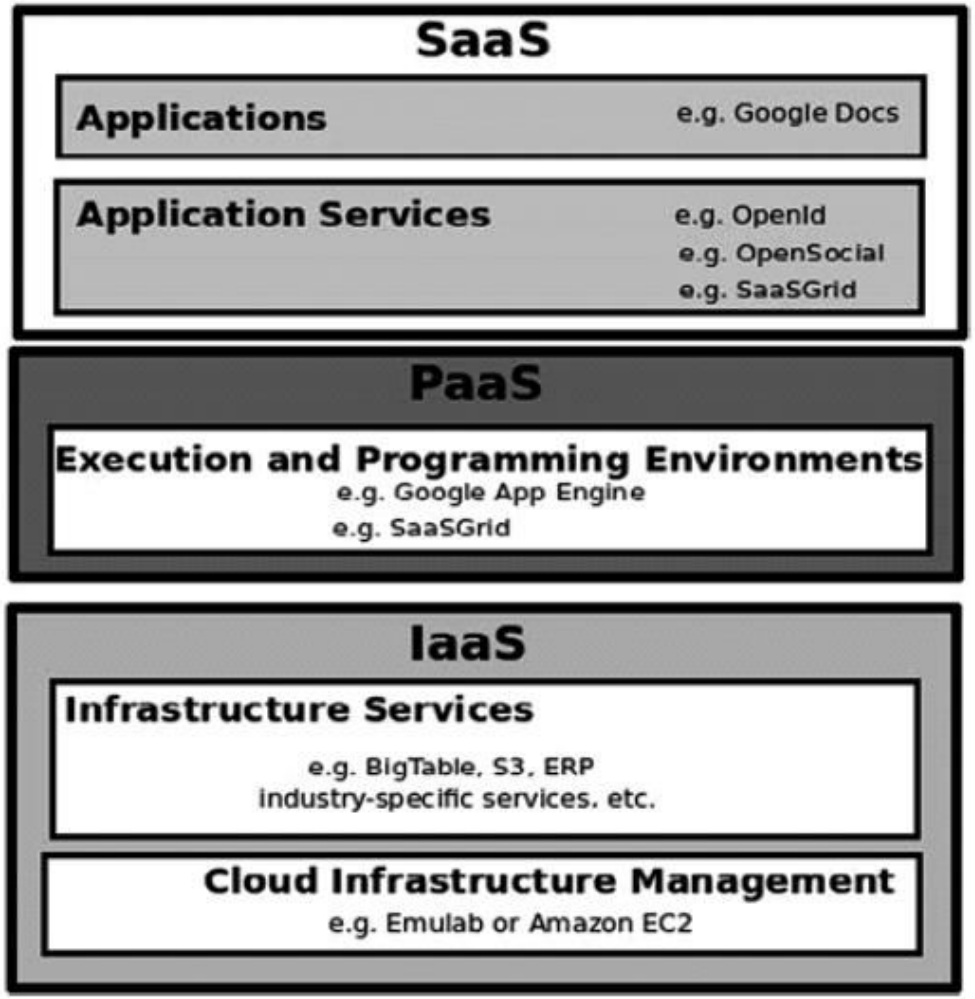
Cloud Services for Projects in Bioinformatics: Technical Considerations and Business

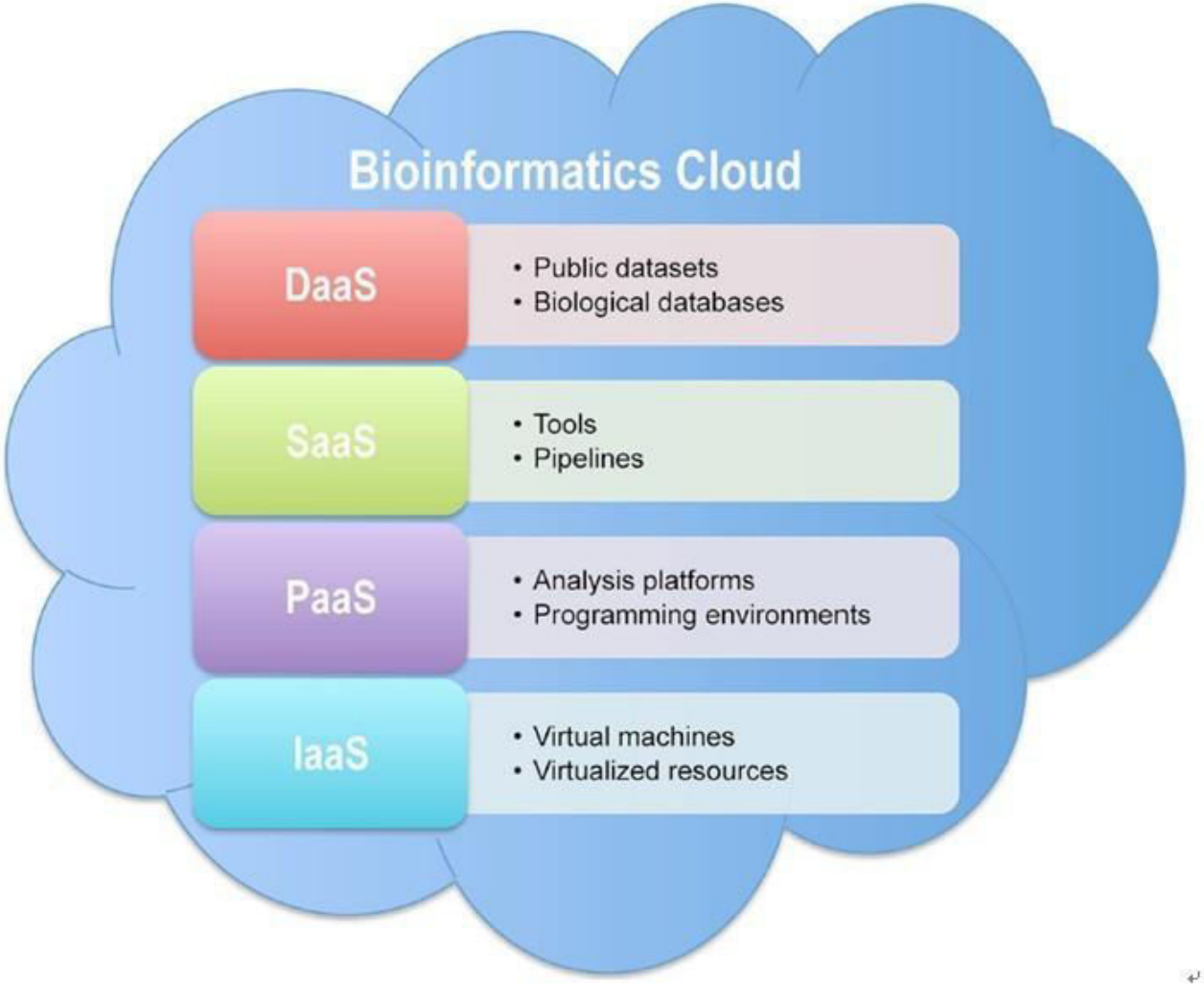
Fernando Barraza
Omicsco
Universidad de San Buenaventura Cali

Agenda

- What is a Cloud Computing?
- Main Architecture of the Cloud and the Bioinformatic
- Business Aspects
- A Bioinformatic Platform on the Cloud
- Concerns and Opportunities

Main Architecture of the Cloud



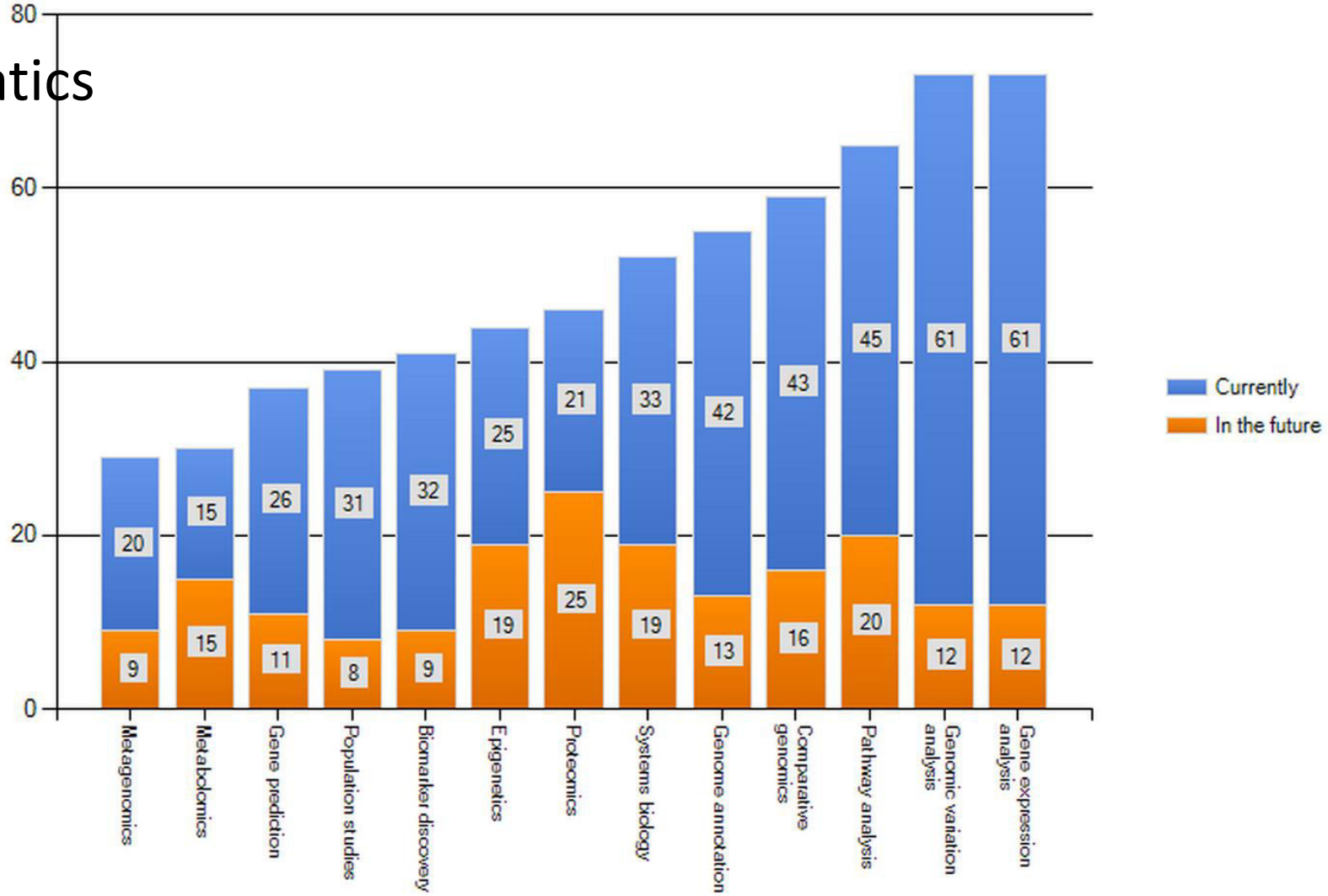


Imaged by Prof. ZHANG Zhang

Business Aspects

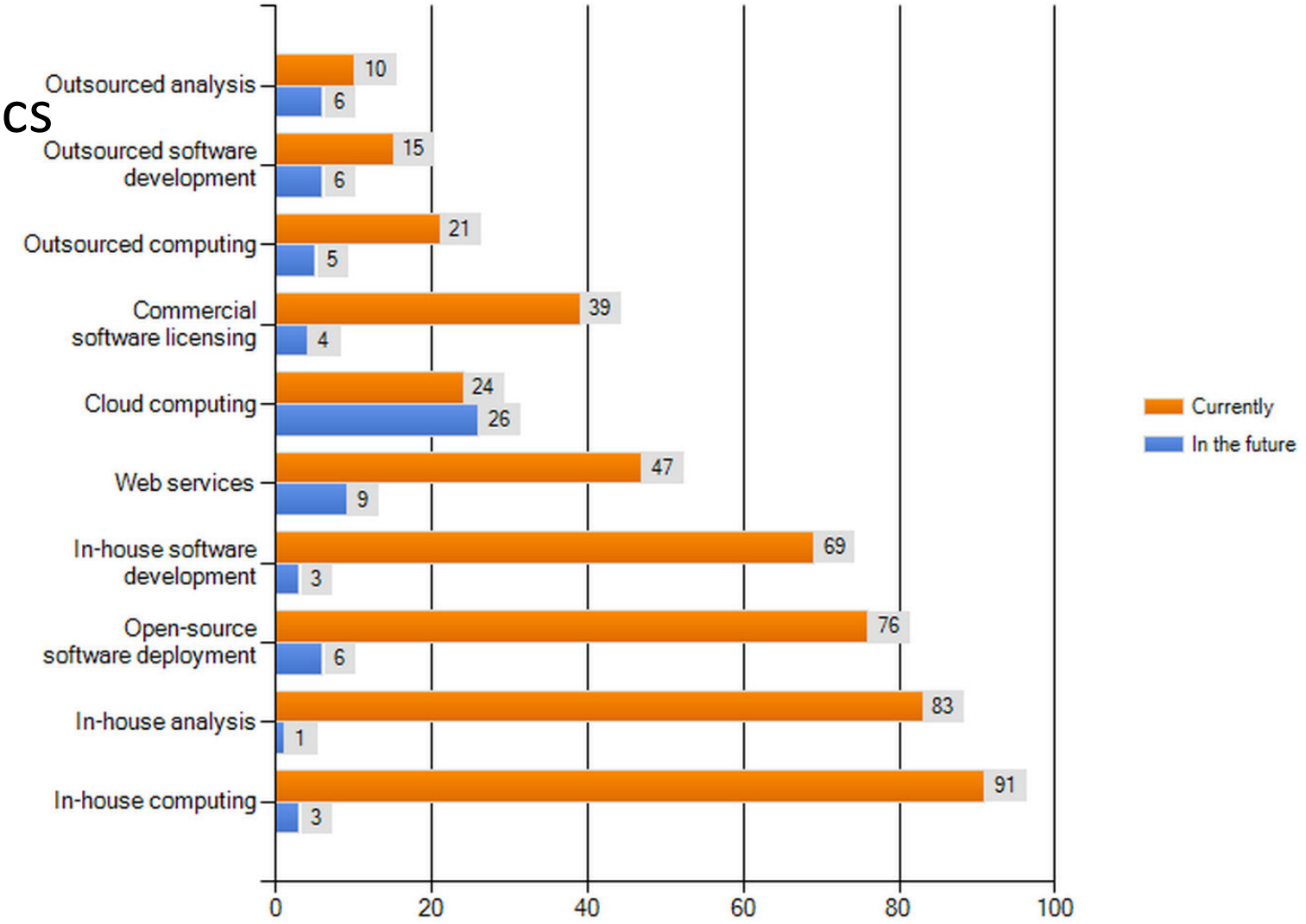
- Where is the Market?
- What are user needs?
- What is my business model?
- Who are the players?

Bioinformatics Studies



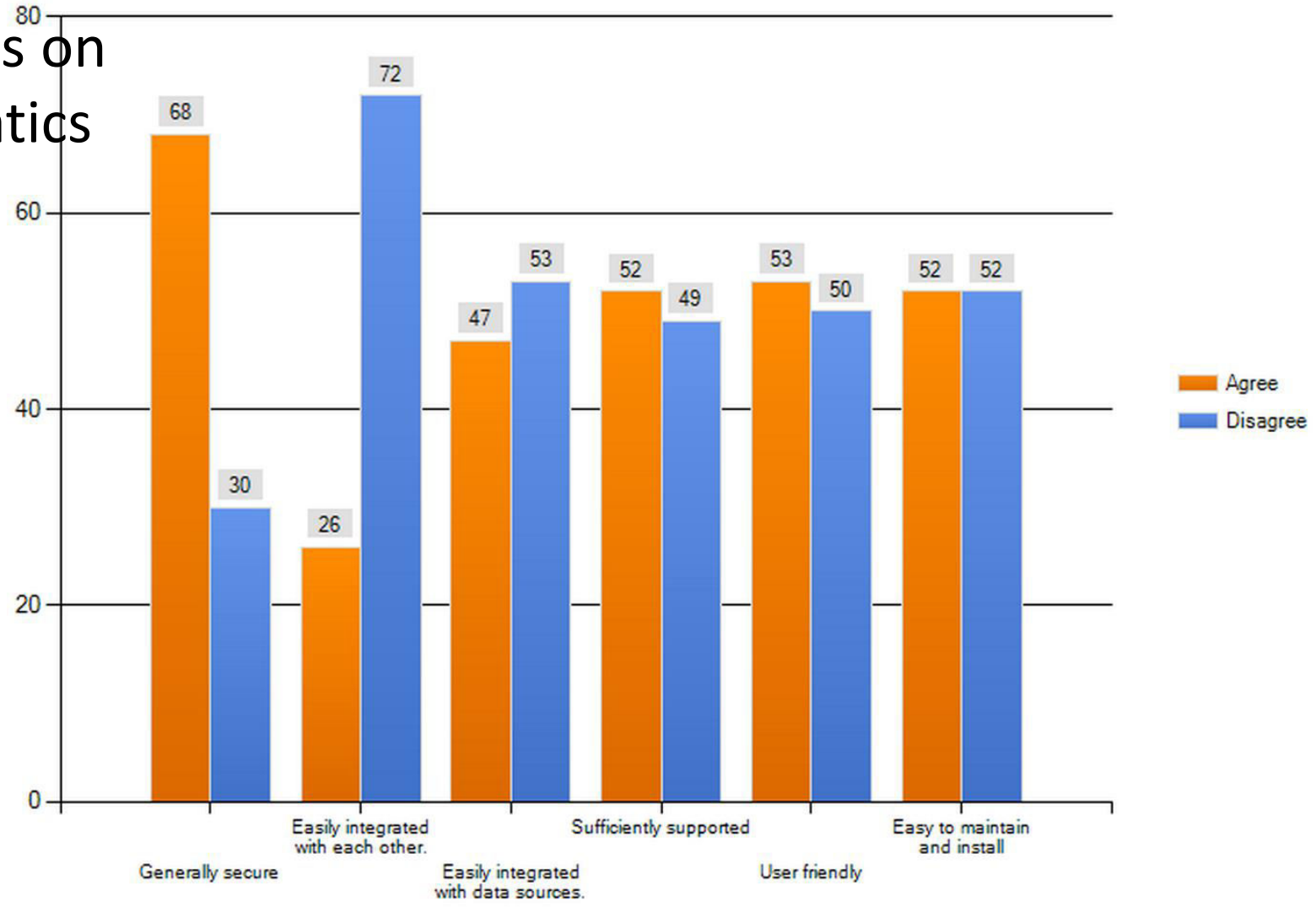
Source: Eagle Genomics

How are Bioinformatics Studies Delivered?



Source: Eagle Genomics

Perceptions on bioinformatics Tools



Source: Eagle Genomics

Cloud Requires Architectural Shift

single instance – multi tenancy

Product vs. Service

	Software as a Product	Software as a Service
Delivery	Installed	Hosted
Development	Longer cycle, "big bang"	Short, continuous cycle
Pricing	Perpetual license + maintenance	Subscription (all inclusive)
Allocation	Capitalized	Expensed
Additional Costs	Installation, maintenance, customization, & upgrades	Configuration
Platform	Multi-version	Single Platform
Updates	Larger, less-frequent	Shorter, frequent
Sales Focus	Close the deal	Prove value in first 90 days
Feedback Cycle	Long	Short
Profits	Initial sale	Ongoing
Success	New license revenue	Lack of churn

Who are the PaaS players

Google App Engine

CLOUD FOUNDRY

Azure

Bitnami Cloud

vmforce
The world's first enterprise Java cloud

CloudBees

CumuLogic

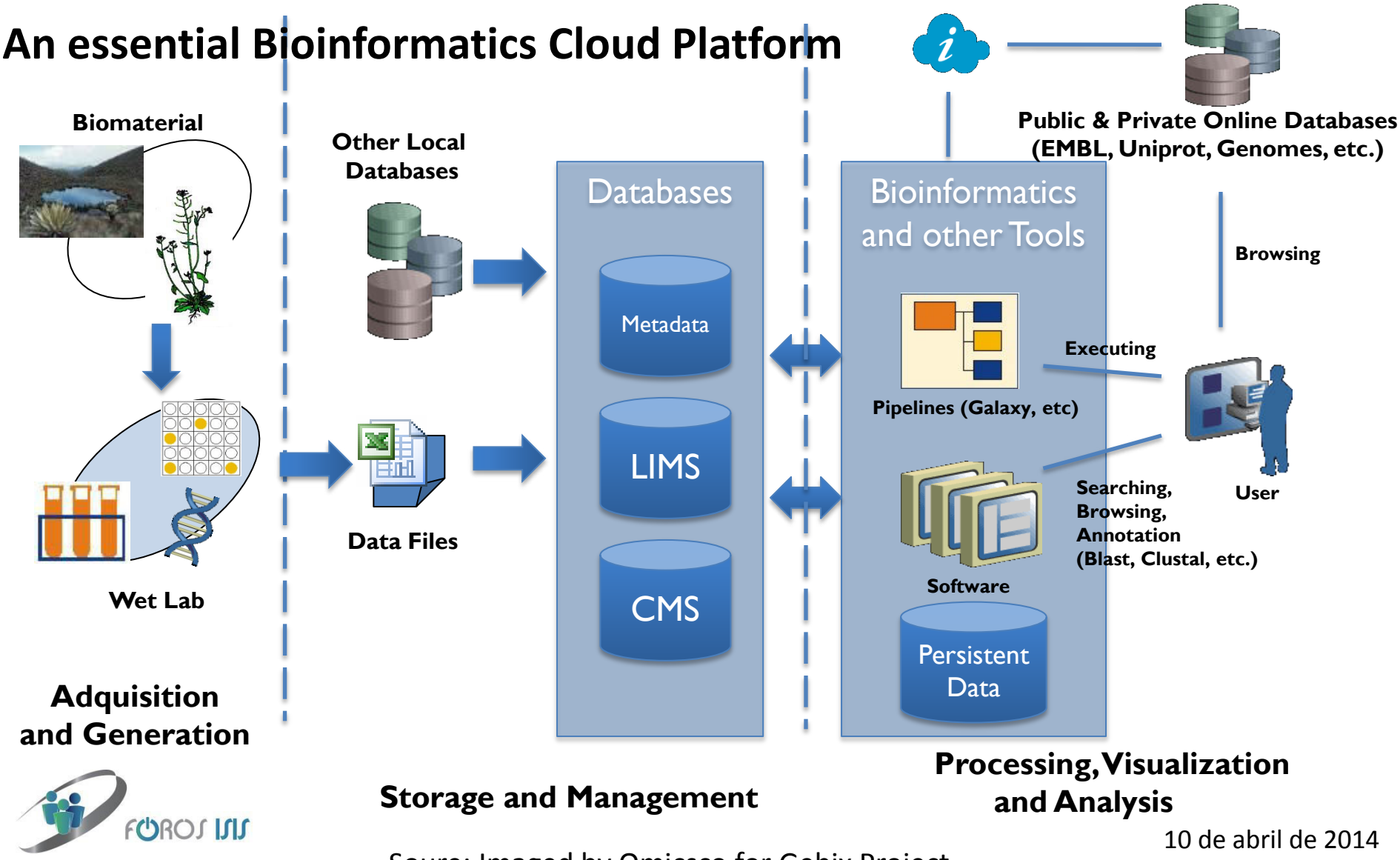
amazon web services

Amazon Elastic Beanstalk

FOROS ISIS

Slide by Srini Kumar, VP MSat

An essential Bioinformatics Cloud Platform

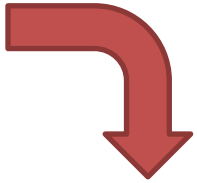


10 de abril de 2014

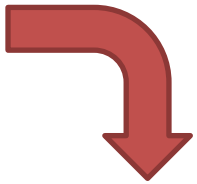
Soure: Imaged by Omicsco for Gebix Project

Bioinformatics in Cloud - Maturity Model

- Started with small and scripted applications experimenting with Cloud Services
- Copy-Paste and Data Files manipulation
- Use of predefined Tools and Pipelines



- Move to a Hybrid model, where the Cloud Services will integrate with Data Sets in-house
- Linked with Public Biological Databases
- Enhanced Pipelines



- Streamlining of pipelines in a Multi-Instance Environment.
- Use of biocomputing 'appliances'
- Persistent and Large Ubiquitous Storage

Expected features on a bioinformatics appliances

- Availability of scientific validation of the tools through publication/peer review
- Computational efficiency and scalability
- Ease of maintenance and installation
- Ease of data manipulation/visualization (Do not discard command-line tool !!)
- Plug & Play for Workflow Platforms (APIs and Standard)

Concerns and Opportunities

- Lack of Control, Security and Data Privacy
- Service Level Agreements & Standards Compliance
- New application model costs/adoption
- Data Manipulation still is a pain !! (Web Semantics technologies ?)
- Needs for new algorithms to improve performance (big data ?)

Questions ?