Emerging Technologies CEOS/WGISS

CEOS Plenary 2014

Tromso, Norway

Tuesday, October 28th 16:00-17:00

Image Source: http://beck-technology.com/

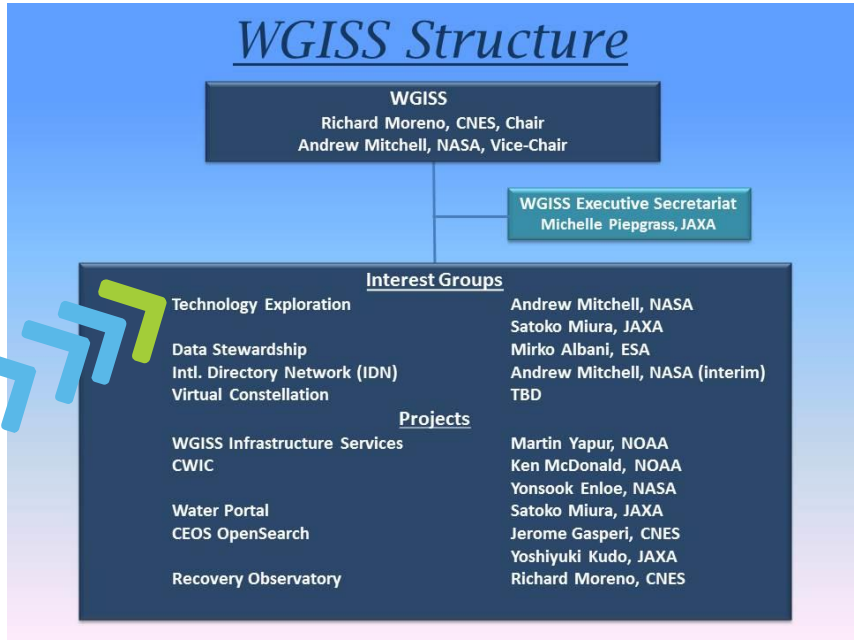
CEOS
Committee on Earth Observation Satellites

Agenda

- **WGISS Technology Exploration Interest Group Introduction**
- **Selected Emerging Technologies: Big Data and Cloud Computing**
  - **What is it?**
  - **What are practical applications for CEOS?**
- **Positive ways the WGISS Technology Exploration Interest Group can serve CEOS data systems/VCs/WGs/Ad Hoc Groups**
- **Open Forum for Discussion**

Technology Exploration Interest Group

- **WGISS created the Technology Exploration Interest Group in 2012 to serve as a forum to discuss technical information and lessons-learned experience about current and trending software technologies and services impacting CEOS agencies.**



Technology Selection

- Before each WGISS meeting, we send a survey to the WGISS members asking them to list and rank in order of priority the emerging technology topic affecting your data system.
- The most popular 3 topics are selected to hold sessions during the WGISS meeting inviting speakers from the CEOS entities and experts in industry.
- During WGISS past meetings, several agencies used this forum to present and discuss topics surrounding Big Data, Cloud Computing, Semantics, Visualization, Authentication and Product Formats.
- After every meeting, we will be releasing a report summarizing the discussions during the sessions.

Past Topics & Presentations

Big Data and Cloud Computing

- INPE – Big Earth observation data analytics for land change monitoring
- ESA – ESA Earth Observation Big Data R&D
- NASA – BEDI (Big Earth Data Initiative)
- ROSCOSMOS - Roscosmos Geoportal services
- NASU-NSAU - SSAU: JECAM & GEOGLAM activities – big data challenge
- CNES - Big Data Technologies (Hadoop)
- Geoscience Australia - Australian Geoscience Data Cube
- CNES - Quality assessment and land cover services in a collaborative clouds environment

Semantics

- CNES – Providing semantic search using RESTo
- NASA - SWEET: Semantic Web for Earth and Environmental Terminology & SESDI: Semantically Enabled Scientific Data Integration

Past Topics & Presentations

Visualization

- **HUNAGI - NASA World Wind Virtual Globe Technology and the WW Europa Challenge**
- **NASA – Global Imagery Browse System / Worldview**

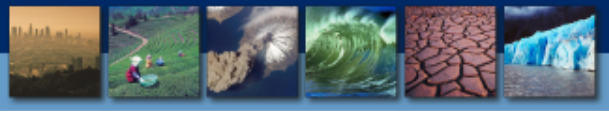
Metadata/Data Formats and Standards

- **ISO TC 211 - Revision of ISO 19115-2**

User Registration (Authentication/Authorization)

- **User Management for GEOSS**
- **ESA - ESA EO Federated User Management**
- **NASA – EOSDIS User Registration System (URS)**

What is Big Data?



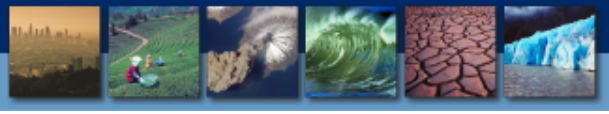
- **Wikipedia defines Big Data as "an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications."** http://en.wikipedia.org/wiki/Big_data

- **Big Data has often been looked at as a problem but it should be looked at as an opportunity.**

- **Big Data enables data systems to process, archive, manage, and manipulate large amounts of disparate data.**

- **The definition of "Big Data" varies depending on the data system's capabilities and the capabilities of the services running at the data system manage the data. What is "Big" to some, may be small to others and what is considered "Big" today will not be in the future.**

Big Data Characterization



- **Volume: How much data**
- **Velocity: Speed of data processing**
- **Variety: Various types of data**

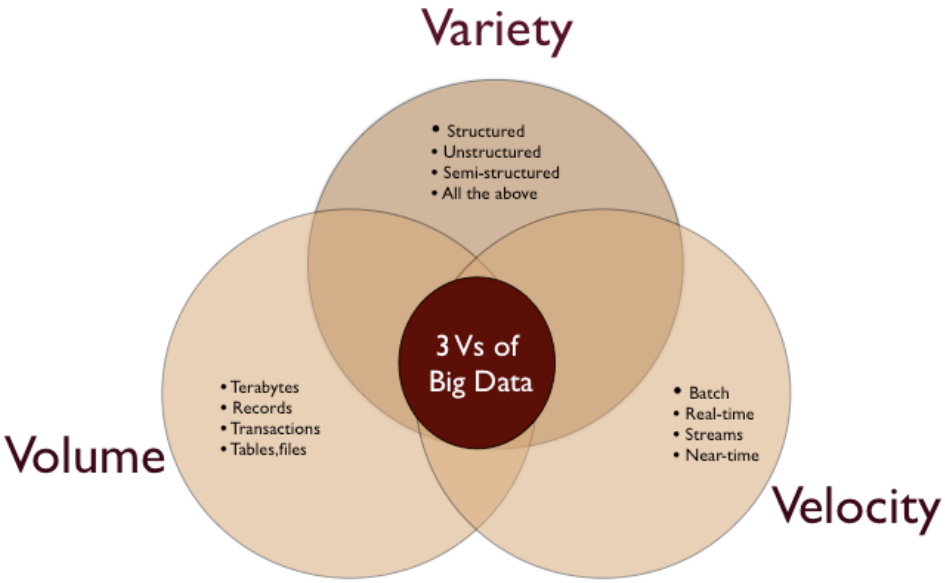


Image Source: http://jaywsalon.com/?q=jwsalon&page=2

Gartner analyst Doug Laney introduced the 3Vs concept in a 2001 MetaGroup research publication, *3D data management: Controlling data volume, variety and velocity*.

Big Data

- **Big Data is NOT a single technology. It refers to multiple technologies and initiatives that involve large complex data sets and infrastructures.**

The real issue is not that agencies are now archiving and processing large amounts of data. It's what you do with the data that counts. Big Data, when harnessed properly, allows real-time analysis and data mining to produce better science.

- **Cloud Computing enables Big Data processing for data systems by relieving a number of complex scenarios that Big Data can introduce.**

What is Cloud Computing?

Image Source: http://cantsaytoomuch.com/cloud-computing/

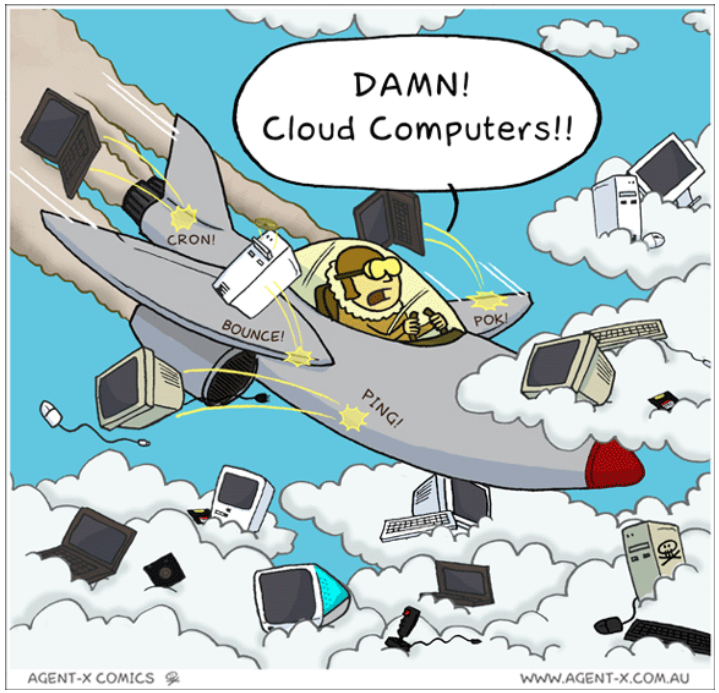


Image Source: http://quartsoft.com/blog/201303/saas-iaas-and-paas-cloud-computing

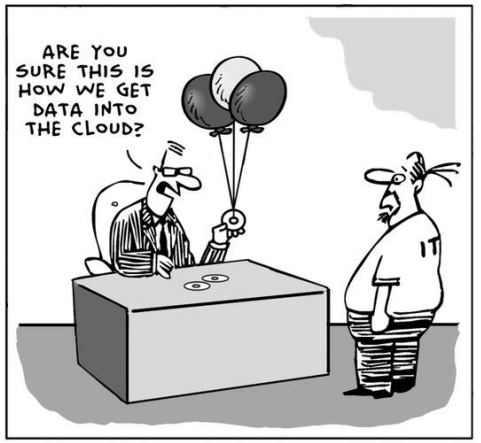


Image Source:
http://blog.inf.ed.ac.uk/sapm/2014/02/21/746/

Cloud Computing

- **Cloud Computing is an emerging information technology and computing architecture that seeks economies of scale for storage and processing based on the incremental use of computing resources**

- **Amazon Web Services defines "Cloud Computing as "on-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing.**

(http://aws.amazon.com/what-is-cloud-computing/)

amazon web services™

Cloud Computing Services – by JAVA Code Geeks

- **Infrastructure as a Service (IaaS):** This is the most basic cloud-service model, which provides the user with virtual infrastructure, for example servers and data storage space. Virtualization plays a major role in this mode, by allowing IaaS-cloud providers to supply resources on-demand extracting them from their large pools installed in data centers.

- **Platform as a Service (PaaS):** In this model, cloud providers deliver to the user development environment services where the user can develop and run in-house built applications. The services might include an operating system, a programming language execution environment, databases and web servers.

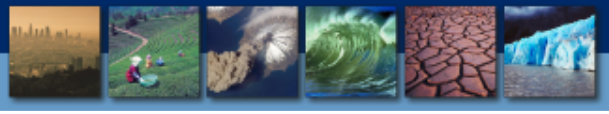
- **Software as a Service (SaaS):** In this model, the cloud provides the user with access to already developer applications that are running in the cloud. The access is achieved by cloud clients and the cloud users do not manage the infrastructure where the application resides, eliminating with this the way the need to install and run the application on the cloud user's own computers.

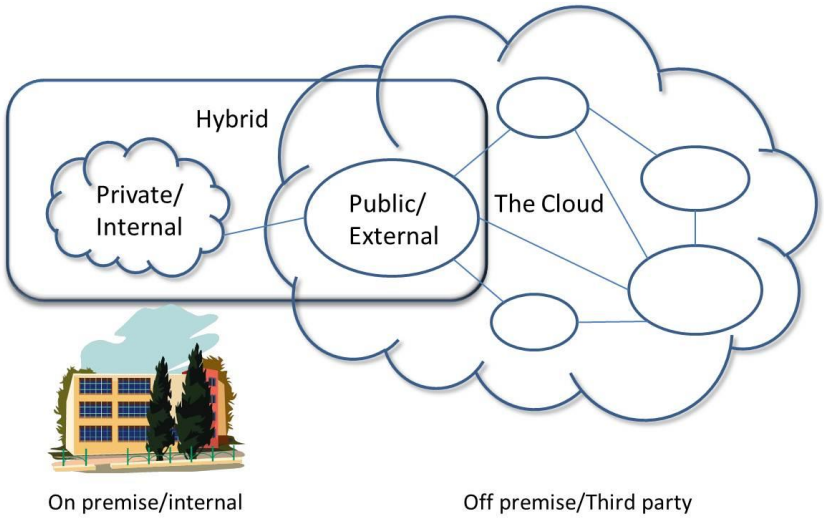
- **Network as a Service (Naas):** The least common model, where the user is provided with network connectivity services, such as VPN and bandwidth on demand.

http://www.javacodegeeks.com/2013/04/advantages-and-disadvantages-of-cloud-computing-cloud-computing-pros-and-cons.html

Cloud Computing Types

(http://www.synergygys.com/Solutions/CloudServices/)





Cloud Computing Types

- **Private clouds are typically deployed within an organization's own internal ecosystem, often leveraging the organization's own private datacenter.**
- **Public clouds are hosted by a third party datacenter located off premise at multiple locations outside of an organization's building. Public clouds are often hosted on virtualized multi-tenancy datacenters where different organizations have access to shared pooled hardware and power resources, yet can run their applications and data in secure, isolated environments.**
- **Hybrid clouds are a combination of using some services delivered via a private cloud internally and other services delivered via a public cloud externally.**

Pros and Cons of Cloud Computing
(by JAVA Code Geeks)

Pros

- Cost Efficiency
- Convenience and continuous availability
- Backup and Recovery
- Cloud is environmentally friendly
- Resiliency and Redundancy
- Scalability and Performance
- Quick deployment and ease of integration
- Increased Storage Capacity
- Device Diversity and Location Independence
- Smaller learning curve

Cons

- Cost Efficiency
- Security and privacy in the Cloud
- Dependency and vendor lock-in
- Technical Difficulties and Downtime
- Limited control and flexibility
- Increased Vulnerability

Image Source: http://blog.activestandards.com/responsive-design-the-pros-and-cons/

http://www.javacodegeeks.com/2013/04/advantages-and-disadvantages-of-cloud-computing-cloud-computing-pros-and-cons.html

Practical Applications of Cloud Computing for CEOS

- **Given the increase of data volumes and number of data records due to new missions and reprocessing cloud computing provides valuable opportunities to manage Earth Observation data.**

- **Cloud Computing can be beneficial to CEOS in the following areas:**

  - **Processing Data**
  - **Ingesting Data**
  - **Archiving Data**
  - **Distributing Data**
  - **Discovering Data**
  - **Analyzing Data**
  - **Visualizing Data**
  - **…**

Examples of how WGISS can help you…

- Open invitation to participate in the Technology Exploration IG and bring request of technology topics for the group to explore.
- The first discussion of the OpenSearch standard, which is now a CEOS Best Practice for making data accessible was held at WGISS 32 (September 2011) in the precursor to the Technology Exploration Interest Group
- Collecting lessons learned from cloud computing/big data use cases and sharing them on the Technology Exploration section of the WGISS webpage.
- Technology Exploration IG will be working with the SEO to investigate options for supporting cloud processing for GFOI and GEOGLAM. The findings will be made available for VCs, WGs and Ad Hoc Teams.
- Deployed a WGISS website to allow agencies to share code and promote their data system's applications. Looking to populate these pages with resources from other CEOS entities.

Benefits of Technology Exploration IG

- Share information about emerging technologies that can be used in EO data systems
  - Cloud computing
  - Big data
  - …
- Agencies can pick up the latest technology trend, hints for our future systems and find possible collaborations (with other agencies)
  - Validation of R&D strategies with peers
  - Lead for better decision making
  - Learning what other agencies are doing
  - Sharing experience and expertise
- WGISS develops white papers, best practices and lessons-learned documents
  - CEOS OpenSearch Best Practice
  - CEOS Interoperability Handbook
  - Long Term Data Preservation guidelines
  - …
- WGISS develop systems that can be operational
  - IDN for data discovery
  - CWIC / FedEO for data access
  - Recovery Observatory for post disasters management

In Summary…

- **We would like to invite all WG, VC and Ad Hoc Group members who have an interest or are being impacted by emerging technologies to participate in the WGISS Technology Exploration Interest Group.**
- **Sign up for the mailing list today!**

Due to the multi-lingual/multi-national nature of CEOS, at the last WGISS meeting it was widely agreed that WGISS should develop a technology glossary that represents information technology definitions used in current projects and interest groups to ensure all entities are 'speaking the same language'.

Image Source: http://ipkitten.blogspot.com/2013/07/the-patent-legacy-of-computer-mouse.html

Discussion Time

- **What is your biggest challenge(s) you face today?**
- **How should WGISS promote emerging technologies among CEOS groups?**
- ?

Image Source: https://salemcc.instructure.com/courses/1/pages/difference-between-discussion-and-announcement

Point of Contact:
Richard Moreno (WGISS Chair)
Richard.Moreno@cnes.fr

Point of Contact:
Andy Mitchell (WGISS Vice Chair)
Andrew.E.Mitchell@nasa.gov

References

- http://aws.amazon.com/what-is-cloud-computing/
- http://en.wikipedia.org/wiki/Big_data
- http://jaywsalon.com/?q=jwsalon&page=2
- http://tdwi.org/articles/2012/07/24/big-data-4th-v.aspx
- http://www.javacodegeeks.com/2013/04/advantages-and-disadvantages-of-cloud-computing-cloud-computing-pros-and-cons.html
- http://www.mongodb.com/big-data-explained
- http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
- http://www.synergygs.com/Solutions/CloudServices/