SANAL ARAŞTIRMA ORTAMLARI
ve
AÇIK VERİLER

Bülent Karasözen, ODTÜ

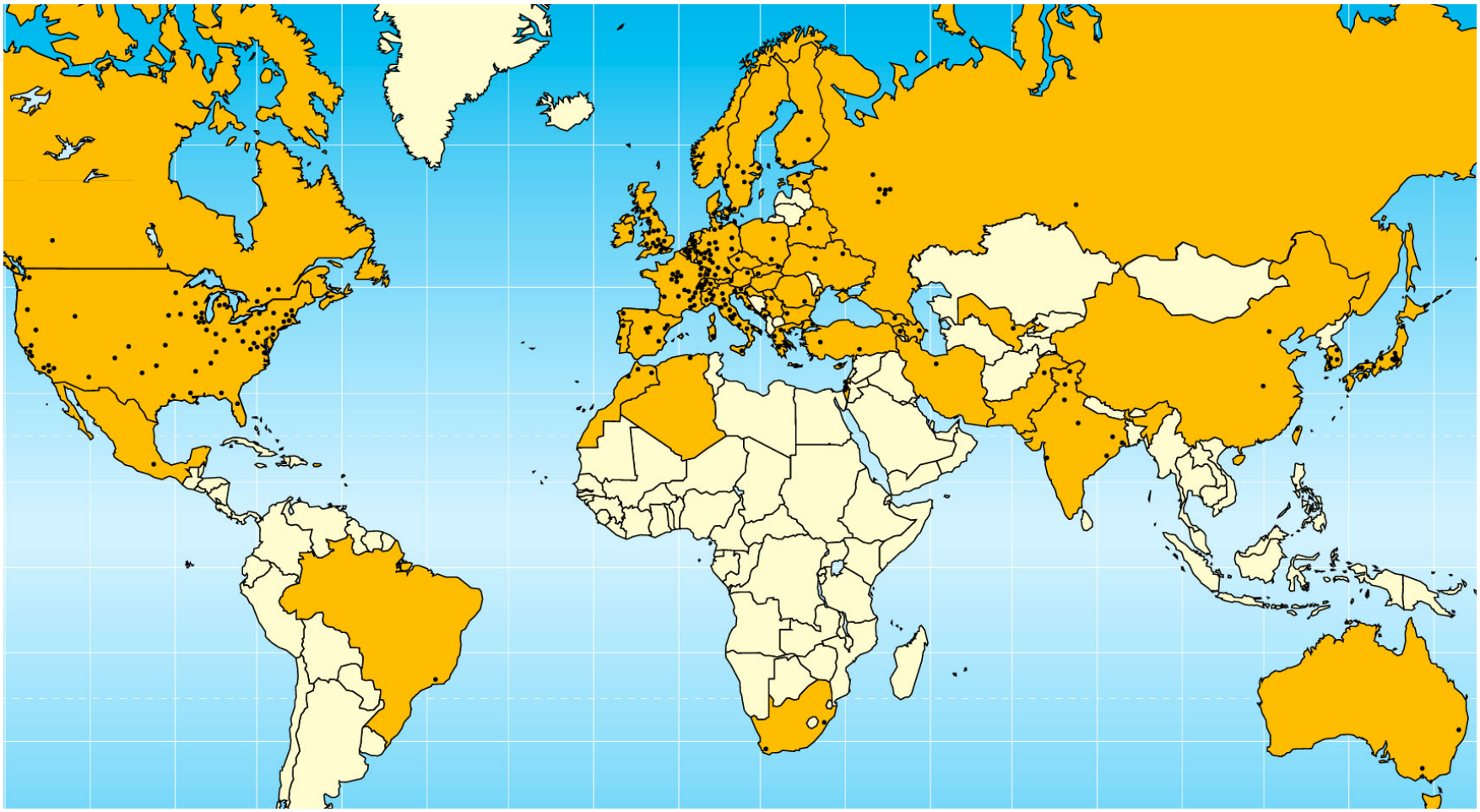
INER-TR 05

9-11 Aralık 2005, Bahçeşehir Üniversitesi

E-bilim

- Bilim ve teknolojide yeni problemlerin çözümü için giderek artan küresel işbirliği ve kaynakların ortak kullanımı
- 1990, Tim Barnes Lee , CERN, World Wide Web
- 1990, J.C.R. Licklider, ARPANET

CERN kullanıcıları



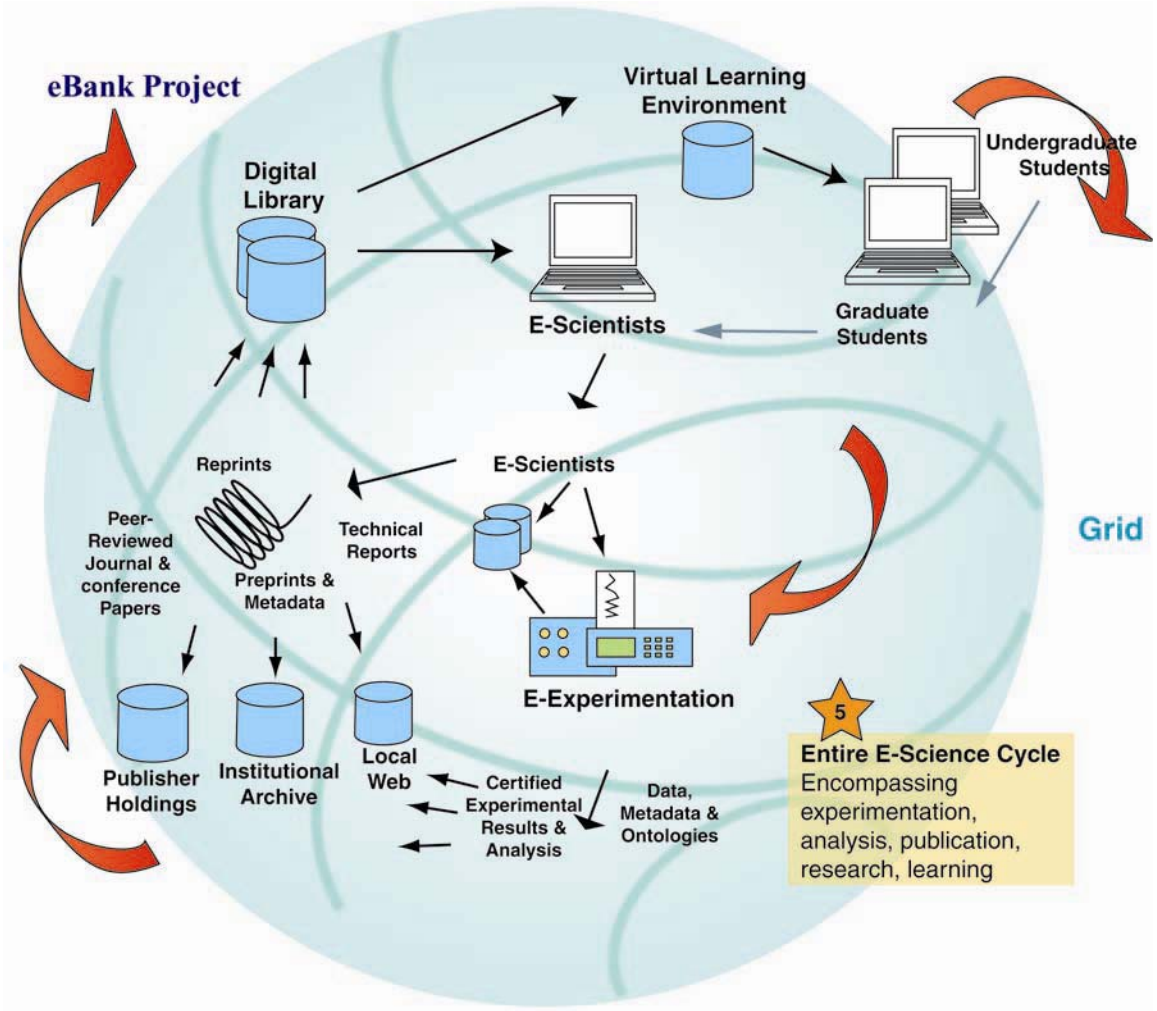
Avrupa: 267 enstitü, 4603 kullanıcı
Diğer ülkeler: 208 enstitü, 1632 kullanıcı

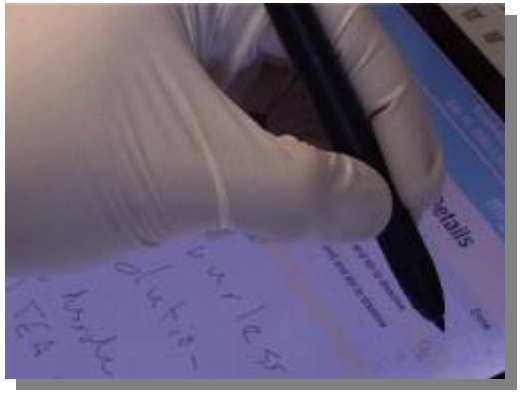
E-bilim projelerinden bazı örnekler

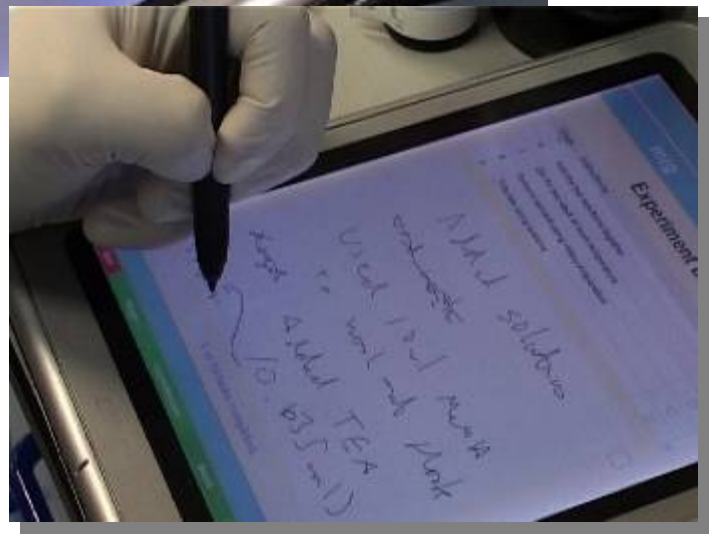
- **Parçaçık fiziği**
  - **Deneylerde elde edilen verilerin küresel düzeyde paylaşımı ve simulasyonlarda kullanımı**
- **Astronomi**
  - **Teleskoplardan elde edilen verilere dayalı ‘sanal gözlem evleri‘ oluşturulması**
- **Kimya**
  - **Deney aletlerinin uzaktan kontrolü ve elektronik labrotuvar kitapçığı**
- **Bioinformatik**
  - **Verilerin bütünleştirilmesi, veri akış planları, veriye dayalı bilgi üğretimi**
- **Sağlık**
  - **normalize edilmiş mammogramların paylaşımı**
- **Çevre**
  - **İklim modellemesi**

E-Bilimi itekliyen ana etkenler

- Büyük çaptaki sayıları az pahalı labratuvar, super bilgisayar ve veri depolarına erişim:
  - CERN LHC
- Açık kaynaklı, kaliteli ‘grid middlewaere’nın oluşturulması:
- - OMII, NMI, C-Omega
- Araştırmacıların karşısına çıkan veri seli:
  - Parçacık fiziği, astronomi, bioinformatik
- Açık erişim hareketi:
  - Bilimsel yayınlara ve verilere açık erişim

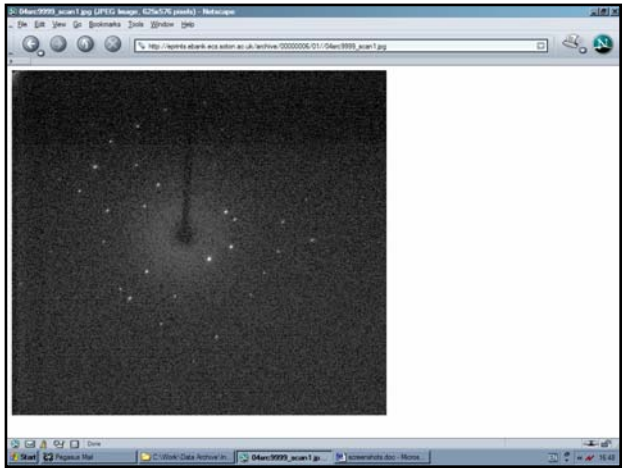

eBank Project
Virtual Learning Environment
Undergraduate Students
Digital Library
E-Scientists
Graduate Students
E-Scientists
Reprints
Peer-Reviewed Journal & conference Papers
Technical Reports
Preprints & Metadata
Grid
E-Experimentation
Publisher Holdings
Institutional Archive
Local Web
Certified Experimental Results & Analysis
Data, Metadata & Ontologies
5
Entire E-Science Cycle
Encompassing experimentation, analysis, publication, research, learning



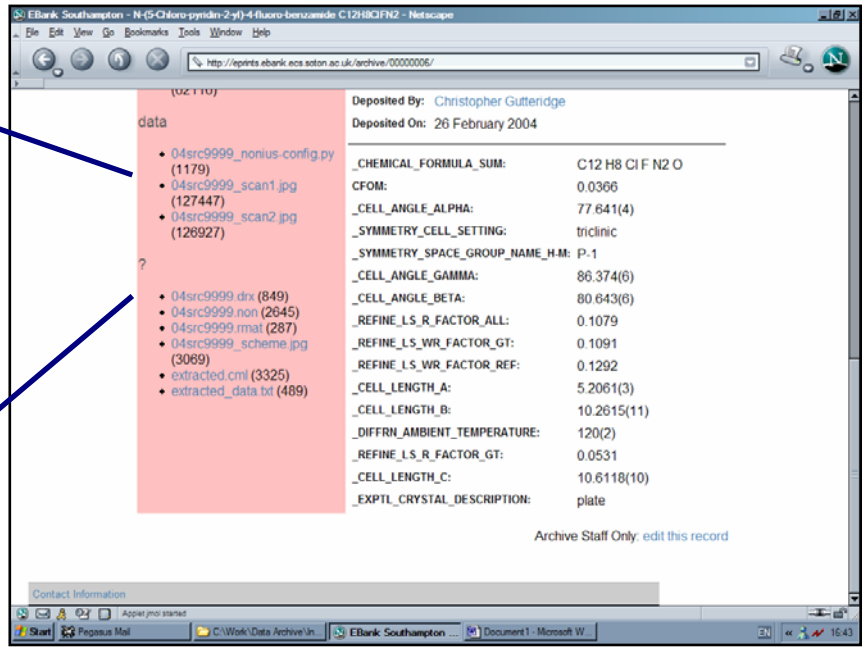


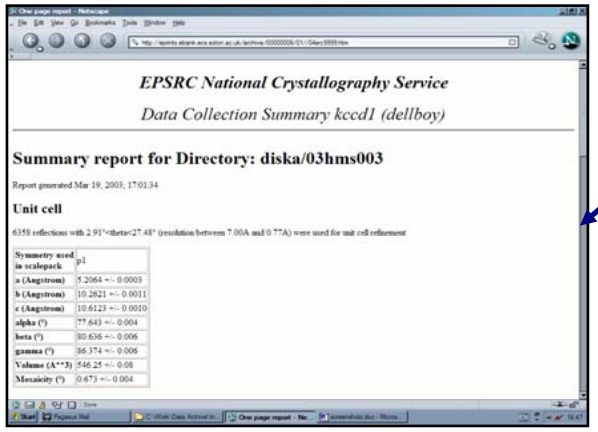
digital lab book

Crystallographic e-Prints



- **Direct Access to Raw Data from scientific papers**





Raw data sets can be very large and these are stored at National Datastore using SRB server

NSF ‘Atkins’ siber altyapı raporu:

- **Bir çok bilim dalında en son buluşlara ve yeniliklere erişim WEB üzerinden gerçekleşmekte**
- **Yüzlerce ve binlerce terabyte bilimsel verinin arşivlenmesi ve erişime sunulması, bilim ve teknolojinin ilerlemesi için artık vazgeçilemez bir gereksinim**

MIT’nin DSpace vizyonu:

- **Araştırmacılar tarafından üretilen makaleler, raporlar, veri setleri, deney sonuçları genellikle kişisel veya bölüm WEB sayfalarında saklanmakta. Araştırmacıların ayrılması veya bölümlerin zaman için de değişime uğramasıyla kaybolabilmekte**

Veri Seli

- **Büyük bir roman: 1 Mbyte**
- **İncil: 5 Mbytes**
- **Bir Mozart semfonisi (sıkıştırılmış): 10 Mbytes**
- **OED on CD: 500 Mbytes**
- **Dijital film (sıkıştırılmış): 10 Gbytes**
- **Hakemli bir derginin bir yıllığı (~20k dergi; ~2M makale): 1 Tbyte**
- **Library of Congress: 20 Tbytes**
- **Internet arşivi (10 milyon sayfa) (1996 – 2002 arası): 100 Tbytes**
- **Yıllık basılı, film, optik ve manyetik medya üretimi: 1500 Pbytes**

Basılı kaynaklar, toplam saklanan bilginin sadece 0.003% 'ünü oluşturmakta

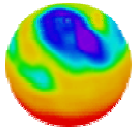
Önümüzdeki beş yılda e-bilim projelerinden, insanlık tarihindeki tüm verilerin toplamından kat, kat fazla petabyte büyüklüklerinde veri üretilecek

Grid, middleware

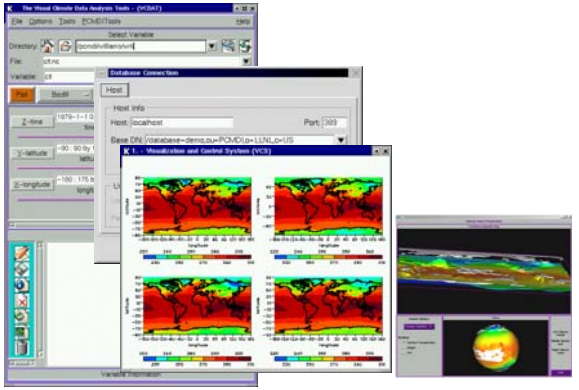
- Grid: birbirinden farklı çeşitli bilgisayarların ortak kullanımının sağlandığı ortam
- Büyük çapta çok disiplinli, araştırmacıların farklı yerlerde olduğu ortak projelerin statik WEB sayfaları aracılığıyla gerçekleştirilmesi mümkün değil
- Middleware: bilgisayar ağıyla, uygulamalar arasında iletişimi sağlayan yazılım

Complexity + Volume + Remote Access = Grid Challenge

British Atmospheric Data Centre

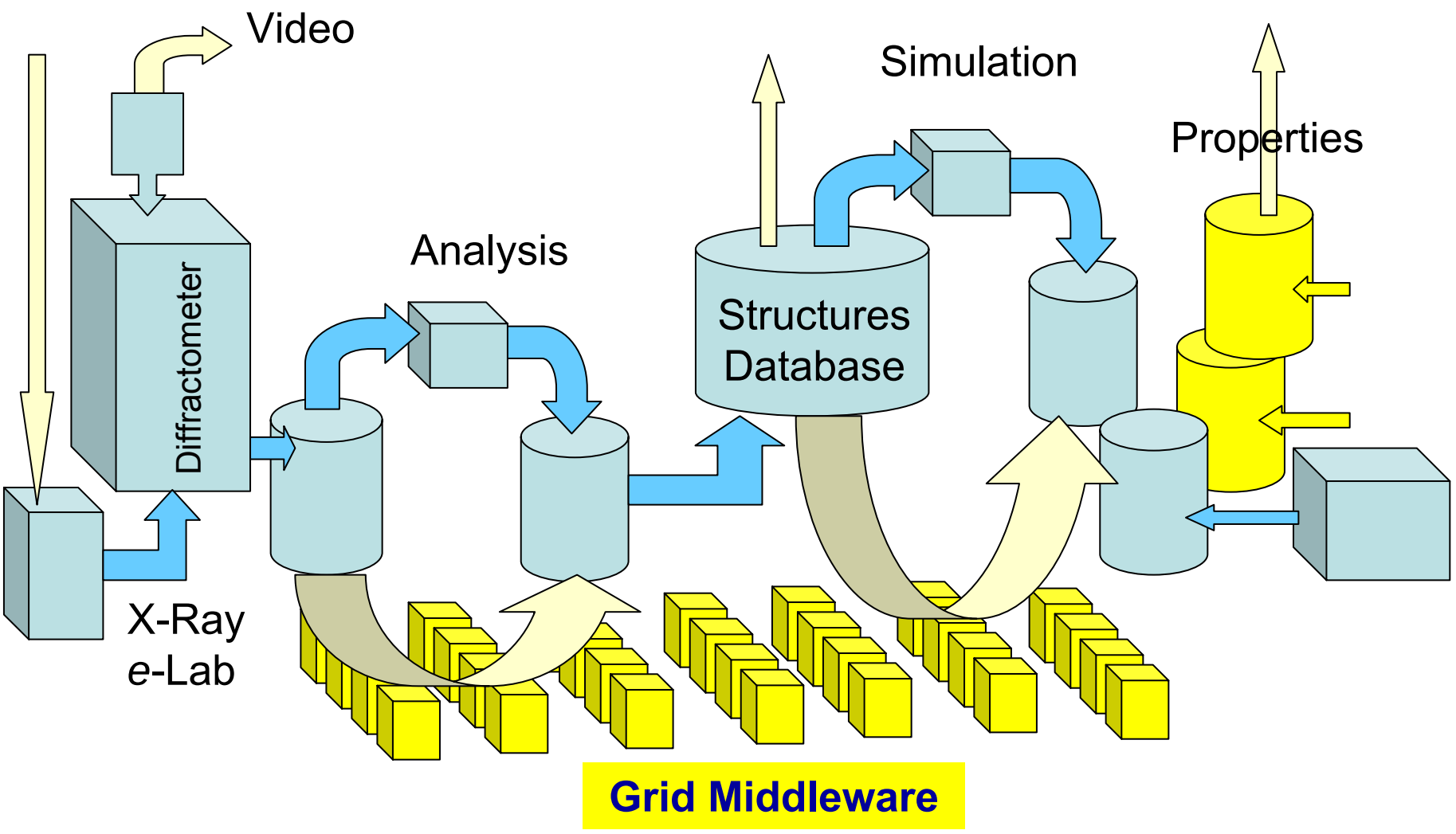


NERC DataGrid



British Oceanographic Data Centre

Comb-*e*-Chem Project



Açık veriler için metadata standartları

San Diego metadata şeması:

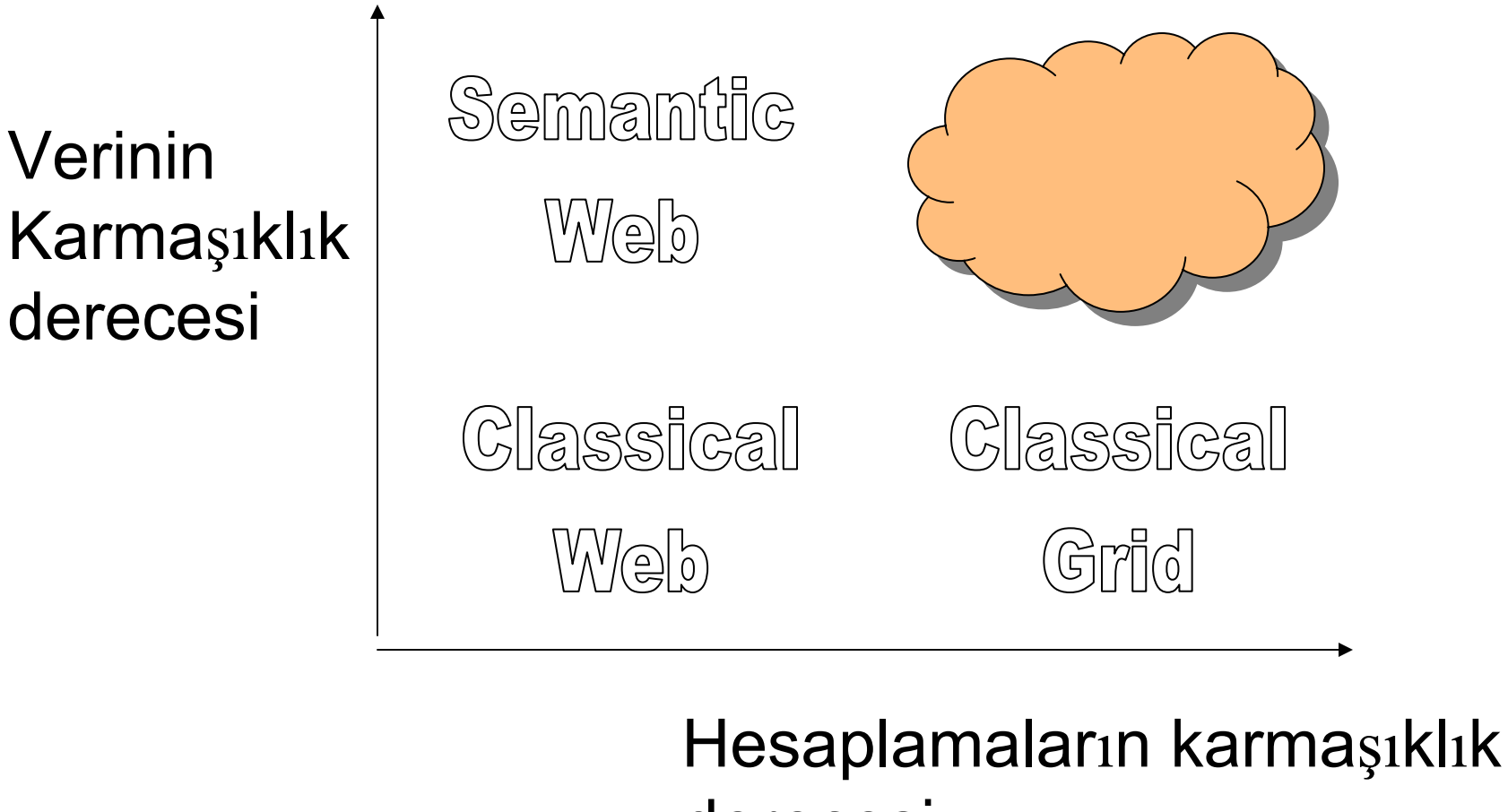
Verilerin saklaması ve erişimi için gerekli sistem metadata

Verinin özelliklerini içeren Dublin Core'a dayalı metadata

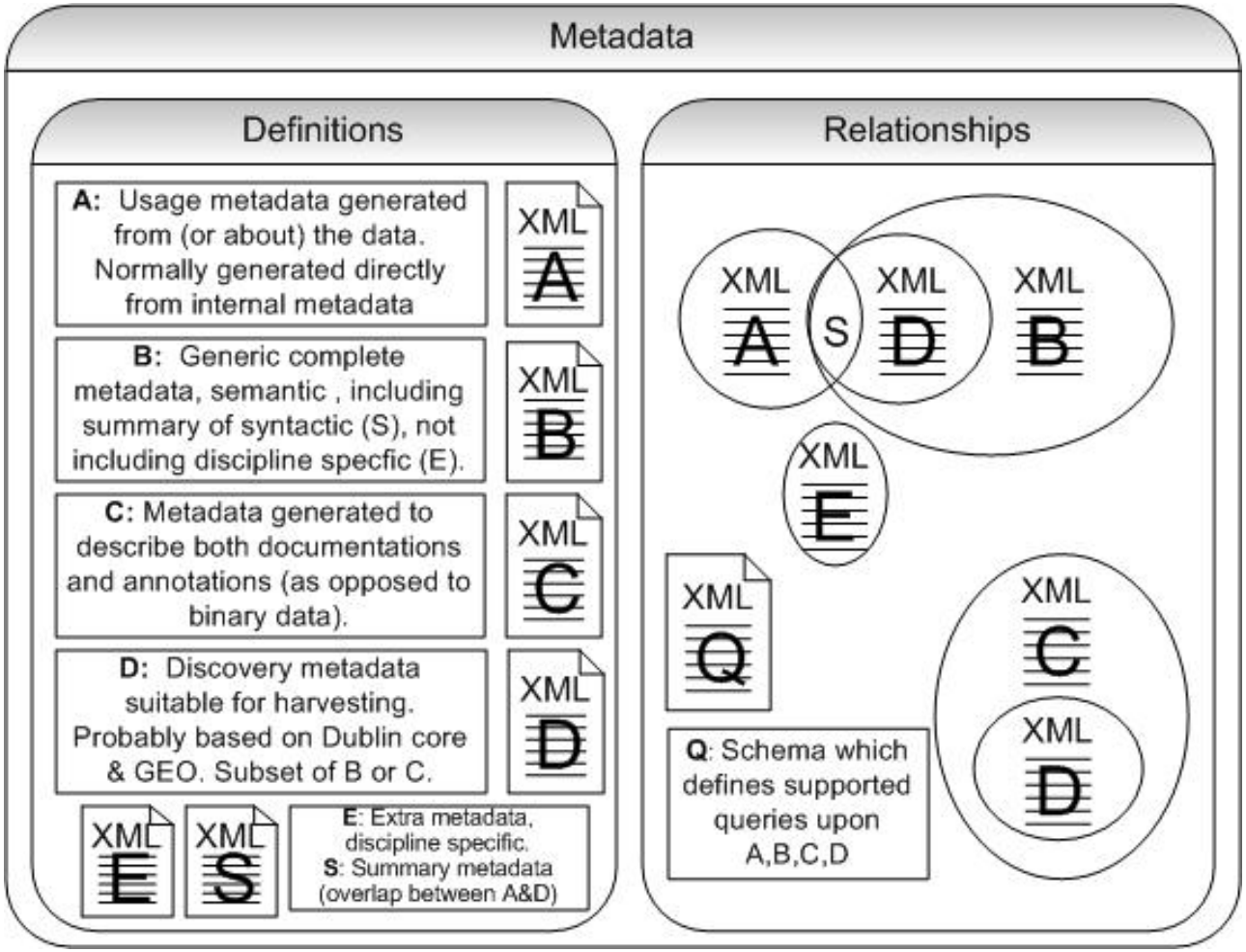
Kullanıcının erişim bilgilerini ve haklarını içeren metadata

Araştırma alanına özgü bilgileri içeren metadata

veriden bilgiye semantik ‘grid’



NERC Data Grid
Metadata Taxonomy



E-bilim ve kütüphaneler

Önümüzdeki on yıl içersinde üniversite kütüphanelerinin en önemli görevi, üniversitede üretilen bilimsel yayınların saklanması ve açık erişime sunulması olacak

Bilimsel yayınların yanı sıra, araştırmalarla ilgili verilerin de erişime sunulması ve saklanması önem kazanmakta

Verilerin bakımı

- Hastahanelerdeki milyonlarca filmin arşivlenmesi, işlenmesi
- Farklı dillerdeki yazılımların ilerde de kullanılabilirliği
- Zaman içinde değişen verilere uygun metadata standartları geliştirilmesi
- Farklı formatlardaki veriler arasında iletişim sağlanması (interoperability)