

# Knowledge Discovery Infrastructure (KDI) – Architecture for Knowledge Discovery from National Healthcare (Medicare, Medicaid) Data

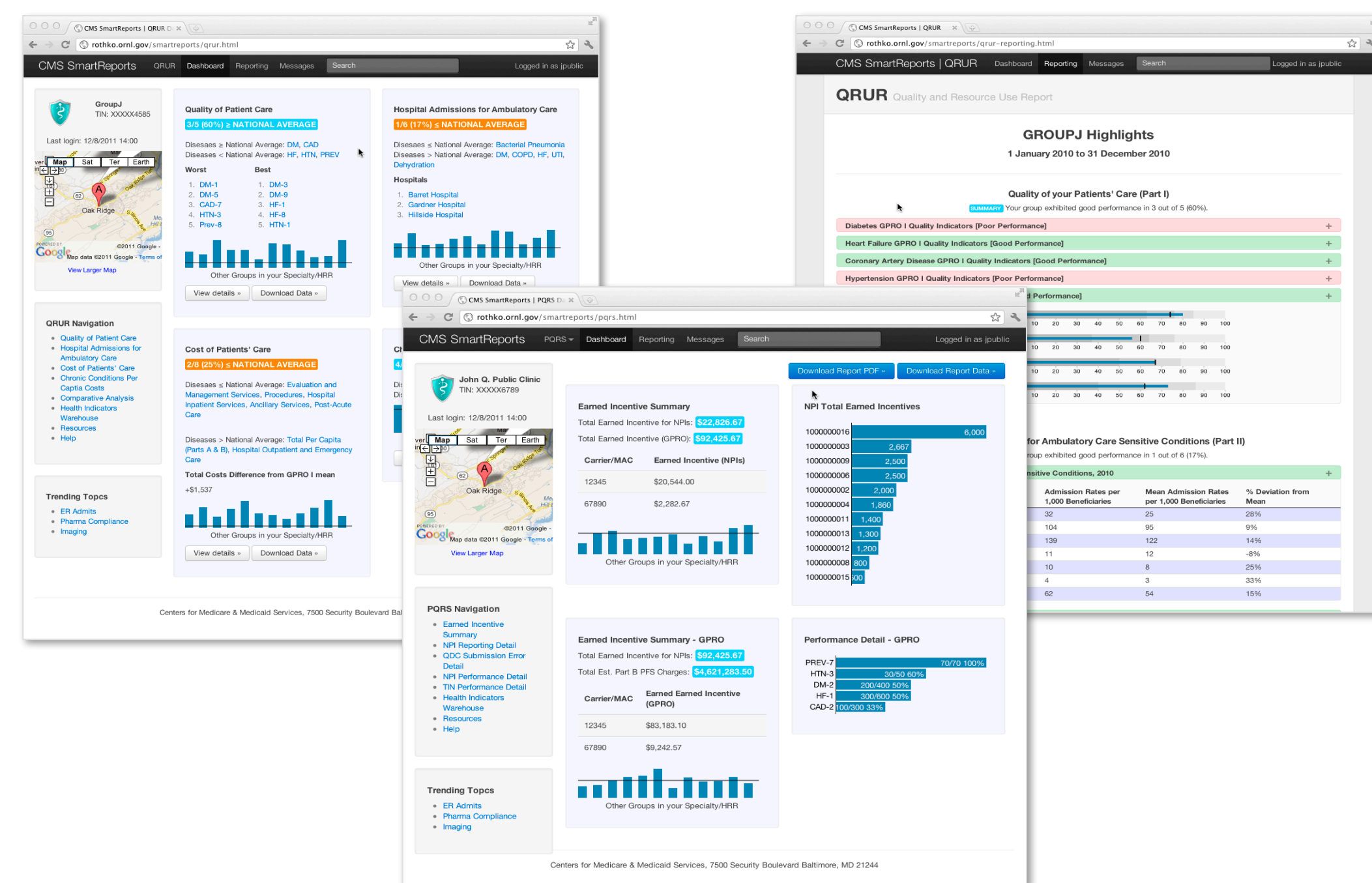
Edmon Begoli, [begolie@ornl.gov](mailto:begolie@ornl.gov), 865.241.1923, Computational Sciences and Engineering Division  
Oak Ridge National Laboratory

## National Healthcare Priorities

- Improve care
- Improve health of the population and healthcare outcomes
- Lowering of the costs of healthcare

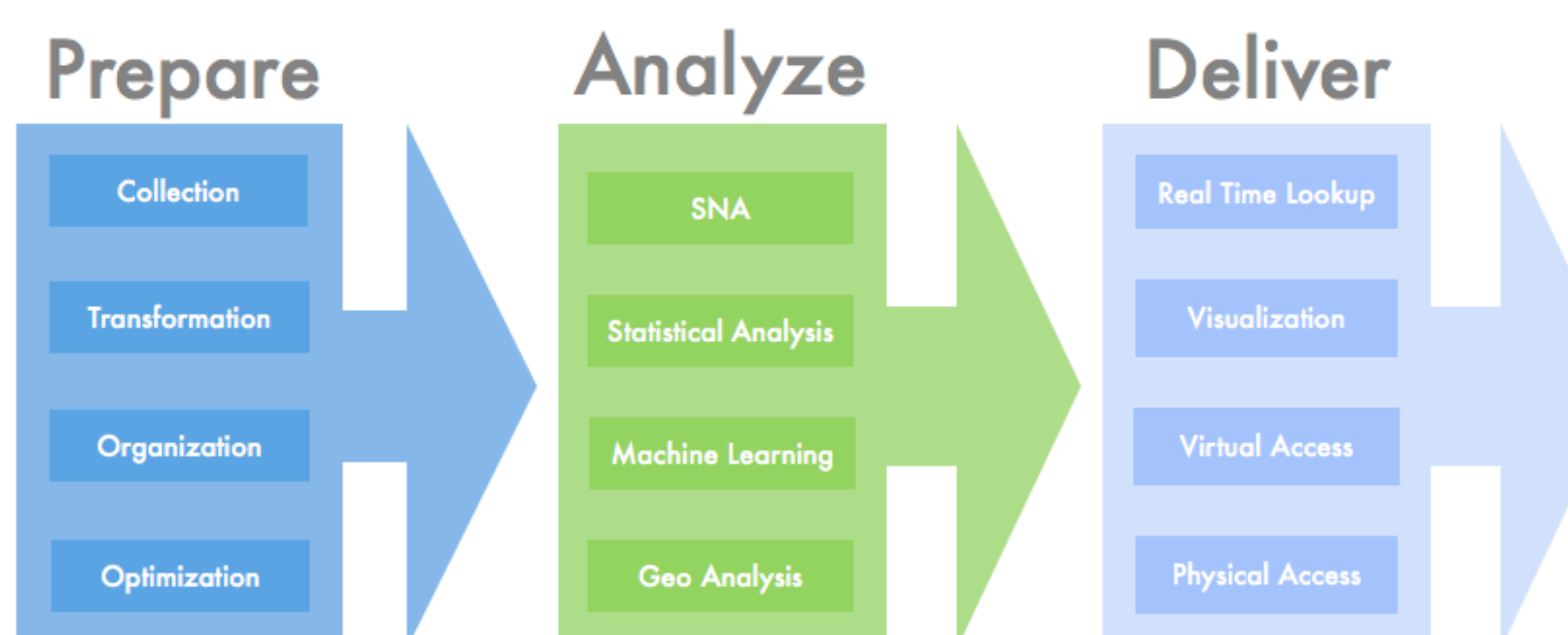
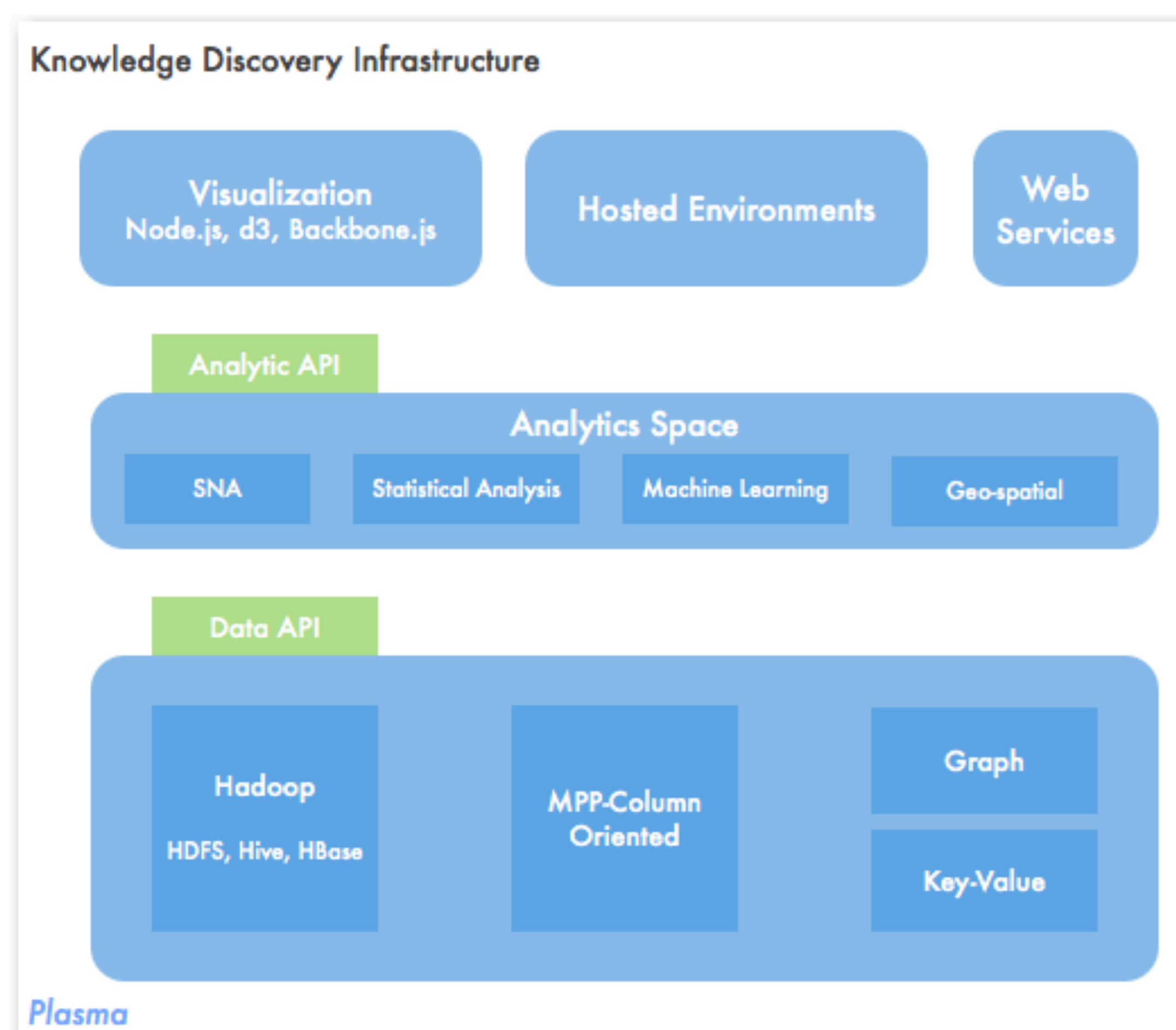
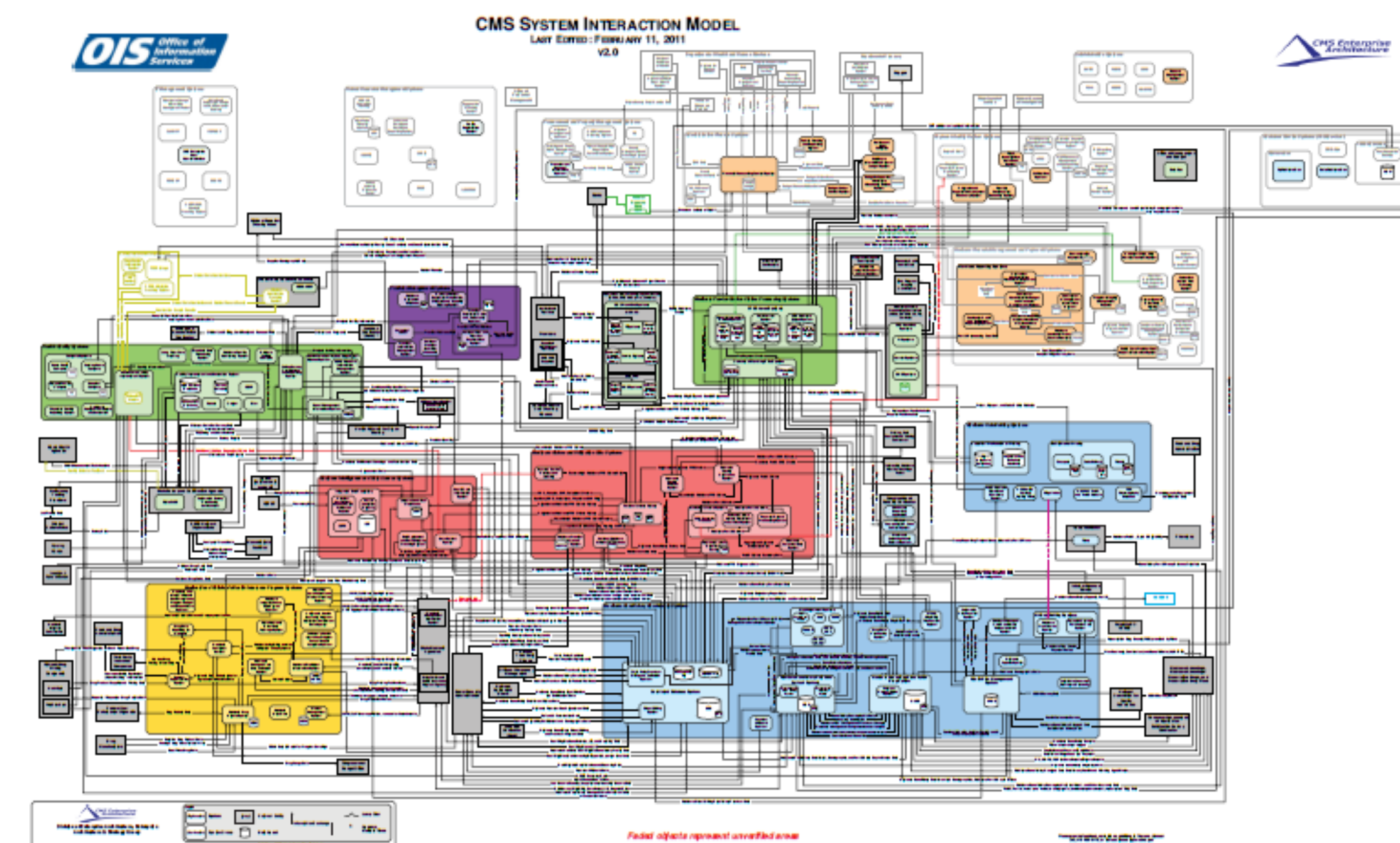
## Architectural Principles

- Flexibility of the data platform and representation
- Comprehensiveness of the analysis
- Universality of the technical implementation (open source, open standards)



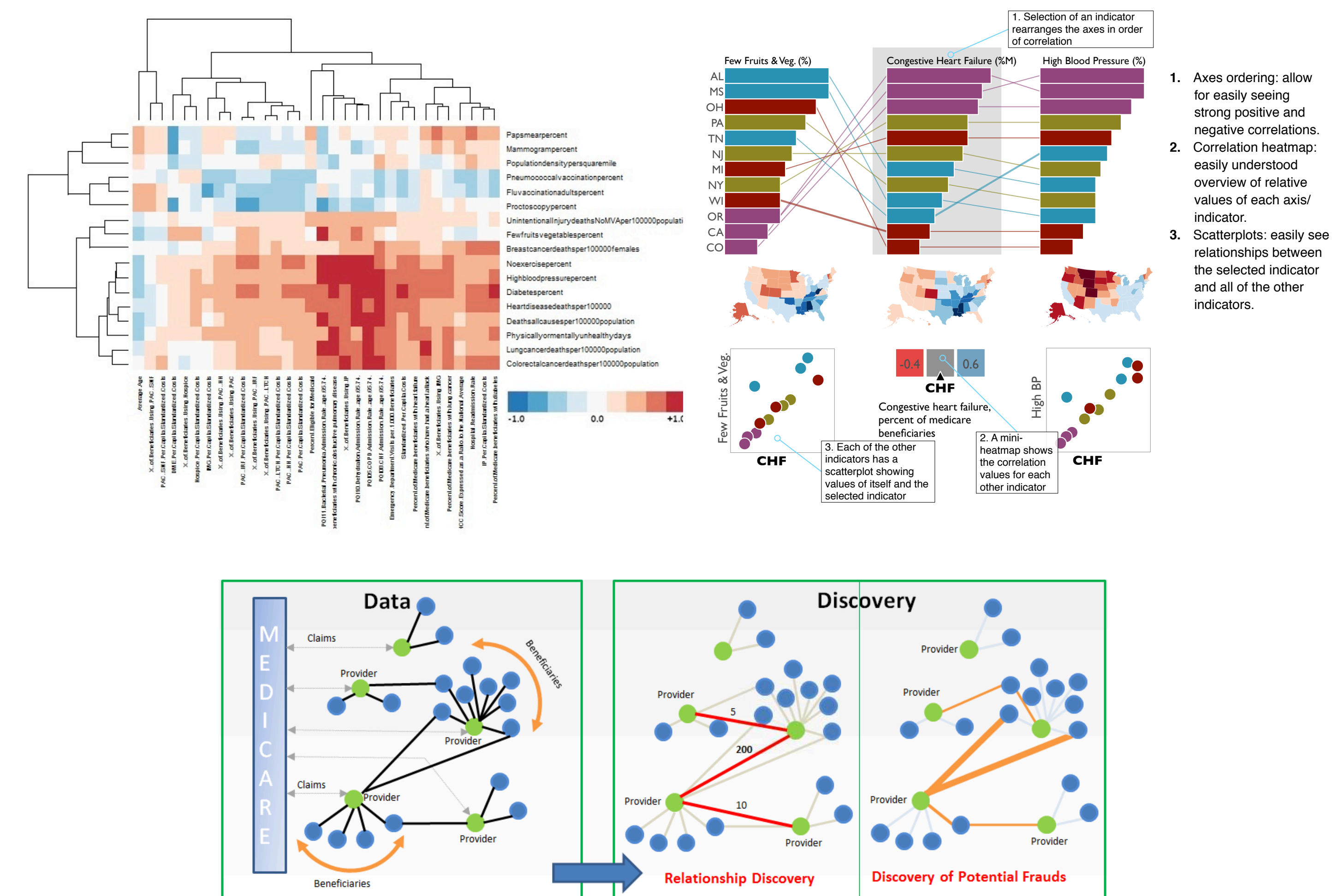
## Scope and Approach

- Use driven organization of data
- All data in one place
- Uniformity of the access (RESTful APIs)
- Analysis-as-a-Service
- Lightweight, cross-platform visualization
- Complete virtualization of compute and data resources



## Knowledge Discovery Problems Addressed

- Uniform and straightforward access to all of the data sets by internal and external consumers of data
- Better understanding and organization of data
- Sharing of analytic tools and methods across the national healthcare enterprise
- Data and analysis-driven decision making, program integrity, policy analysis, feedbacks and benchmarking of care



## Architecture Components and Frameworks

- *Cosmos* - Semantic representation and linked data
- *Aurora* – Visualization framework
- Analysis Space – SNA, SA, Predictive Analysis, ML, Geo
- *Plasma*– flexible, large scale, adaptive and elastic platform (Hadoop, MPP, KV, Graph, REST, Cloud)
- Cloud-get - automated resource and data provisioning in the cloud (Cloudstack, Python, REST)