

INL/CAES Big Data Meeting- Idaho State University

Keith T. Weber, GIS Director

Randy Gaines, CIO

Mark Norviel, NetCom Manager

Our Definition of Big Data

- Digital data that when compiled for a given research project exceeds 1 TB in size
- May be structured or unstructured

Challenges with Big Data

- Storage (*where can we store terabytes of data?*)
- Movement (*how long will it take to transfer these data over our network?*)
- Access (*once stored, how do we access it in the future?*)
- Performance/computation (*current big data projects need to actively use these data*)

Research at ISU

- Not all research at ISU can be considered “Big Data”
- But, some of our research is...
 - Geographic Information Science
 - Imaging
 - Data Visualization

GIScience

- Satellite imagery
 - NASA RECOVER Wildfire DSS
 - MODIS, SMAP, and Landsat

Spatial Domain 1: "Great Basin States" (Idaho, Utah, Arizona, Nevada, California, Oregon): ~ 113 Scenes										
Volumes shown in TB										
# Bands # Years	1	2	3	4	5	6	7	8	9	10
1	0.40	0.79	1.19	1.59	1.98	2.38	2.78	3.17	3.57	3.97
2	0.79	1.59	2.38	3.17	3.97	4.76	5.55	6.35	7.14	7.93
3	1.19	2.38	3.57	4.76	5.95	7.14	8.33	9.52	10.71	11.90
4	1.59	3.17	4.76	6.35	7.93	9.52	11.10	12.69	14.28	15.86
5	1.98	3.97	5.95	7.93	9.91	11.90	13.88	15.86	17.85	19.83
6	2.38	4.76	7.14	9.52	11.90	14.28	16.66	19.04	21.42	23.79
7	2.78	5.55	8.33	11.10	13.88	16.66	19.43	22.21	24.98	27.76
8	3.17	6.35	9.52	12.69	15.86	19.04	22.21	25.38	28.55	31.73
9	3.57	7.14	10.71	14.28	17.85	21.42	24.98	28.55	32.12	35.69
10	3.97	7.93	11.90	15.86	19.83	23.79	27.76	31.73	35.69	39.66
11	4.36	8.72	13.09	17.45	21.81	26.17	30.54	34.90	39.26	43.62
12	4.76	9.52	14.28	19.04	23.79	28.55	33.31	38.07	42.83	47.59
13	5.16	10.31	15.47	20.62	25.78	30.93	36.09	41.24	46.40	51.55
14	5.55	11.10	16.66	22.21	27.76	33.31	38.86	44.42	49.97	55.52
15	5.95	11.90	17.85	23.79	29.74	35.69	41.64	47.59	53.54	59.49
16	6.35	12.69	19.04	25.38	31.73	38.07	44.42	50.76	57.11	63.45
17	6.74	13.48	20.23	26.97	33.71	40.45	47.19	53.93	60.68	67.42
18	7.14	14.28	21.42	28.55	35.69	42.83	49.97	57.11	64.25	71.38
19	7.53	15.07	22.60	30.14	37.67	45.21	52.74	60.28	67.81	75.35
20	7.93	15.86	23.79	31.73	39.66	47.59	55.52	63.45	71.38	79.32

Max Volume
= 79.32 TB

Looking at data volumes with different combinations of # 1-band images (columns) and # years (rows)
 Assume (365/16) = ~23 overpasses per scene in one year
 Assume <=160MB for each 16-bit band
 Total Data Volume = # scenes in AOI * 23 overpasses per year * # years * # bands * 160MB

GIScience (cont'd)

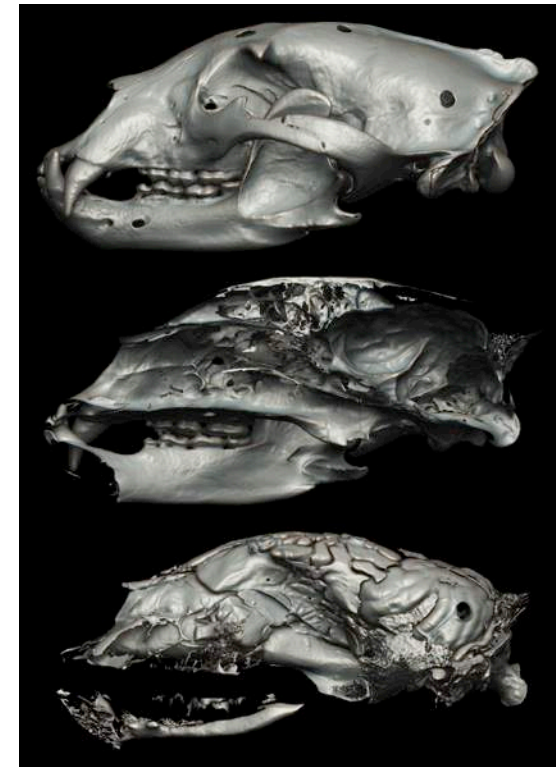
- These Big Data can be stored on a **very large** server
 - *Space can be saved* by compressing with ZIP
 - But ready access then becomes an issue
 - Tape drives are not an option for *active research*

Geospatial Imaging

- UAV/UAS
 - 20GB per daily collection
 - 50 collections = 1TB of raw, unprocessed data
 - Current ortho process requires 2-3 days for each collection
 - This project is computationally challenging because it is Big Data

Imaging

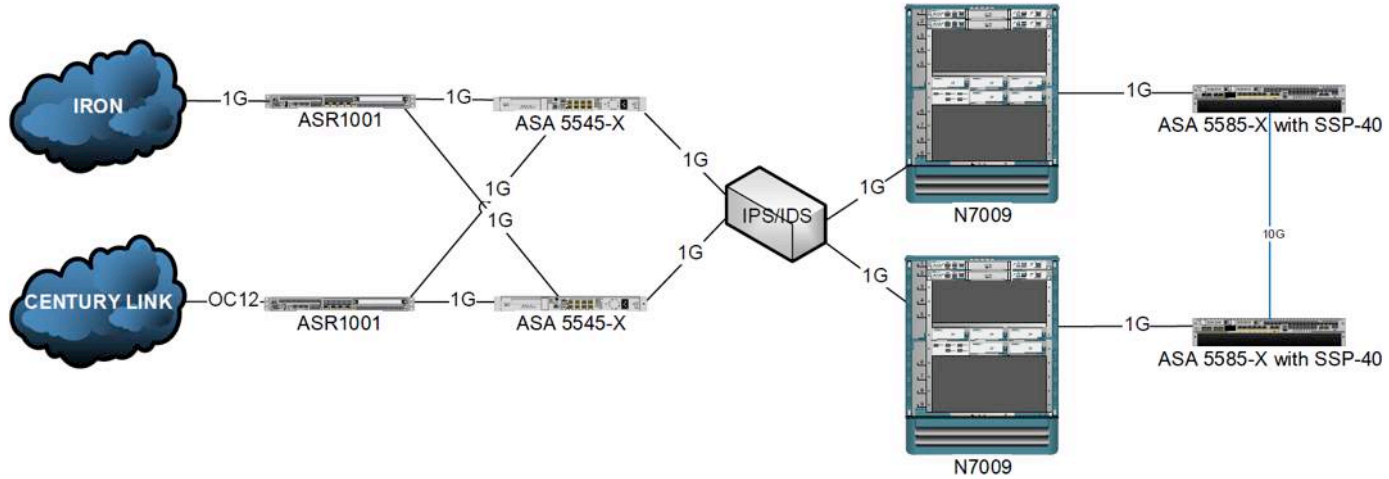
- IMNH Virtual Museum imaging
 - 11 TB of imagery data
 - Used for research, analysis, and reference



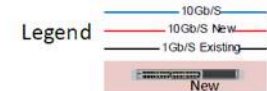
Progress Toward Challenges

- Research data center
- Science DMZ
- Network improvements

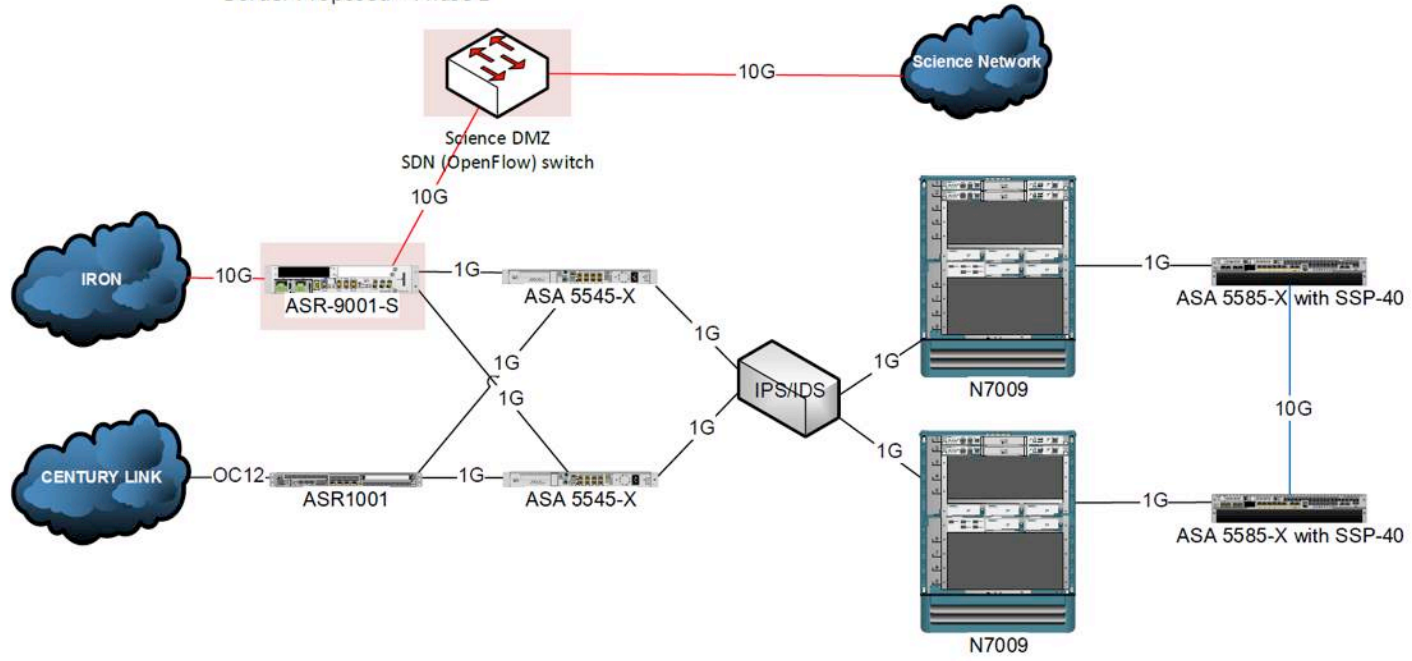
Border - Existing



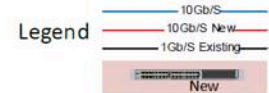
Revision.1 :5/13/2015
ISU NeTel JB



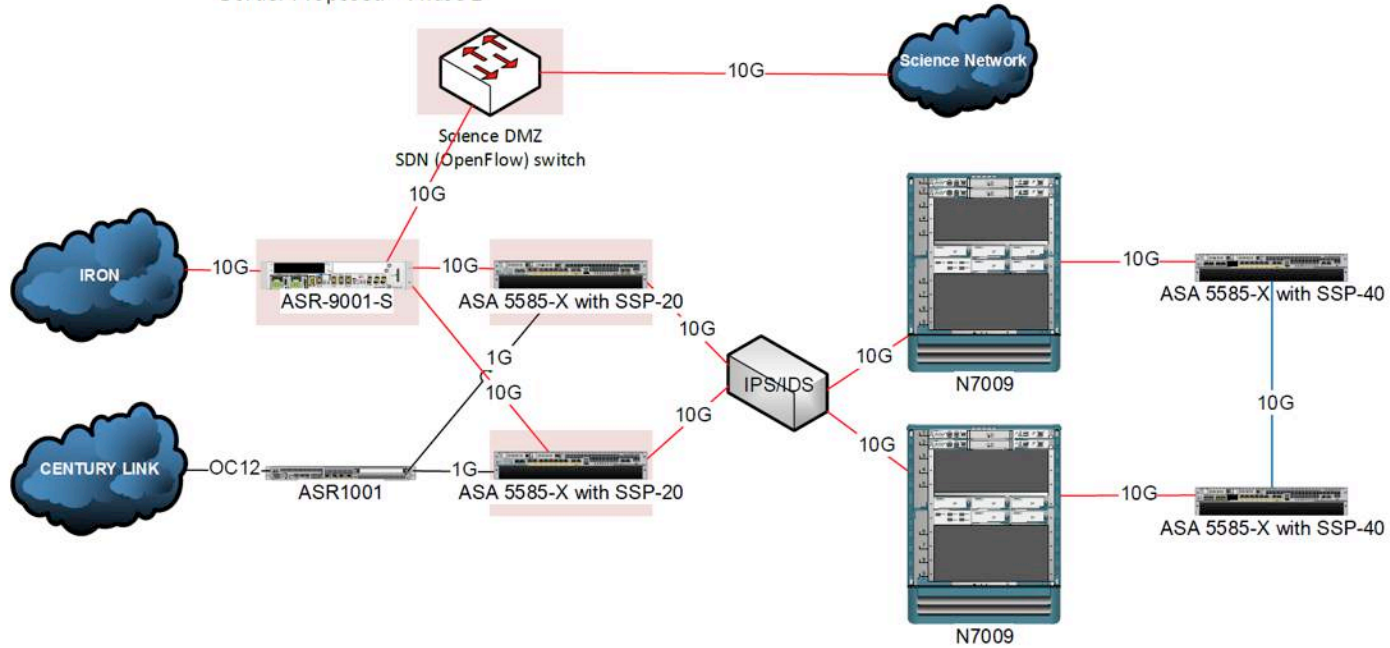
Border Proposed – Phase 1



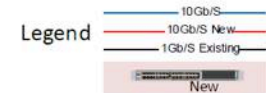
Revision.1 :5/13/2015
ISU NeTel JB



Border Proposed – Phase 2



Revision.1 :5/13/2015
ISU NeTel JB



Questions?

- Keith T Weber (webekeit@isu.edu)
- Randy Gaines (gainrand@isu.edu)
- Mark Norviel (norvmark@isu.edu)