EMBL Identity & Access Management

Rupert Lück
IT Services
EMBL Heidelberg

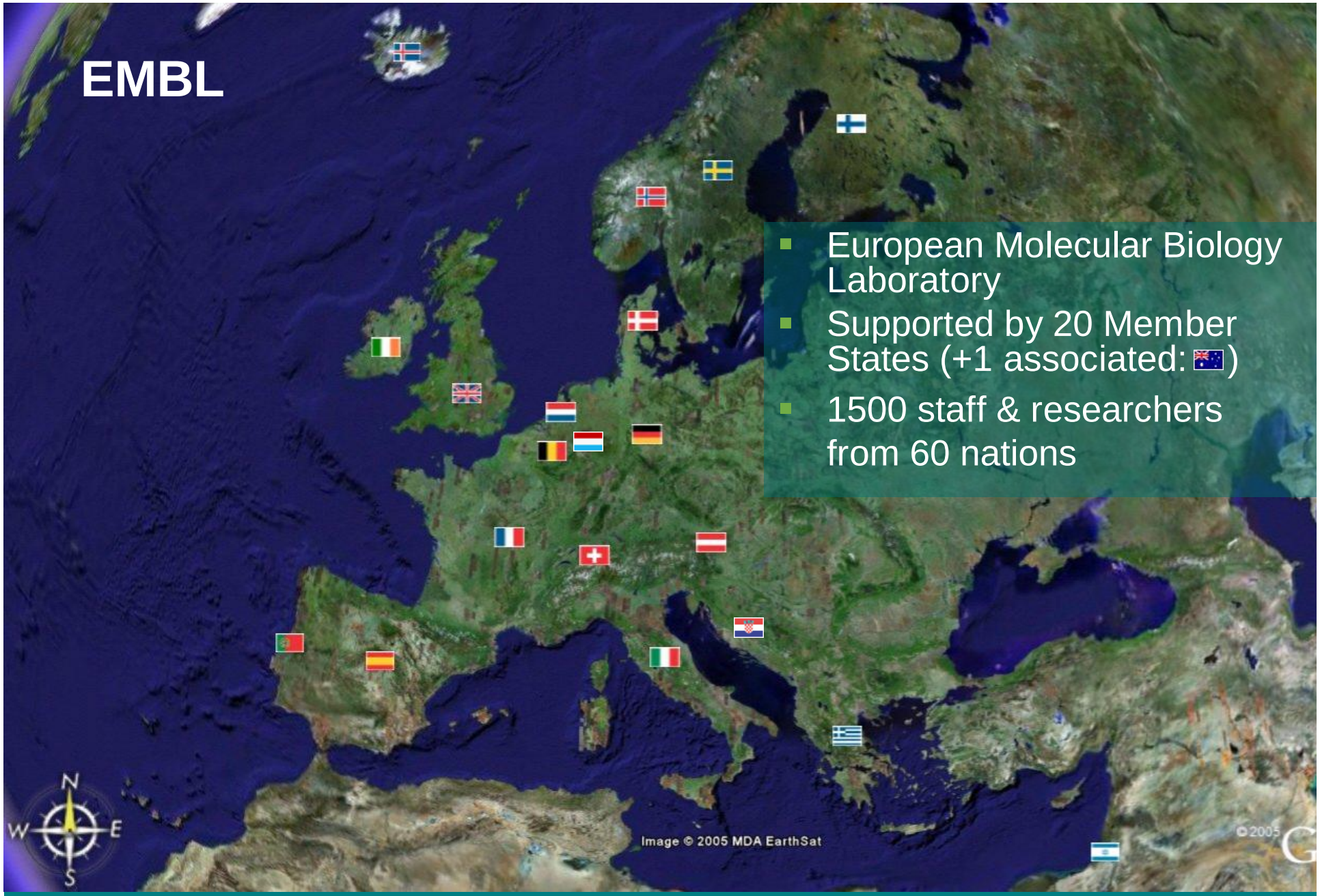
e-IRG Workshop Zürich
Apr 24th 2008

EMBL

Outline

- **EMBL Overview**
- **Identity & Access Management for EMBL**
  - **IT Requirements & Strategy**
  - **Project Goal and Features**
  - **Defining the scope**
  - **Integrated User Management**
  - **Benefits**

EMBL



- European Molecular Biology Laboratory
- Supported by 20 Member States (+1 associated:)
- 1500 staff & researchers from 60 nations

EMBL Sites

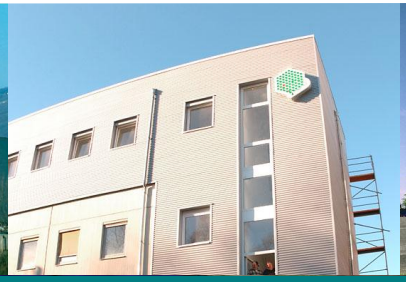
Heidelberg, DE:
Main Lab, Basic molecular biology research

Hinxton, UK:
EBI, Bioinformatics databases, research & services

Grenoble, F & Hamburg, DE:
Structural Biology

Monterotondo, I:
Mouse Biology

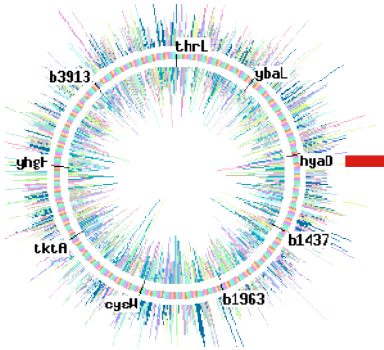


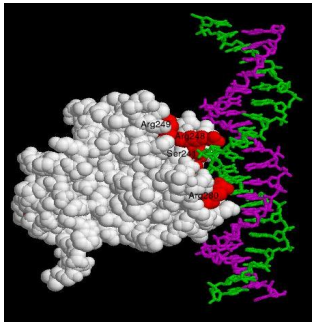

EMBL's Mission

- Flagship Lab for Basic Research in Molecular Biology
- Instrumentation & Technology Development
- Services
- Advanced Training
- Technology transfer

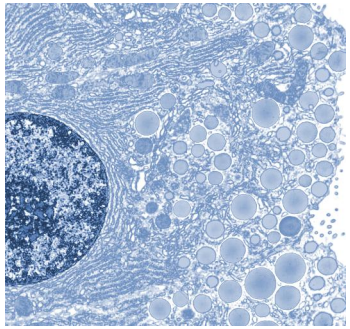
Systems Biology: From Molecules to Organisms



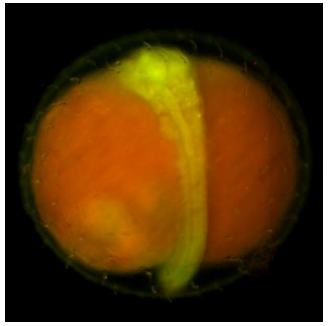
Genome



Protein/DNA



Cell



Embryo

Fruitfly

Mouse

Human Development,
Ageing, Disease

Systems Biology

- Understand Cell Function as a dynamic biological system
  - Away from *one gene – one function* concept
  - Towards quantitative understanding of living systems

- Involves
  - Interdisciplinary Research across scientific domains
    - Collaboration infrastructures
    - Data sharing & data integration
  - Quantitative studies & Integration of information
  - Technologically complex experimentation
  - Computational approaches
    - modeling and simulation
    - Highly compute and storage intensive (Grid technology)

Instrumentation & Technology Development

- NG Sequencing, microarrays, databases, screens…
- Light Microscopy *(4D confocal microscopy, cell assays screening, …)*
- Electron & Synchrotron tomography
- High throughput proteomics and structure analysis
- Modelling of biological processes
- small animal imaging

  - Large amounts of heterogeneous data (PetaByte+ range)
  - Significant needs for Network, Compute & Storage Resources
  - Scalability of IT

EMBL Services

- More than **2000 Facility Users** per year
  use the radiation sources for structural biology

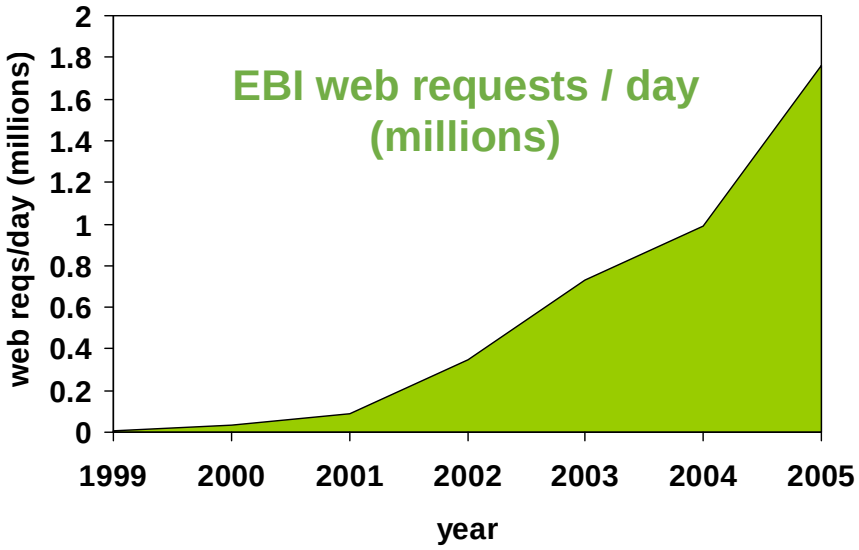
- More than **200,000 scientists** per year
  from all life sciences branches
  use the EMBL bioinformatic data resources

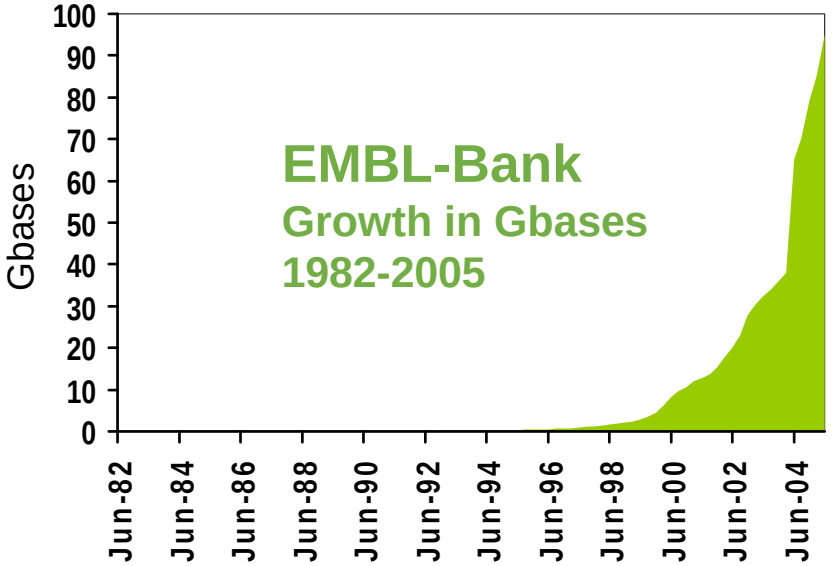
- More than **1000 visitors** per year
  benefit from state-of-the-art equipment
  learn new techniques
  carry out collaborative projects

EBI Services

- Reference site for biological data
  - 150 different databases
  - 120+ different tools.
  - 9 different data submission systems.
  - 8 major query interfaces.
- User base
  - Rapidly growing
  - > 100.000 different Users / Month
  - Scientific community
  - Pharma & Biotech Industry
- Trends
  - Rapid growth of data
  - Faster than Moore's law

    => Service oriented architecture
    - Web Service based access
    - Database Federation
    - Grid approach





[Source: Peter Stoehr, EBI]

Outline

- EMBL Overview
- Identity & Access Management for EMBL
  - IT Requirements & Strategy
  - Project Goal and Features
  - Defining the scope
  - Integrated User Management
  - Benefits

IT Requirements & Strategy

- IT Requirements
  - Collaboration IT Environment
    to support Interdisciplinary research
  - Scalability, Efficiency & Reliability
    of IT infrastructure and processes

- Strategy
  - **Institution-wide Collaboration Platform**
  - **Identity & Access management solution**
  - Consolidation
  - IT Standards

Project: Identity & Access Management for EMBL

- Project goal
  - Provide an EMBL-wide user database – ***EMBL Network Passport***

- Key features
  - Based on an LDAP
  - Identity management and provisioning infrastructure
  - *Unified Login* and *Single-Sign-On where reasonable*
  - Automated fine-grained provisioning of resources to different user populations
  - Balanced implementation effort and cost
  - Future flexibility

Defining the scope

- Resources
- User & Client populations
- Access roles
- IT Security domains

IT Resource Landscape

- HPC Clusters
  - Several 1000 CPU cores
  - mainly in Heidelberg and at the EBI
  - NIS
- Storage Systems
  - > 700 TByte primary storage
  - on NetApp and BlueArc NAS
  - 3 PB secondary storage
  - NIS, AD
- Network
  - WLAN (Radius)
  - VPN (Radius)
  - Multiple VLANs
  - Inter-campus VPN
- Applications
  - Small to enterprise level application server based
  - Web apps and native clients
  - Scientific and commercial line of business systems
  - LDAP, individual access silos
- Database systems
  - Oracle
  - MySQL
- Desktop and Server Systems
  - Operating systems (Windows, MacOS X and Linux)

User / Client populations

- Named users
  - Staff:
    - ~1500 across 5 different EMBL sites
    - 9yr contracts max.
  - Visitors: >1000 / Year
  - Facility users: >2000 / Year
  - Contractors & Consultants
  - *e*-Collaborators: >500
  - Alumni: >4000
  - Industry: collaborations & programme

- High fluctuation
- Even between populations

- Public access:
  - Scientific tool and content DB user populations (200.000+)

Access Roles
(selection)

- VPN Access

- Unix / NIS Account
- Windows / Active Directory Account

- Email Account
- Access to Intranet
- Access to shared workspaces
- Access to resource booking system (Microscopes, Rooms, etc.)
- SAP: can use online shopping module (SRM)
- SAP Modules X, Y, Z: can manage data
- Access to scientific application X,Y,Z

- Oracle DB user / access roles

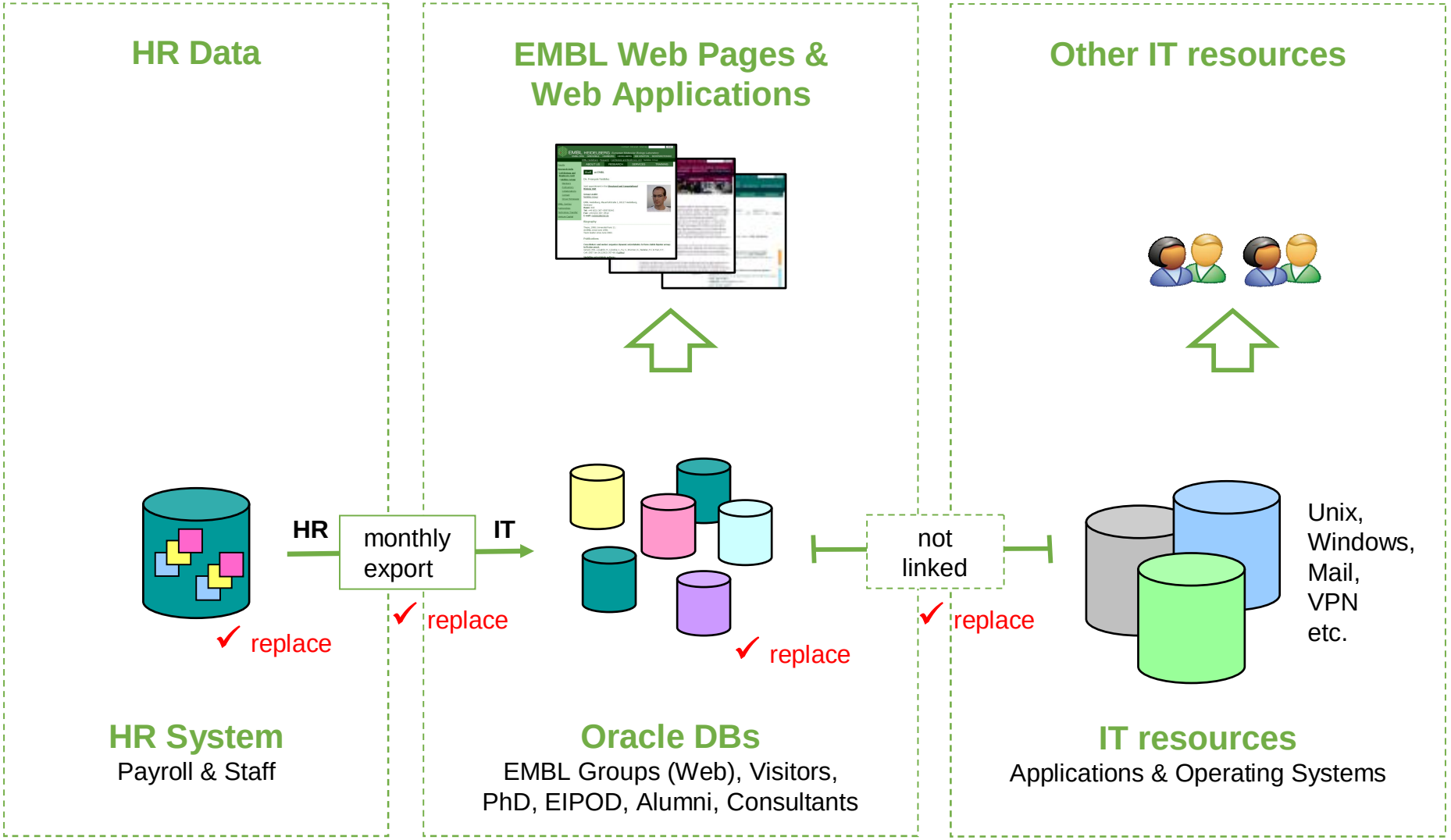
IT Security domains

- EMBL's organization is distributed across 5 sites

- Individual IT Services organizations
  - Responsible for local IT management
    (Site in Rome, managed from Heidelberg)
  - Local IT security
  - Inter-site security as a joint effort

  - ➢ Split user domains
  - ➢ Blocks efficient collaboration

User Management ▸ Until 2007



User Management ▸ Short comings

- Many different identities in different systems
- Huge efforts
  - to manage individual identities and access profiles
  - To achieve a reasonable level of consistency
- No fine-grained assignment of access patterns
- By default only access to IT infrastructure of users EMBL home site
- Many existing (self developed) systems cannot be integrated with others

Integrated User & Access Management ▸ 2008+

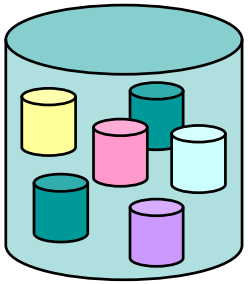
Master Data
(one central resource)

Payroll, Staff
EMBL Groups (Web),
EIPOD, PhD, Visitors,
Alumni, Consultants

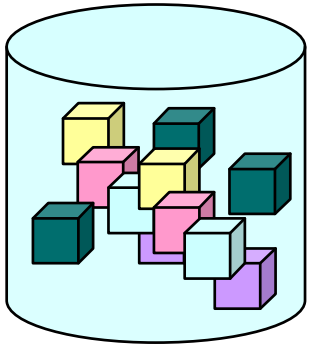
User & Identity Management

Access Management

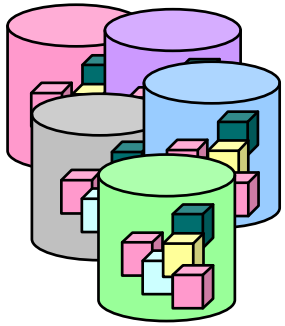
Template based
Provisioning



HR → sync → IT



sync



Unix,
Windows,
Mail,
VPN
Web CMS,
SAP,
Oracle,
etc.

SAP HR / OM

LDAP / Oracle IM
EMBL User Directory &
Identity Management

IT resources
Applications & Operating Systems

Integrated User Management ▸ Benefits

- One central user directory (LDAP)
  - for all people associated with EMBL
  - from all sites
  - not only staff
- Automation of access rights management and provisioning to IT resources
- Real time information displayed on the EMBL web
- LDAP is a standard component
  - Easy Integration in future projects
  - Can also be used by any application developer within EMBL
  - Integration projects costs significantly lower

Integrated User Management ▸ Collaboration Benefits

- EMBL-wide unified login (username & password)
  e.g. NIS, Windows, SAP, Storage systems,…
- Ability to login while visiting another EMBL site
- Access to remote (expensive) analysis tools e.g. via Terminal Server
- Secure sharing of data with EMBL colleagues from remote sites
- Resource booking and checking peoples availability across the organization

Integrated User Management ▸ Technical Benefits

- Provisioning templates allow fine-grained access management
  - i.e. a user population could get access to many resources
  - Others only could be assigned email-only access
- Why a commercial solution
  - Vendors like Oracle provide out-of-the-box connectors to other access infrastructures, e.g.
    - Active Directory
    - LDAP (various vendors)
    - UNIX, NIS
    - SAP (various modules)
  - Allows faster and cost effective integration of other infrastructures
  - Federations:
    - Supports Liberty alliance standard
    - Federations across organizations also to industry partners

Summary

- Systems biology at EMBL requires a collaborative, scalable and secure IT environment to enable research and to protect IP
- The introduced an identity management and provisioning infrastructure is one of the key components to support this requirement
- It allows automated fine-tuning of individual access scenarios
- Allows fast and cost effective integration of other infrastructures