam, and EventDate. The data shows five records for 'Casey Stone' in 'Yosemite' county, with dates ranging from 1948 to 1984. A new record (ID 17) is being added, with SiteID 120 and SiteName empty.

ID	SiteID	SiteName	Sampled_frc	Fieldnotes	north_latitu	west_longit	quadrangle	county_nam	EventDate
13	119	Casey Stone	B	-1	37.25071	-84.767135	Yosemite	Casey	9/14/1948
14	119	Casey Stone	B	-1	37.25071	-84.767135	Yosemite	Casey	2/17/1977
15	119	Casey Stone	B	-1	37.25071	-84.767135	Yosemite	Casey	9/21/1984
16	119	Casey Stone	B	-1	37.25071	-84.767135	Yosemite	Casey	10/3/1984
17	120								
*	(New)								

Populating a relational database



Access (or similar) forms: easy to use front-end / requires setup

- long-term use, reduce data entry error, inexperienced users, less-flexible deployment
- incorporate other processes (this one converts and moves scanned images)

The collage illustrates the process of populating a relational database using Microsoft Access. It includes:

- Microsoft Access Form View:** Shows the 'Process the Scanned Images' form with fields for 'Scan Location' (i:\scanned_images\), 'Operator' (operator), and 'Originals' (0283.8 MEGs). It also features a 'STANDBY MODE' button and a 'Click AFTER Creating CD: T10325' instruction.
- VBA Code Editor:** Displays the code for the 'Process New Images' macro, which handles file operations, database updates, and error handling.
- Flowchart:** Details the logic for copying TIFF files from the 'unprocessed_images' folder to the 'processed_images' folder, checking for existing files, and updating the 'Scanlist.Status' to 'DUN'.

```
Microsoft Visual Basic for Applications - SCANS_TO_PDF_MOTHCOPY - [Form_Process New Images (Code)]

'Value = 7 digit
'Pvalue = 3 digit
RECNoPageNo = Rvalue & Pvalue

'checks if is in scan_list already - if so, send a warning - and DON'T PROCESS on
'checks RECNoPageNo - this catches both old and new scans
'will basically skip this document
'If DCount("[status]", "dbo_scanlist", "[ImageName]=' ' & ImageName & ' ' > 0 Then
'If DCount("[status]", "dbo_scanlist", "[RECNoPageNo]=' ' & RECNoPageNo & ' ' > 0 Then
IsMoved = False
Print #1, "ERROR: Image File: " & Rvalue & " Page: " & Pvalue & " (" & ScanName & "
Else
'Set rstImages = dbs.OpenRecordset("dbo_scanlist")
'this adds the values to the database: ![FieldName] in scanlist:
With rstImages
.AddNew
.RECNoPageNo = Rvalue & Pvalue
!CNumber = "DISK" ' "DISK" for all images - will update upon archive
!whole_eleg = is Elog
!ImageName = Trim(ImageName) 'updates with the TRIMMED image name: Dvalue
!PageNo = Rvalue
!Type = dbType
!Status = "NEW" 'this tracks through the process
!image_is = Pvalue
!image_length = size 'not tracking this anymore
!image_width = image_size(scan_loc & ScanName, "W") 'not tracking this anymore
!dpi = image_size(scan_loc & ScanName, "D") 'not tracking this anymore
!file_size = image_size(scan_loc & ScanName, "S") 'not tracking this anymore
!Updated = Format(Now(), "short date")
!work_loc = 0
.Update
counter = counter + 1
'End With
'move scan to unprocessed_images folder - on Lithos2!
'names using the scan name (which was put back together using ImageName & PageNo)
fs.MoveFile scan_loc & ScanName, "\\lithos2\OG_Scans\unprocessed_images\" & ScanName & ".tif"
```

Flowchart Logic:

```
graph TD
    Start([COPY TIFF file from  
\\lithos2\OG_Scans\unprocessed_images  
to  
\\lithos2\OG_Scans\processed_images\[accordinRECNO]\]) --> Decision{Check if Elog TIFF And Elog JPEG exist in  
\\lithos2\OG_Scans\processed_images\[a  
ccordinRECNO]\}
    Decision -- Yes --> Update([Update  
Scanlist.Status = DUN  
also:  
ModifyDate = current date  
QCPerson  
QCLevel = "INIT"  
Updated = current date  
for the page])
    Decision -- No --> Message[Message  
r = TRUE]
```


Populating a relational database



Web forms: easy to use front-end / may require extensive programming

- long-term use, reduce data entry error, inexperienced users, flexible deployment
- need to host on a server or via web services

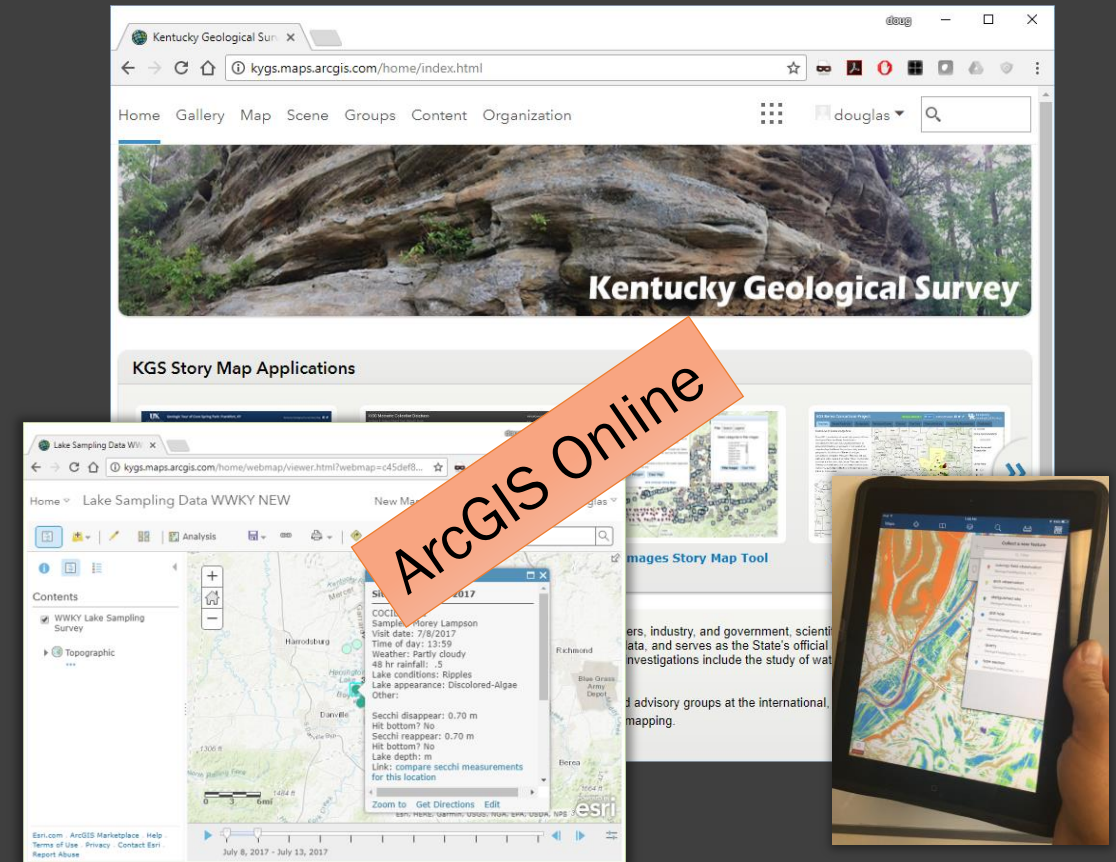
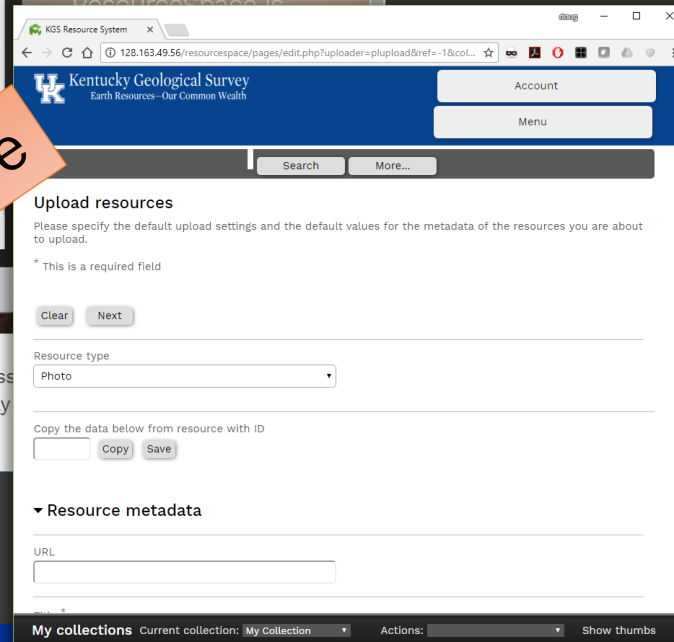
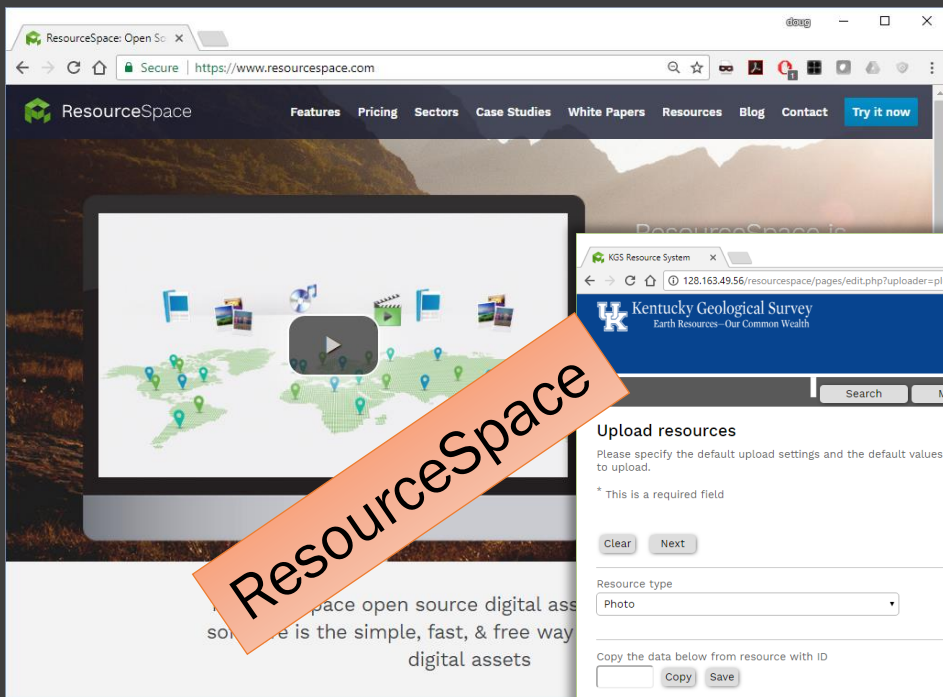
```
'update the DescriptionSource Table:
'only do if there was no error with the user profile:
If updateVer = TRUE then
    dim updatesql
    updatesql = "INSERT INTO DescriptionSource (SourceType,SourceRank,SourceDescription,ForeignID) "&_
        "VALUES ('Person',6,'"&src_name&"','&pro_scribe&');"
'Execute the statement
sqlReturn = executeConnection(oConnUP,updatesql)
If sqlReturn= TRUE then
    updateMsg = updateMsg&"<strong>***A NEW ENTRY INTO THE DescriptionSource TABLE WAS MADE!</strong>"
Else
    updateMsg = updateMsg&"<strong>***database update script stopped**</strong>"
    updateMsg = updateMsg&"<br><strong>***UPDATE OF THE DescriptionSource TABLE FAILED!</strong>"
    updateMsg = updateMsg&sqlReturn
    updateVer = FALSE
End If
End If

'Now requery to check to get the description SourceID:
sqlUP = "SELECT DescriptionSource.* FROM DescriptionSource WHERE (((DescriptionSource.ForeignID)='&pro_s_
sqlReturn = oRsOpenFunction(oConnUP,oRs_up,sqlUP)
If sqlReturn = TRUE Then
    If NOT oRs_up.EOF Then
        'just get the sourceID:
        srcID = trim(oRs_up("SourceID"))
    Else
```

Populating a relational database

Cloud web services: relatively new frontier - lots of options

- launch relatively quickly, no hardware/software maintenance
- data maintenance / entry tools available, long term use (just pay the fee!), flexible deployment
- depending on service – may not be clear on how data is stored – but, do you care?



Why build a relational database?

- Organize and sustain your related data
- Typically server-based: centralize and serve to organization / public
 - Bonus: can secure your data
- Can export data into “flat tables” using queries
- Can make “views” to simplify data visualization

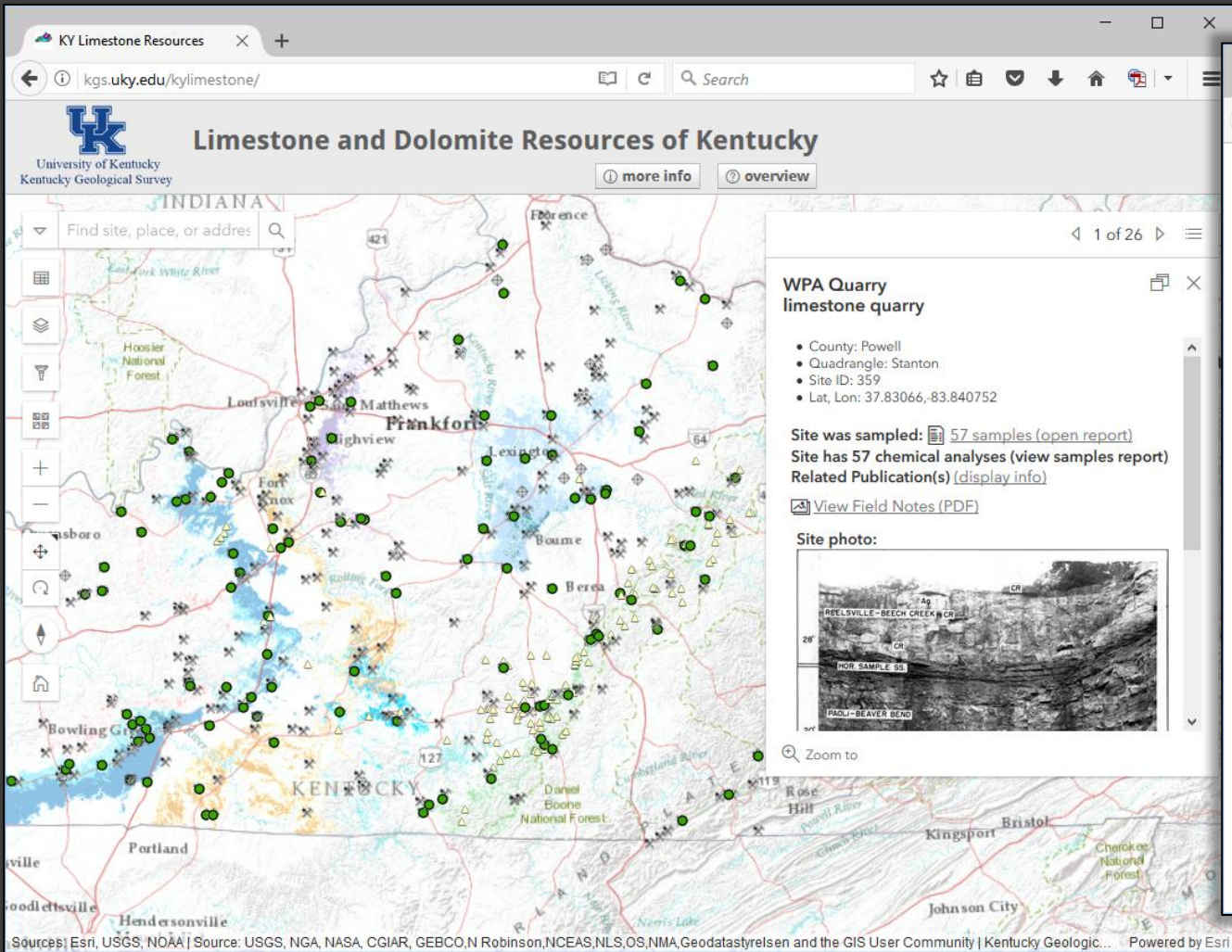
But:

- Careful planning / knowing your data is essential
- Helpful skills:
 - programming (SQL, Python, VB, etc)
 - server hardware / software knowledge
 - spreadsheet wrangling

KGS IT / Data Infrastructure:

- In-house data storage:
 - SQL server 2014 (virtualized Win Server 2012) – relational database
 - File server (virtualized Win Server 2012)
- Data Management:
 - Microsoft Suite (Excel, Access, etc)
 - Adobe Suite (Acrobat, Photoshop, Illustrator)
 - ArcGIS
 - ResourceSpace (newbies)
- ArcGIS Server / ArcGIS Online org account
- Web Presence:
 - IIS (virtualized Win Server 2008)
 - Backend: ASP classic / PHP
 - Frontend: HTML 5 (Javascript/HTML/CSS / Dojo, JQuery, Highcharts, etc)
 - ESRI Javascript API (map services)

Biggest benefit for KGS: Data Dissemination



KY Limestone and Dolomite

University of Kentucky
Kentucky Geological Survey

Limestone and Dolomite Resources of Kentucky Site Report

info / definitions

Site report for:
WPA Quarry limestone quarry

- County: Powell
- Quadrangle: Stanton
- Site ID: 359

Site Events

Ledge Chemistry

Increment Sample Average
Site sampled from the bottom.

export to CSV

Section	Ledge	Unit	Bottom (ft)	Top (ft)	No. of Samples	CaCO3	MgCO3	Al2O3	Fe2O3	SiO2	P	S	Lithology	Descripti
1	11	Paint Creek Limestone	66	68	2	97.46	0.92	0	0.22	1.46	0	0		
1	10	Paint Creek Limestone	45	66	21	88.76	5.53	0.55	0.22	4.37	0	0		
1	9	Renault Limestone	40	45	5	97.87	0.45	0.02	0.28	1.15	0	0		
1	8	Renault Limestone	30	40	10	90.78	1.45	0.53	0.4	6.1	0	0		
1	7	Renault Limestone	23	30	7	98.37	0.44	0.04	0.26	0.73	0	0		

Thanks!