Portable Data Management Cloud for Field Science

Yuma Matsui, Aaron Gidding, Thomas E. Levy, Falko Kuester, Thomas A. DeFanti
California Institute for Telecommunications and Information Technology (Calit2), UC San Diego
{yumatsui, agidding, tlevy, fkuester, tdefanti}@ucsd.edu

Managing Big Data in Archaeology

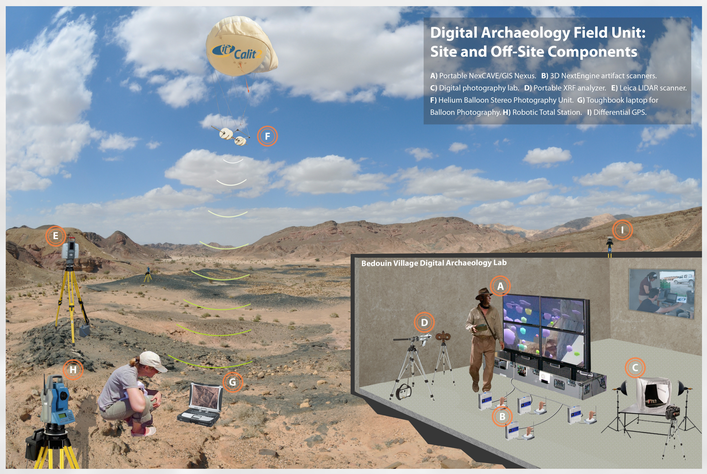
• ***A modern field science such as archaeology is heavily data-driven*** using various kinds of state-of-the-art measurement instruments. It requires sophisticated computer infrastructure to manage large amounts of heterogeneous data.

e.g.) Geospatially referenced ground and airborne images, 3D geospatial data from total stations and laser scanners (LiDAR), and various metadata for archaeological sites and artifacts (artifacts' shapes, spectrums, temporal information, inventory)

• The concept of cloud computing provides a flexible cyber infrastructure for large-scale data management. However, ***A problem unique to field research is that researchers often work at remote field sites*** with limited computer and network resources.

• For a data management system that has to work in the campus cloud and under vastly different field conditions, ***portability of computer infrastructure and common data access methods*** are essential requirements.

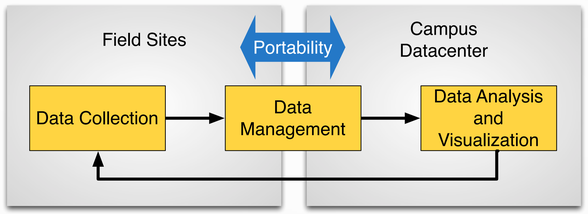
• We developed portable data management system on IaaS Cloud and used it in a recent archaeological expedition.



Digital archaeology concept

Between Cloud and Ground

• The goal of our system is to streamline data collection, data management, analysis, and visualization for cyber-archaeology over a campus datacenter and field sites.



Life cycle of field research data

• We adopted the IaaS model to take advantage of its fully controllable virtualized environment and realize portable data management infrastructure.

➜ Data and programs are easily transferable.

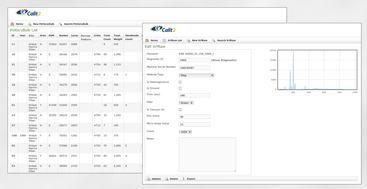
Data and Programs

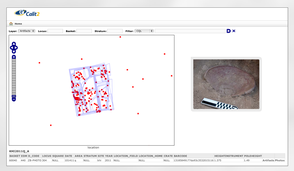
• Representation of field research data

- Raw measurement data: LiDAR, XRF, and FTIR
➜ Stored in an object storage
➜ Accessible with Amazon S3-compatible REST API

- Structured data: artifact inventory data, artifact/site metadata, and total station data
➜ Stored in a database
➜ Accessible with JSON REST API

• Web-based data management application to archive all the data



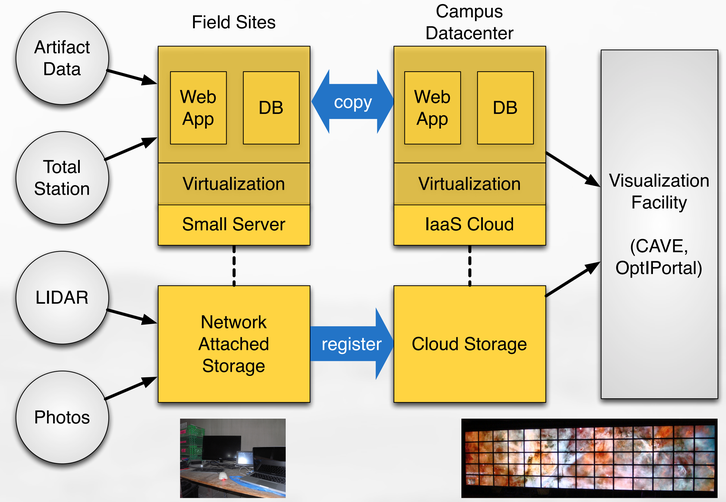


System in the Wild

• System components

- On field sites: measurement instruments and a data management server on small-scale virtualized environment

- On campus: a data management server on IaaS cloud and large-scale visualization facilities



System components of portable field data management system

• System workflow

- On field sites:
Various data are collected with measurement instruments in excavations. Structured data are put into the database through the web application. Raw file data are temporarily stored in network-attached storage.

Synchronization
(VM copy and cloud storage registration)

- On campus:
Data and programs from fields are moved to the campus cloud infrastructure. Data analyses and visualizations are executed with the collected data on high-performance computers.