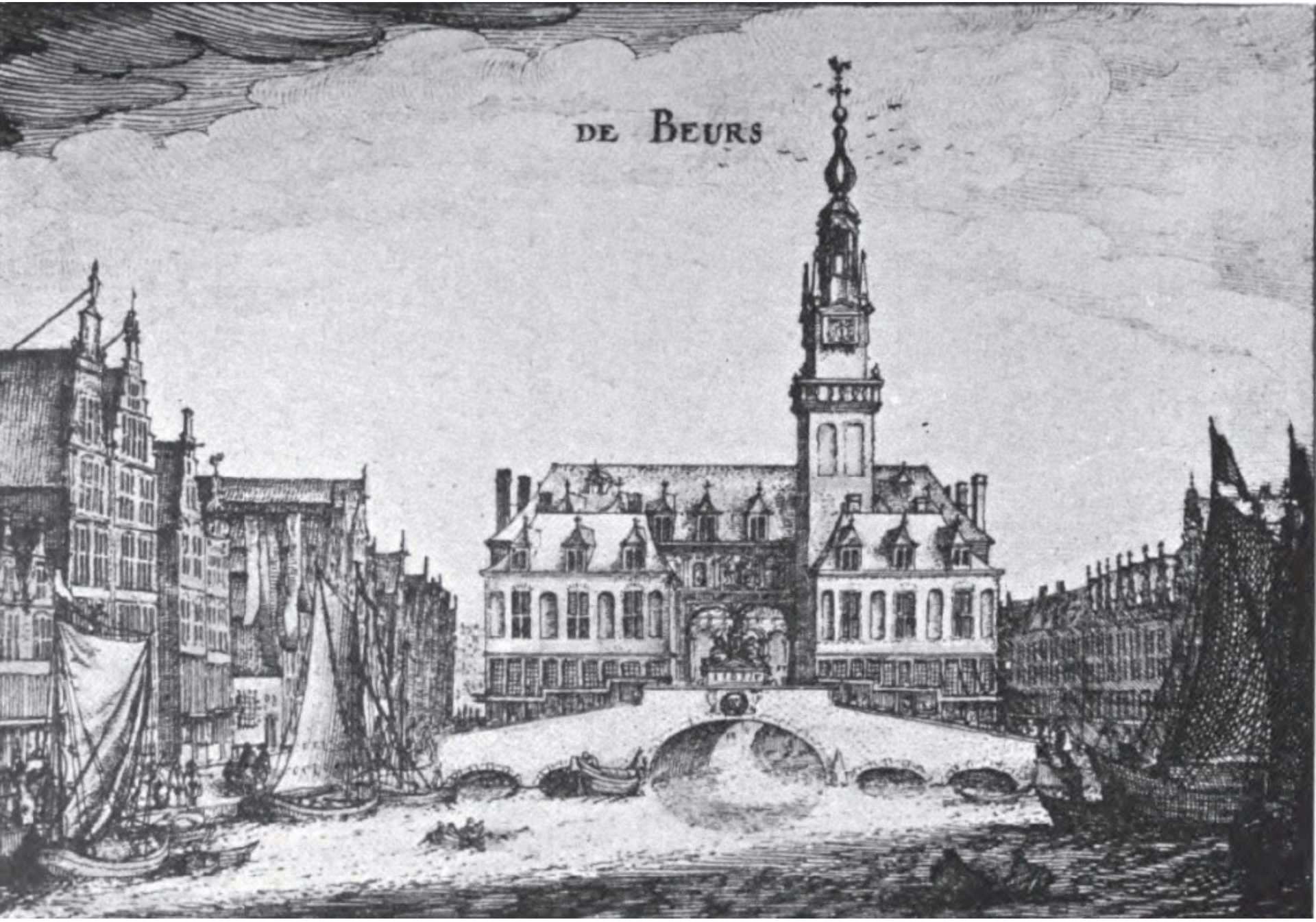
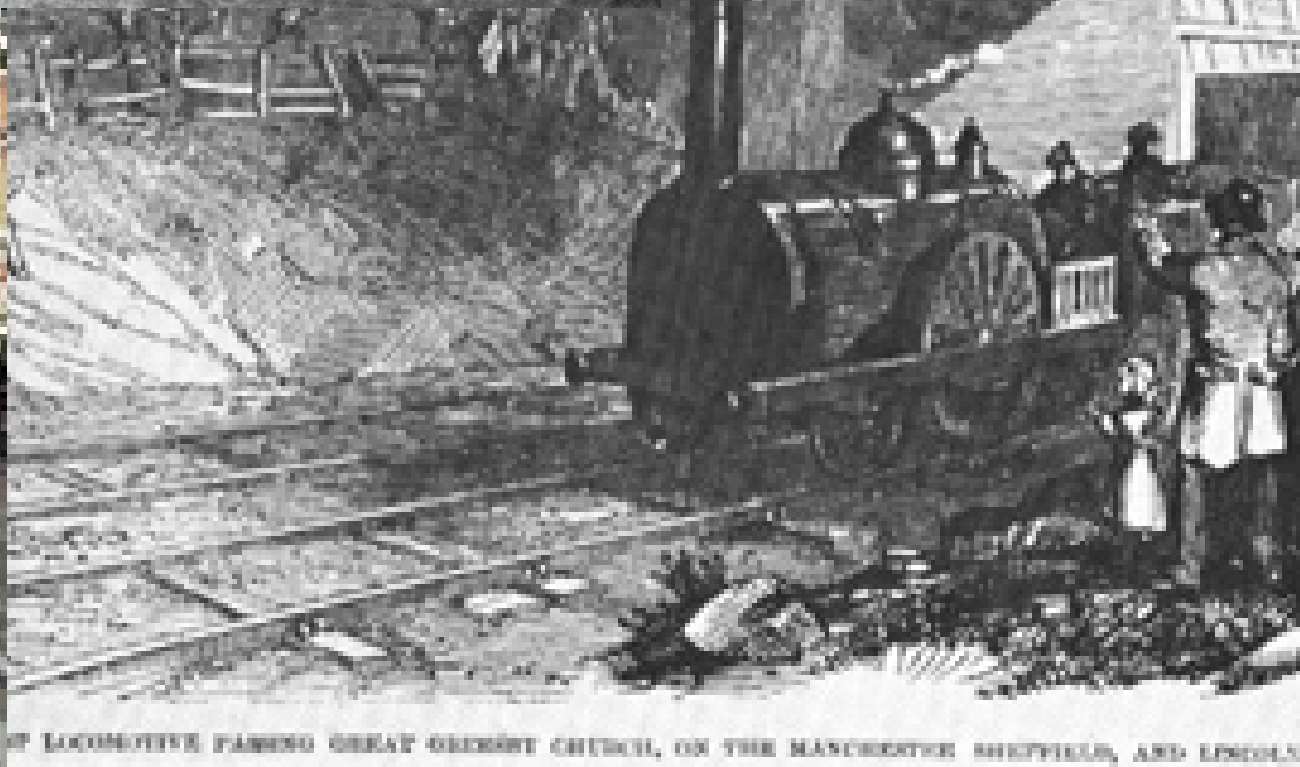

DE BEURS

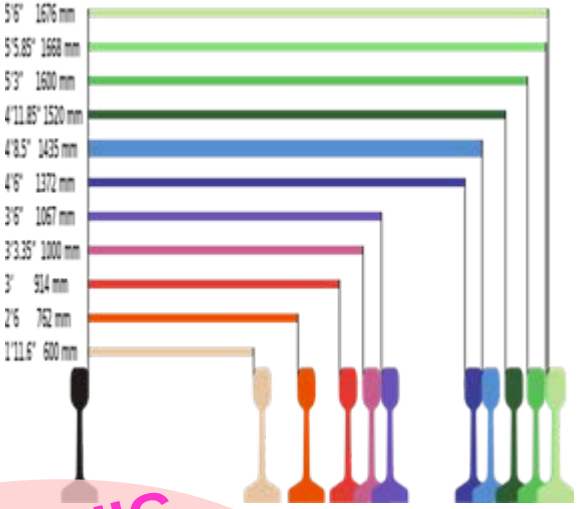
CHT PASSAGIERS NAAR LONDEN, DUS DRIE EXTRA-VLIEGTUIGEN. 1920. Foto FAVIER



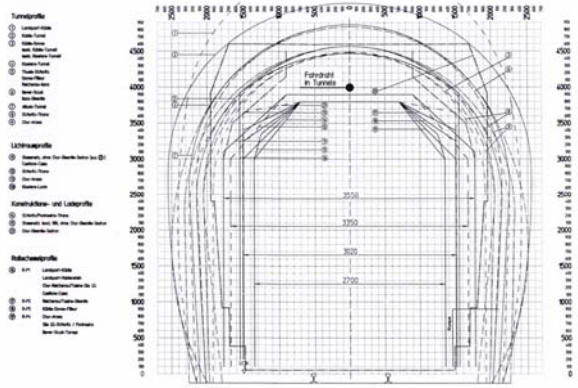
LOCOMOTIVE PASSING GREAT GEORGE CHURCH, ON THE MANCHESTER SHEFFIELD, AND LINCOLN

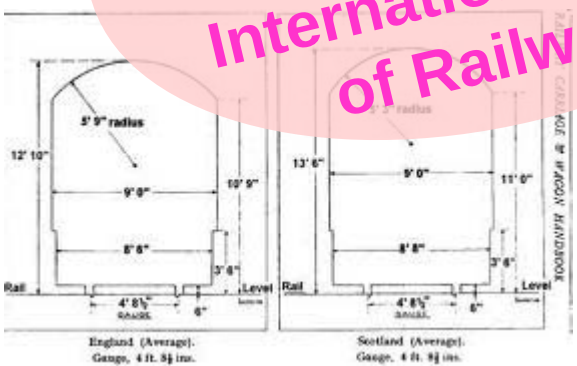
Railway track and loading gauges

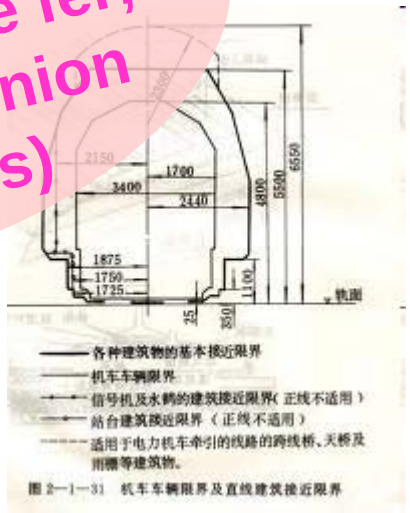
1,676 mm (5 ft 6 in)	Indian gauge
1,668 mm (5 ft 5 2⁄3 in)	Iberian gauge
1,600 mm (5 ft 3 in)	Irish gauge
1,588 mm (5 ft 2 1⁄2 in)	Pennsylvania Trolley Gauge
1,581 mm (5 ft 2 1⁄4 in)	Pennsylvania Trolley Gauge
1,524 mm (5 ft)	Russian gauge
1,520 mm (4 ft 11 5⁄6 in)	Russian gauge
1,435 mm (4 ft 8 1⁄2 in)	Standard gauge
1,372 mm (4 ft 6 in)	Scotch gauge
1,067 mm (3 ft 6 in)	CAP gauge or Cape gauge
1,000 mm (3 ft 3 3⁄8 in)	Metre gauge
950 mm (3 ft 1 3⁄8 in)	Italian metre gauge
891 mm (2 ft 11 1⁄10 in)	Swedish narrow gauge



Governed by UIC (Union Internationale des Chemins de fer, International Union of Railways)








Hamburg
Amsterdam
Hanover
Berlin
Schiphol
Dortmund
Rotterdam
London
Antwerp
Duisburg
Essen
Brussels
Cologne
Lille
Liege
Aix-la-Chapelle
Frankfurt
Luxembourg
Frankfurt Airport
Mannheim
Roissy CDG
Reims
Metz
Karlsruhe
Nuremberg
Le Mans
Paris
Stuttgart
Rennes
Marne-la-Vallée
Nancy
Massy
Strasbourg
Linz
Angers
Freiburg
Munich
Vienna
Tours
Salzburg
Nantes
Basel
Zurich
Innsbruck
Dijon
Poitiers
Berne
Lausanne
Angoulême
Geneva
Lyon
Bordeaux
Valence
Nimes
Avignon
Montpellier
Aix-en-Provence
Marseille
High speed services:
ICE
TGV
eurostar
THALYS
TGV lyria
A Railteam journey is an international journey, direct or with a connection, on the Railteam network


EEF - European E-Infrastructure Forum
Analysis of ESFRI projects' e-infrastructure requirements by the European e-Infrastructure Forum
Bob Jones, CERN
bob.jones@cern.ch
Bob Jones
EEF - European E-Infrastructure Forum
First pass analysis of ESFRI requirements (all sectors) identified the following common areas
• Single sign-on – consistent access to resources
• Virtual organisations (collaboration)
• Persistent storage – long term preservation of data and its access
• Data Management services
• Standards – web services
• Workflows – support of access to HPC/grid/network resources (compute & data) across Europe for the
• Training & consultancy
• Global scope – beyond Europe
Bob Jones
11

LIFEWATCH

An e-Science infrastructure for biodiversity research

Alex Hardisty

LifeWatch Technical Construction Team
and
Director of Informatics Projects
School of Computer Science & Informatics

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

What is LifeWatch?

- An e-Science infrastructure
  - Exploration of patterns of biodiversity and processes of biodiversity across time and space
  - What causes species diversity?
- A European Research Infrastructure
  - Distributed observatories / sensors
  - Databases, processing and analytical tools
  - Computational capability and capacity
  - Collaborative environments
  - Support, training, partnering, fellowship
- Open access, single portal

Intergovernmental Platform on Biodiversity and Ecosystem Services

"Representatives of 86 governments recommend that UNGA 65 should be invited to … take appropriate action to establish the platform [IPBES]"

Supported by:

Based on a presentation given by Ibrahim Thiaw, UN Environment Programme

Film: Introduction & 2 case studies

- No.1 European research infrastructure for biodiversity
  - Represents a new methodological approach to understanding biodiversity as a whole interacting system
  - Integrating across scales: Genomic; organism; habitat; ecosystem; landscape
- Bird strike monitoring
  - Understanding the patterns & behaviours of bird movements can help improve aviation safety
- Urban sprawl
  - Achieving balance between development of urban areas and conservation of biodiversity

Mission

The mission of LifeWatch is to construct and operate a distributed infrastructure for biodiversity and ecosystem science based upon Europe-wide strategies implemented at the local level: individuals, research groups, institutions, countries.

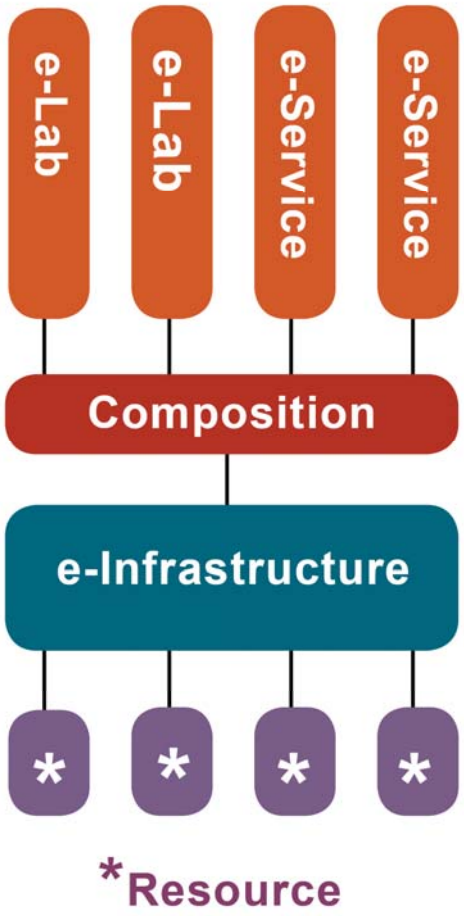
In cooperation with National LifeWatch Initiatives, LifeWatch provides:

- Organisation;
- Technical direction & governance;
- Core ICT infrastructure;
- Management of the LifeWatch “Product”; and,
- Community support.

Aspiration: An integrating "*Infrastructure*" for biodiversity research

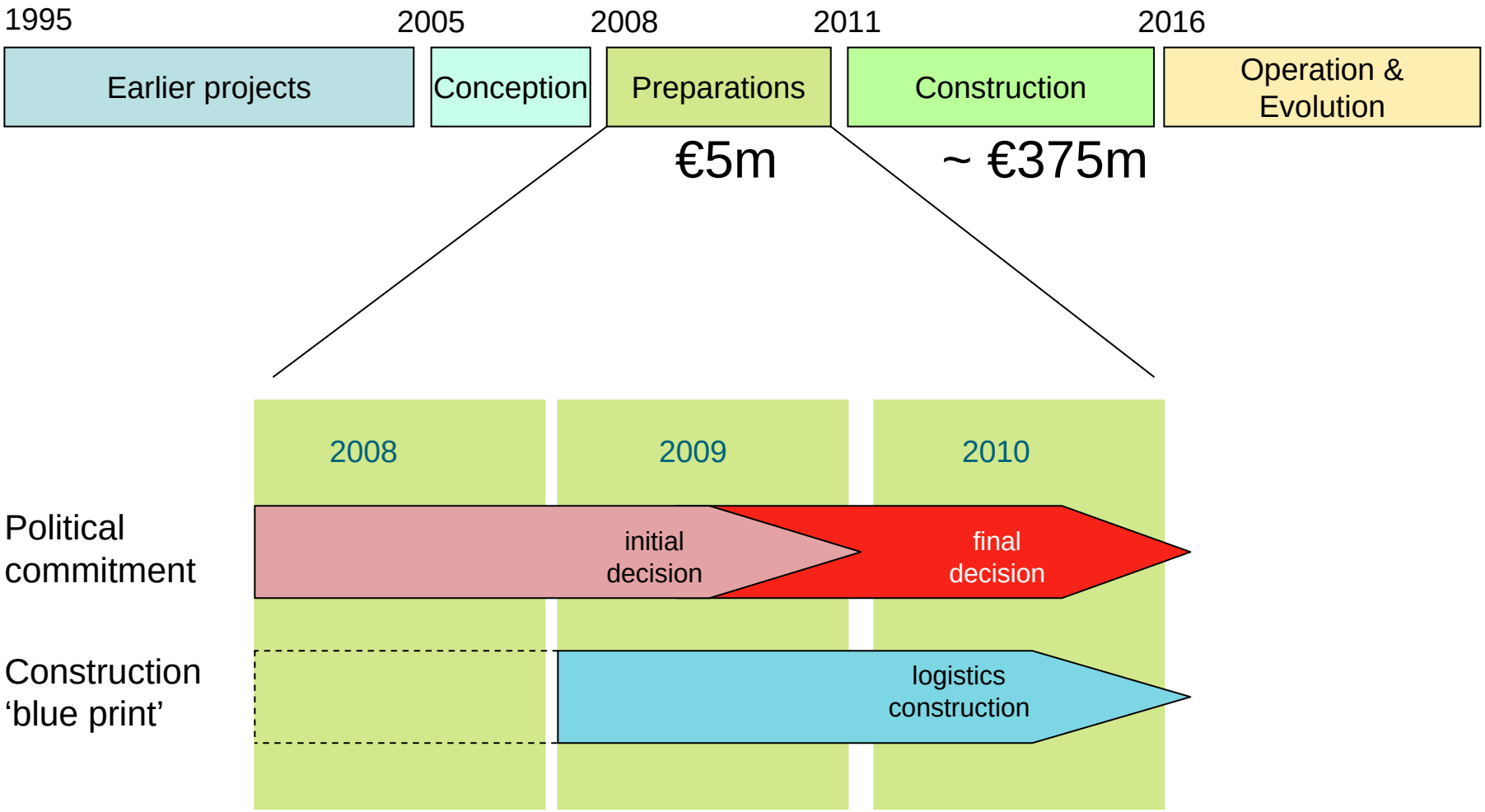
- Full range of functions across multiple scales
  - Data gathering and generation; data management, integration and modelling; diverse applications
  - Genomic; organism; habitat; ecosystem; landscape
- Benefits to the research community
  1. Discovery and access to a wide variety of data – species, genetic, ecological and abiotic – to support biodiversity research
  2. Manage / merge data from multiple sources
  3. Taxonomic support e.g., authoritative species lists and taxonomic classifications, digitisation-on-demand
  4. Spatial mapping of data; INSPIRE compliance
  5. Sharing of workflows, collaboration and community-building

A community driven e-Infrastructure



- Centres, distributed across countries offer services to users
  - ICT oriented (computer centres, data centres), human oriented (service centres), or a combination
- User projects create their own e-laboratories or e-services
- They share their data and algorithms with others, while controlling access

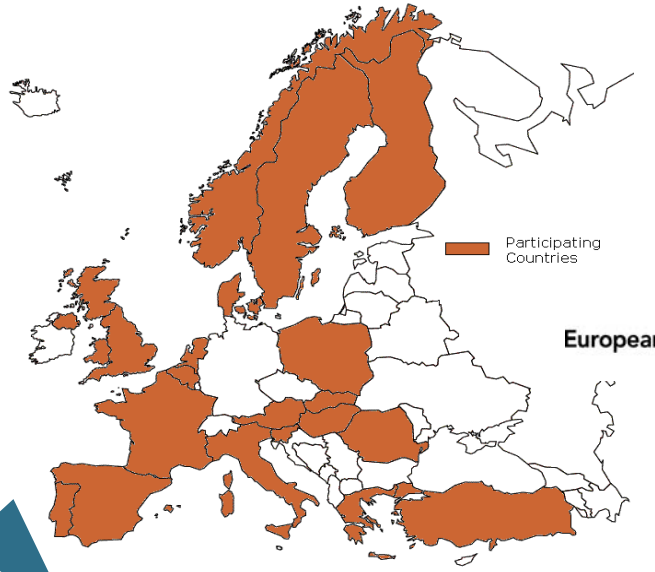
Status: The LifeWatch timeline



Status: The Preparatory Project

Contracted participants

- Countries
- Scientific networks
- 27 executive partners



8 countries negotiating the start-up

European Environment Agency

esa

Other partners

- Data networks
- International infrastructures
- User sectors
- Industry

ALTERNet

BioCASE

EDIT European Distributed Institute of Taxonomy

GBIF

ENBI

Oceans

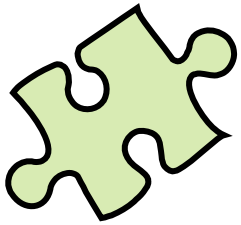
marbef

SYNTHESYS

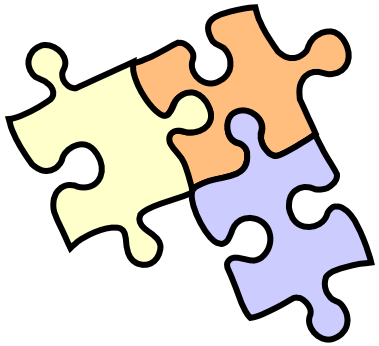
genomics

Our requirements

C	**Conditional Generic Requirements**
C1	Reliable e-infrastructures (quality, availability, common authentication & authorization, uniform configuration)
C2	Common procedures for individual users to cooperate in Virtual Organisations, while working from different countries
C3	Access to distributed computational capacity
C4	Grid access rights for institutional users with a Shibboleth-like approach
C5	Personalized access rights for not NGI/institution-based users. (Many users in env. sciences are not NGI connected)
C6	Harmonization/Standardization of character sets for web sites and databases (i.e. Unicode)
C7	Common database infrastructure and management, across domain specific ESFRI projects, including support for format conversions and reading tools, that enables efficient archiving of data, data mining, access to data and exchange of data.
O	**Operation and Usability**
O1	Digital tools to create personalised work spaces (virtual collaborative environments)
O2	Management of complex data records (fuzzy data, data calibration, data validation, data transformation, data fusion, privacy issues)
O3	Fast access to and retrieval of selected distributed and large heterogeneous data sets and tools
O4	Operation of smart/intelligent observation/sensor networks
O5	Workflow enhancement, depending on distributed data/tools retrieval speed and distributed computational capacity. Options for caches
O6	Virtualization access to clusters/supercomputers/ grid from a PC
O7	Development of real-time instrument (i.e. aircraft/ground, oceanic floats/ground) data links
O8	Availability of fast networking where these don't exist at the data generation sites
O9	Intelligent sensor networks and autonomously experimental adjustments in response to changing environmental conditions
S	**Services for environmental facilities**
S1	Repositories for data, workflows, and workflow results
S2	Support metadata/annotations as hosted externally to the core data. This metadata may be contributed by users who do not have access to the original data
S3	GUIDs, Unique Identifiers for data, files and associated metadata
S4	Repositories for modeling algorithms, software, tools
S5	Repository for tracking provenance and citation of data and workflows (+annotation services)
S6	Software libraries for supporting capabilities (data format conversions, pre-processing (filtering etc), data transformation (parameters such as coordinates) data post processing (statistics, storage)
S7	Interoperability (of heterogeneous data and tools), resource wrapper development, semantic interoperability services, retrieval of ontologies
S8	Integration of arbitrary external data sources (and outside the Grid)
S9	Super fast analysis and modeling services
D	**Application development and community involvement**
D1	Capabilities for user involvement in open-source data generation + software development (workbenches)
D2	Data resource wrapper development kits
D3	Facilitating workflow development, while providing knowledge representation and metadata updates
D4	Ability to view, visualize, explore discovered data and software resources, with reporting of usability and interoperability (and improvement suggestions)
T	**Training needs & Capacity building**
T1	Support long-term ICT development strategies
T3	Targeted information and training on new e-Infrastructure capabilities



Jigsaw of challenges



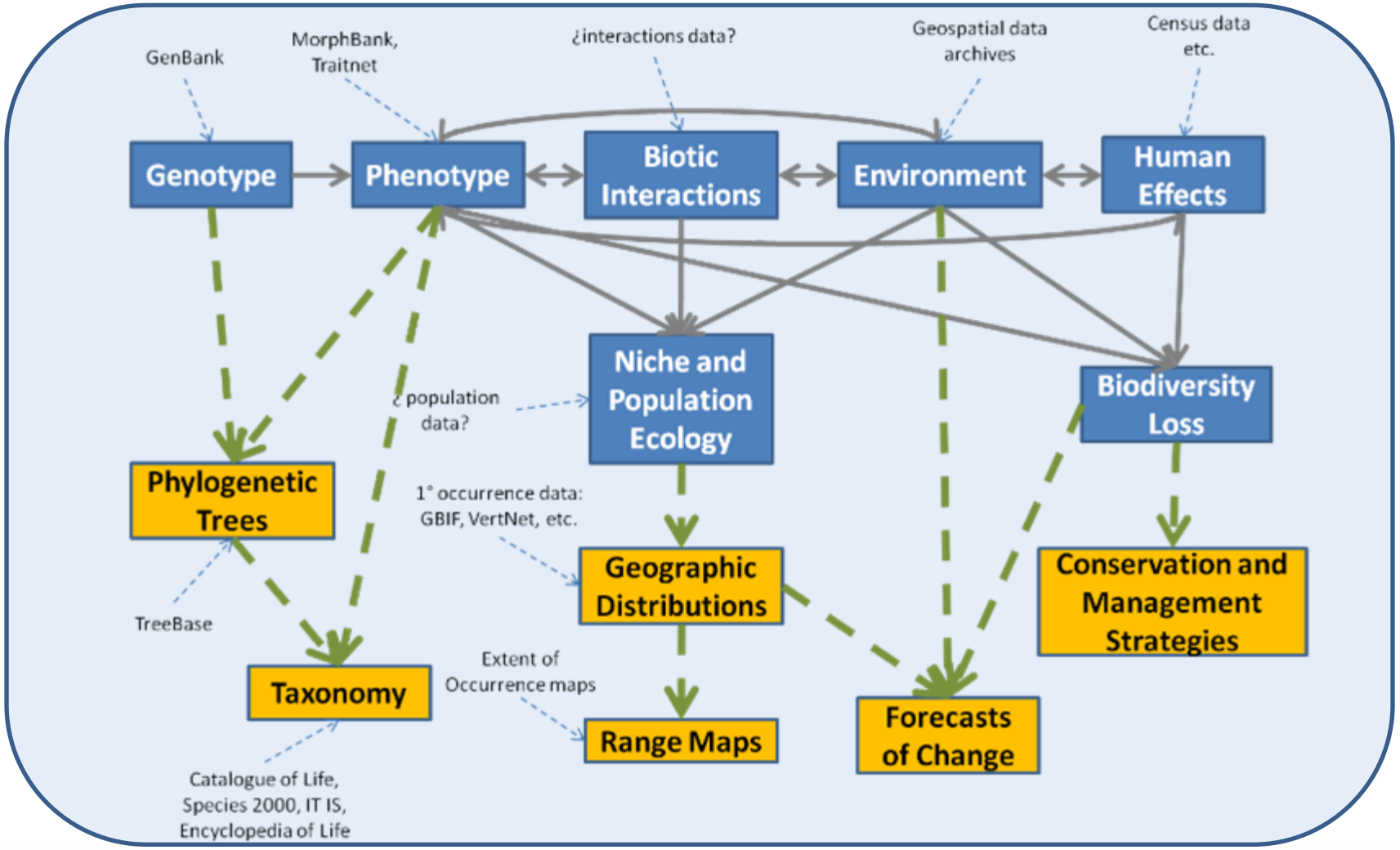
- All the usual:
  - Technical
  - Fitness-for-purpose and ease of use
  - Integration of multiple resources
  - Open and based on industry standards
  - Existing technological solutions as far as possible
  - Operational at the earliest opportunity
  - Staged; not everything available on ‘day 1’

- HETEROGENEITY, GAP, SCALE, PACE, FIT

5 challenges (and 5+ solutions)

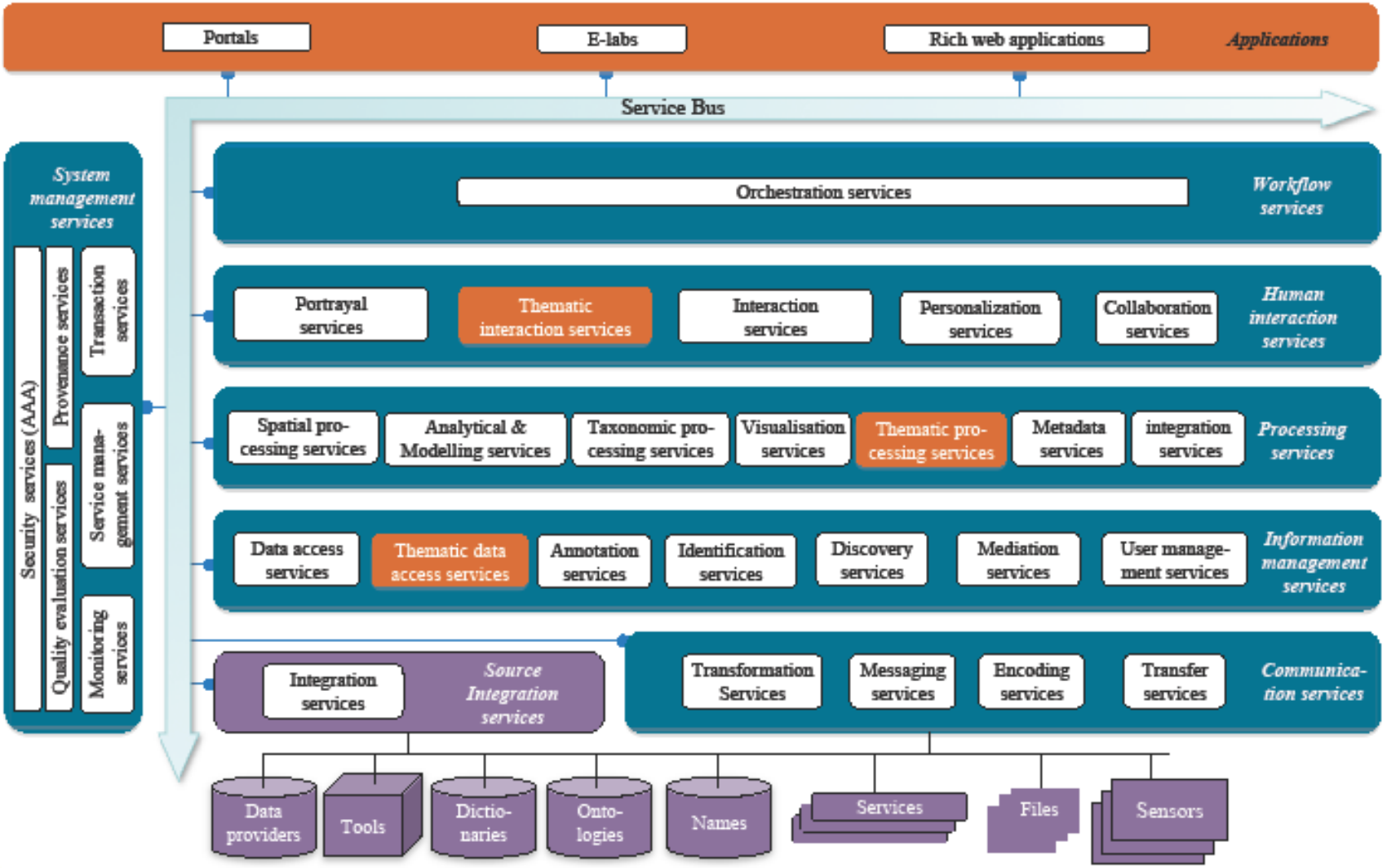
- HETEROGENEITY of the community's requirements, its data resources and tools

Challenge of HETEROGENEITY: Interconnected nature of biodiversity ideas, outputs, repositories

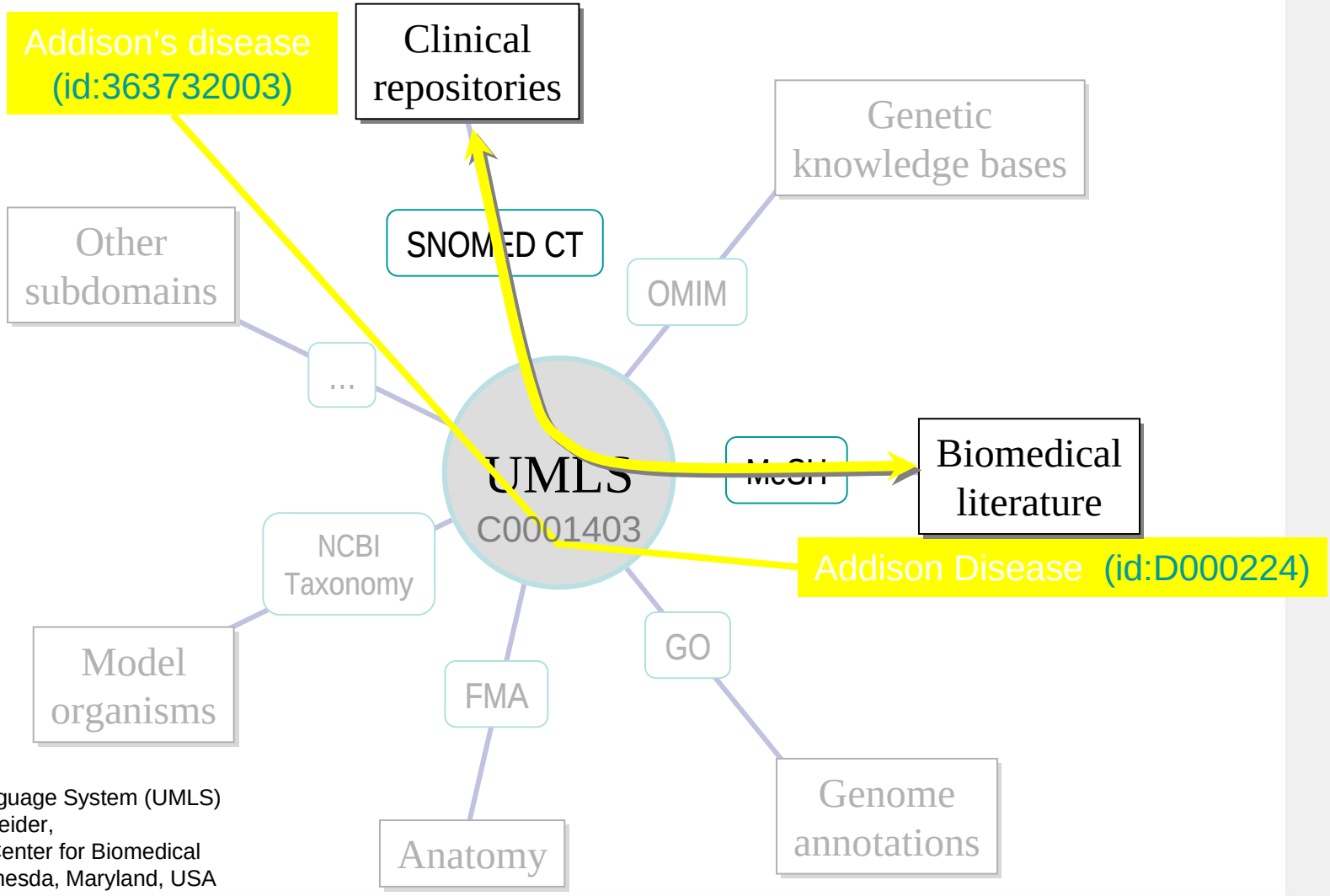


From Peterson et al (2010), Syst Biodivers 8(2), 159-168
From Guralnick and Hill (2010), http://www.slideshare.net/robgur/ievobio-keynote-talk-2010

Solution for HETEROGENEITY: An SOA approach

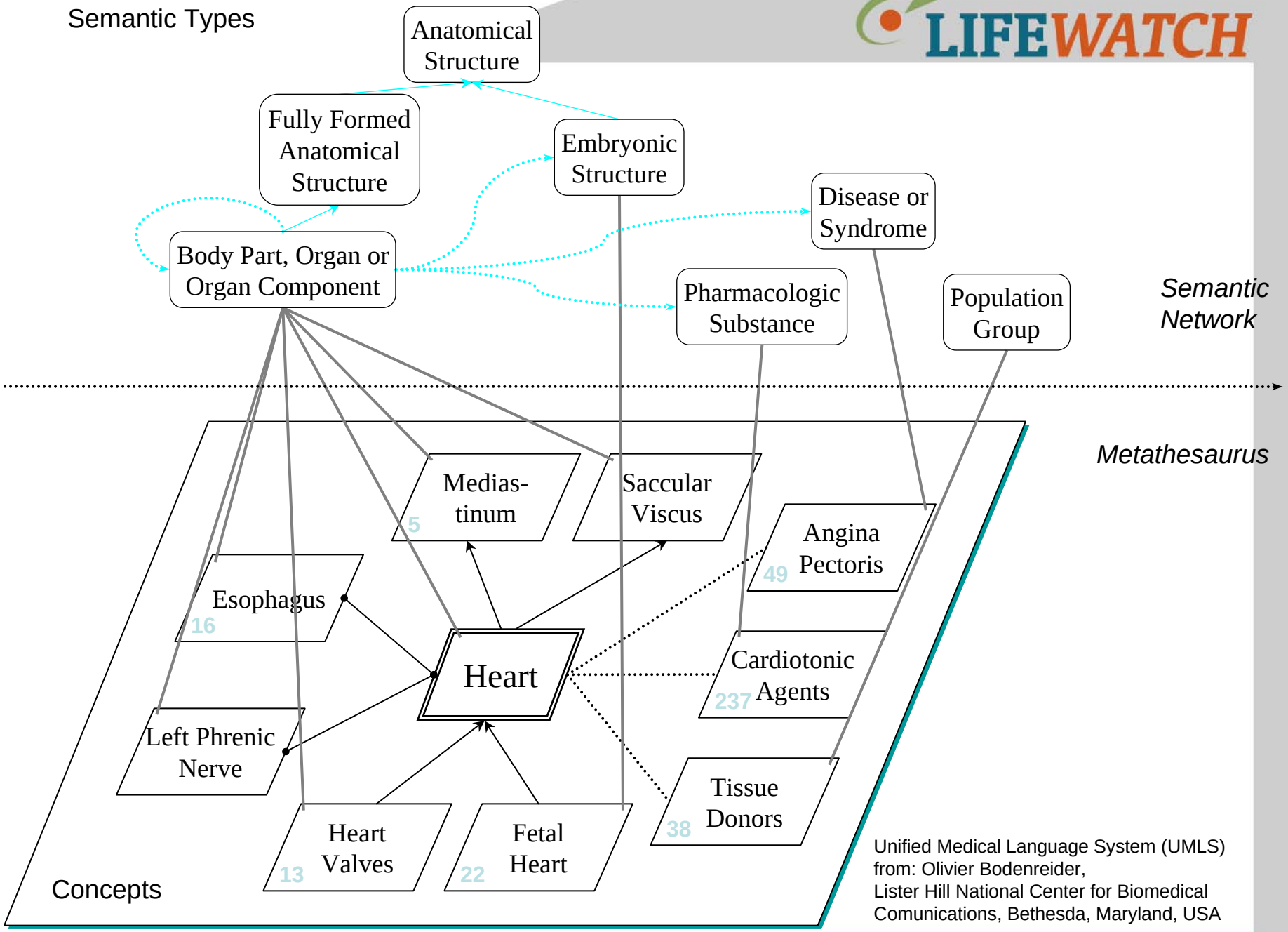


Solution for HETEROGENEITY: Semantic interoperability through knowledge management



Unified Medical Language System (UMLS)
from: Olivier Bodenreider,
Lister Hill National Center for Biomedical Comunications, Bethesda, Maryland, USA

Semantic Types

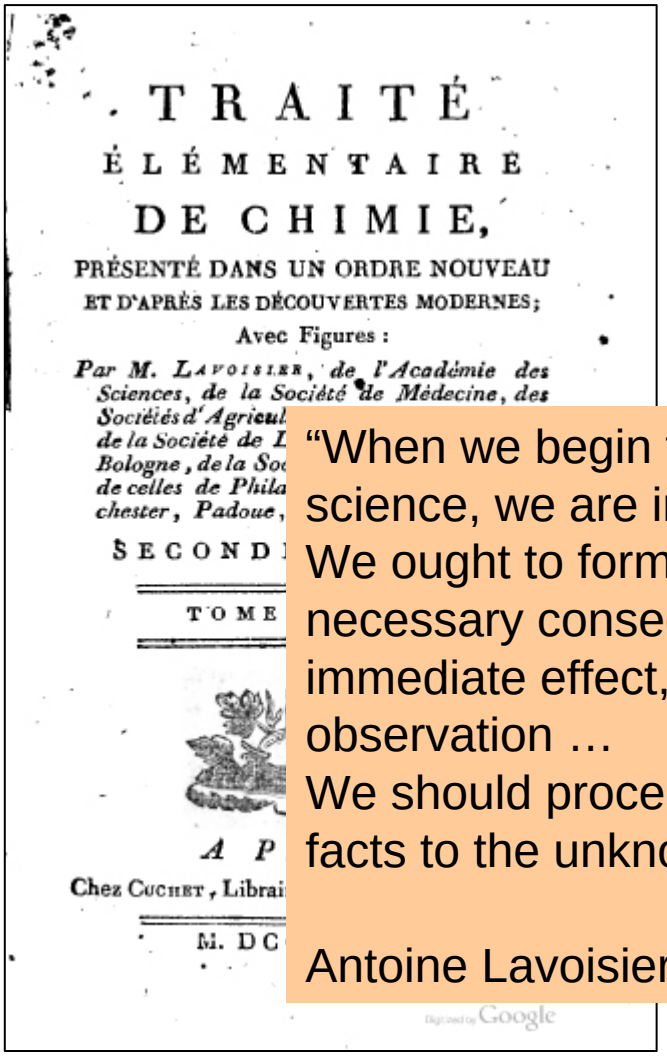


Unified Medical Language System (UMLS)
from: Olivier Bodenreider,
Lister Hill National Center for Biomedical Comunications, Bethesda, Maryland, USA

5 challenges (and 5+ solutions)

- HETEROGENEITY of the community's requirements, its data resources and tools
- GAP between current practice and future vision

GAP: Between current practice and future vision


TRAITÉ
ÉLÉMENTAIRE
DE CHIMIE,
PRÉSENTÉ DANS UN ORDRE NOUVEAU
ET D'APRÈS LES DÉCOUVERTES MODERNES;
Avec Figures :
Par M. LAVOISIER, de l'Académie des Sciences, de la Société de Médecine, des
TOME
M. DC
Digitized by Google

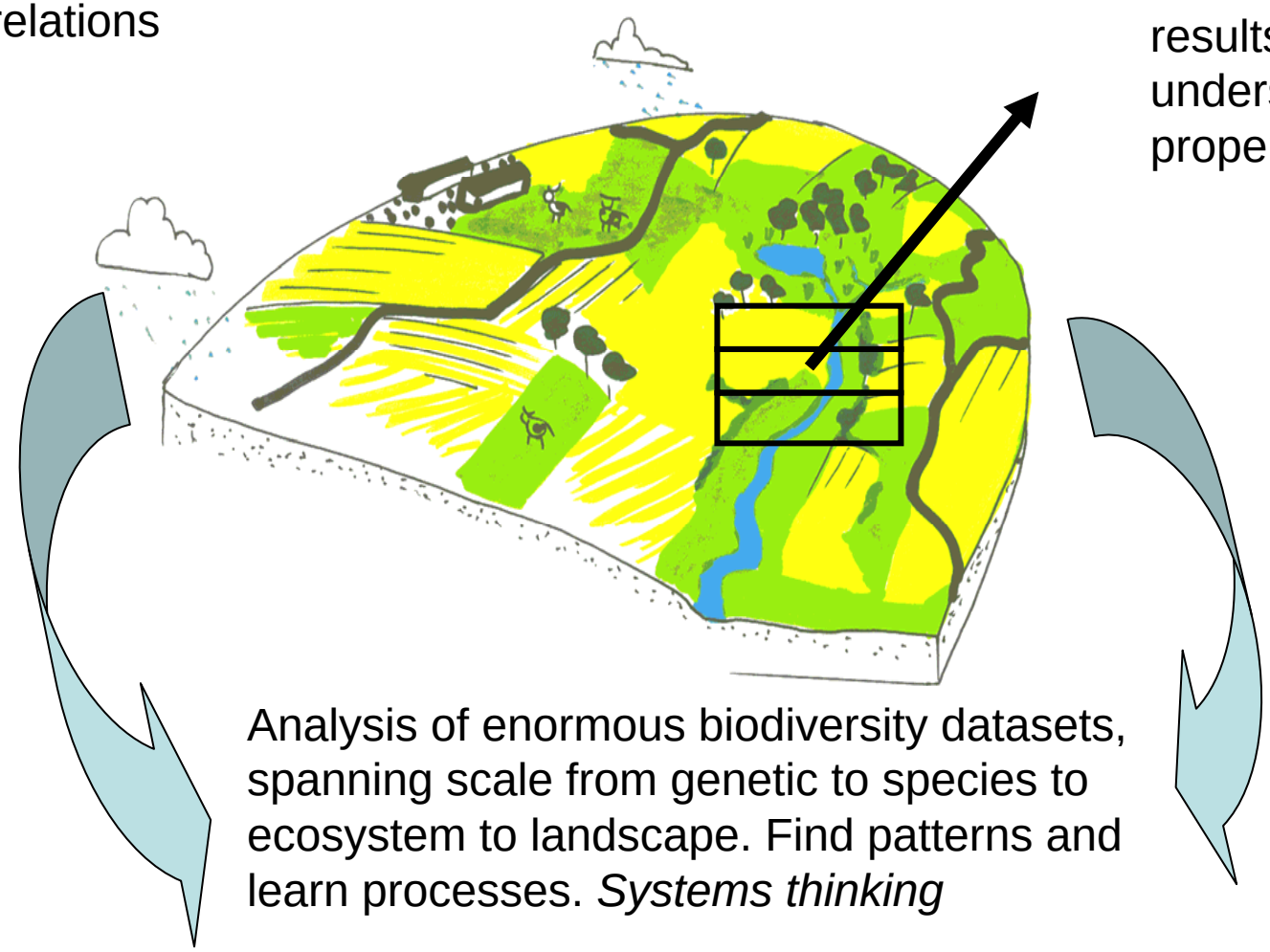
"When we begin the study of any science, we are in the situation, …
We ought to form no idea but what is a necessary consequence, and immediate effect, of an experiment or observation …
We should proceed from the known facts to the unknown"

Antoine Lavoisier, 1789

"collaborative, distributed research methods that exploit advanced computational thinking"

Malcolm Atkinson, 2007

The biodiversity *system* cannot be described by the simple sum of its components and their relations



Experimentation on a few parameters is not enough. There are limits to scaling results in order to understand *system* properties.

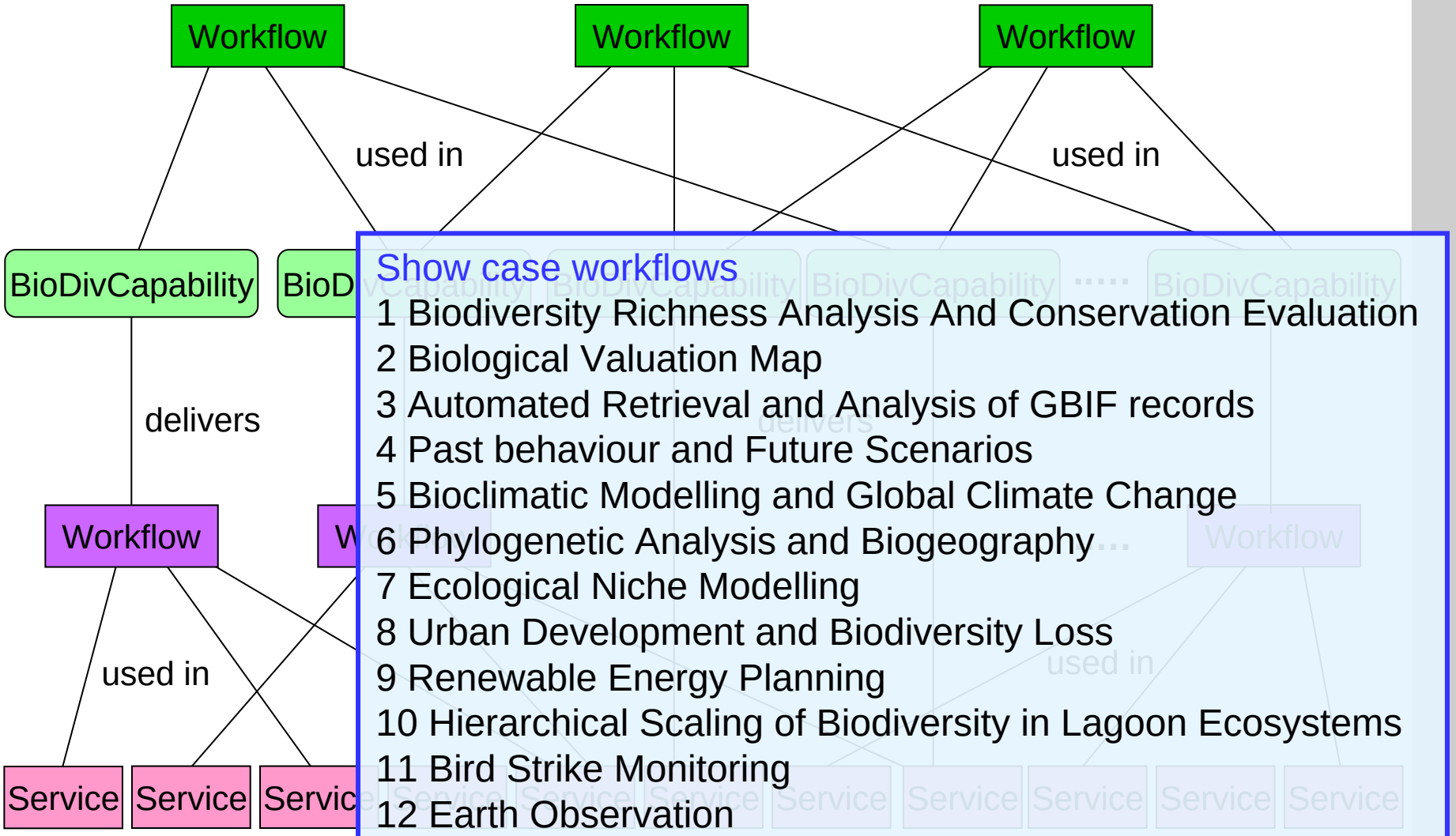
Analysis of enormous biodiversity datasets, spanning scale from genetic to species to ecosystem to landscape. Find patterns and learn processes. *Systems thinking*

Compare with: systems biology, human physiome

Virtual Physiological Human network of excellence

Source: W.Los, modified by A.Hardisty

GAP solution: Workflow paradigm



Show case workflows

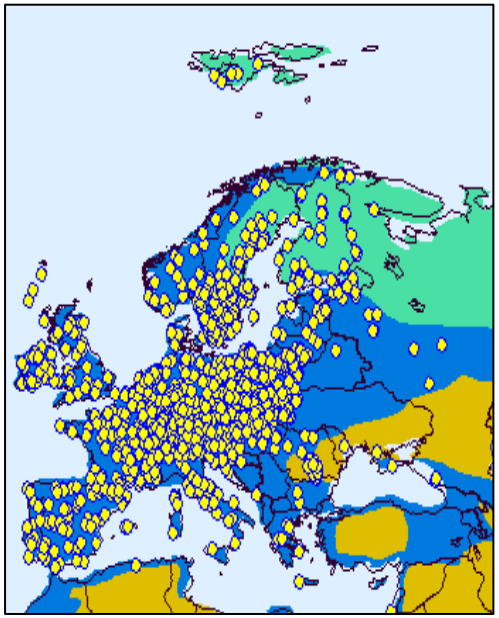
1 Biodiversity Richness Analysis And Conservation Evaluation
2 Biological Valuation Map
3 Automated Retrieval and Analysis of GBIF records
4 Past behaviour and Future Scenarios
5 Bioclimatic Modelling and Global Climate Change
6 Phylogenetic Analysis and Biogeography
7 Ecological Niche Modelling
8 Urban Development and Biodiversity Loss
9 Renewable Energy Planning
10 Hierarchical Scaling of Biodiversity in Lagoon Ecosystems
11 Bird Strike Monitoring
12 Earth Observation

5 challenges (and 5+ solutions)

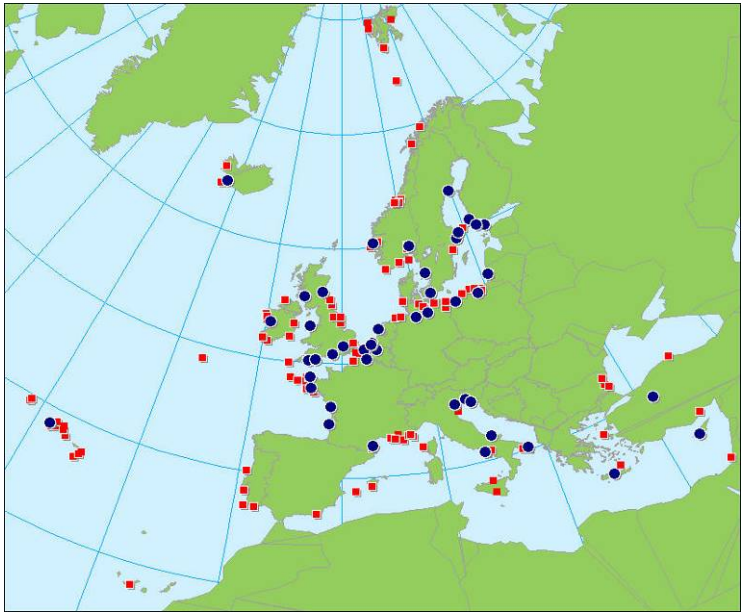
- HETEROGENEITY of the community's requirements, its data resources and tools
- GAP between current practice and future vision
- SCALE of implementation of a pan-European infrastructure, €375m, >25,000 users

Challenge of SCALE: Users and data generators in the large Networks of Excellence

Terrestrial Long-Term Ecological Research (LTER) sites



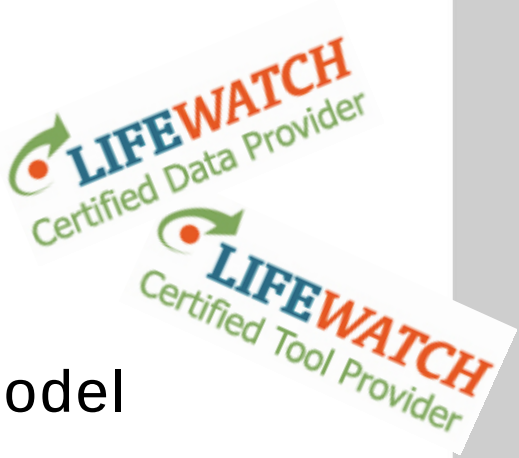
Marine reference and focal sites



Natural science collections

SCALE solution:
Thinking globally, acting locally

- Organisation
  - Top-down financial and legal governance model
  - Project Office
- Technical direction and governance
  - LifeWatch Reference Model
  - Processes to support compliance
  - Bottom-up community governance model
- Core ICT infrastructure
- Management of the product
  - Product Management Board & Release strategy
- Support to the community
  - Service Centre(s), Technical operations support



Core ICT (e-)Infrastructure

- Essential 'central' components
  - Single portal access for all users
  - Datasets & services / tools catalogues (registries)
  - Access to computational resources
  - Security (AAA)
  - Provenance and citation tracking repository
  - Annotations repository
  - Virtual Collaborative Environments / VO / BTCN
  - Workflow composition, execution and management
- Data & tool resources
  - New data resources to be ‘admitted’
  - Statistical, analytical & modelling tools
- Innovation Lab
- Intellectual property management

5 challenges (and 5+ solutions)

- HETEROGENEITY of the community's requirements, its data resources and tools
- GAP between current practice and future vision
- SCALE of implementation of a pan-European infrastructure, €375m, >25,000 users
- PACE of innovation in ICTs

Challenge of PACE: Of innovation in ICT

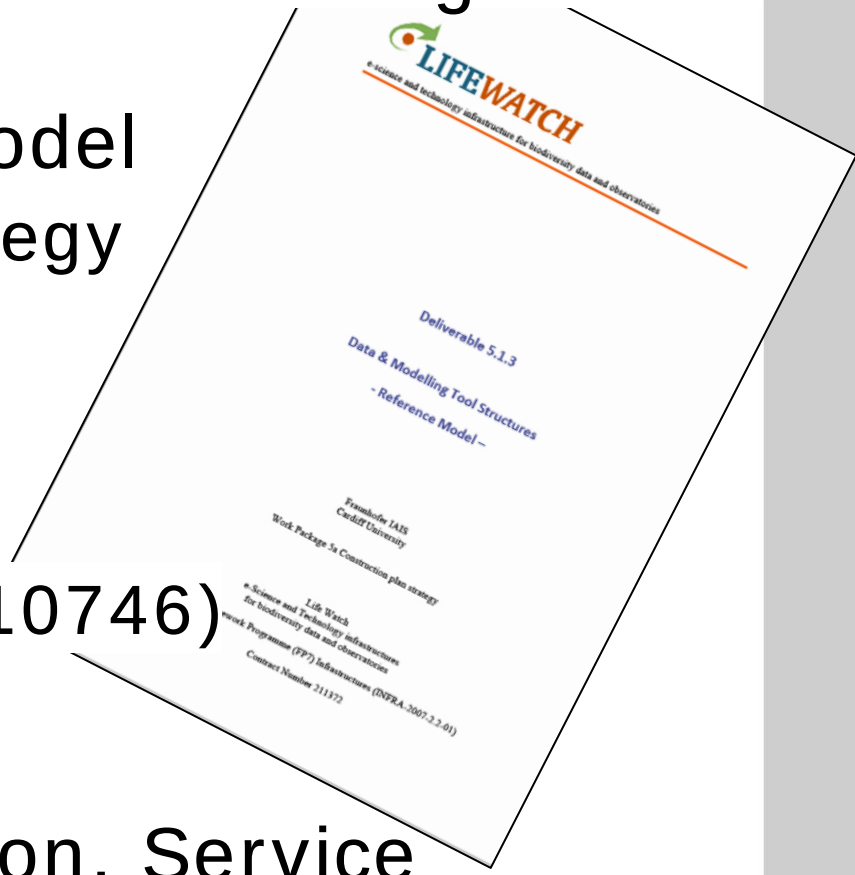
- New technologies, products, services, possibilities, every day
  - Seeing the wood for the trees

- Technology decisions
  - 2 years ago for construction that won't start until next year
  - that have to last for 10 years?

PACE solution:

Divorce functionalities from technologies

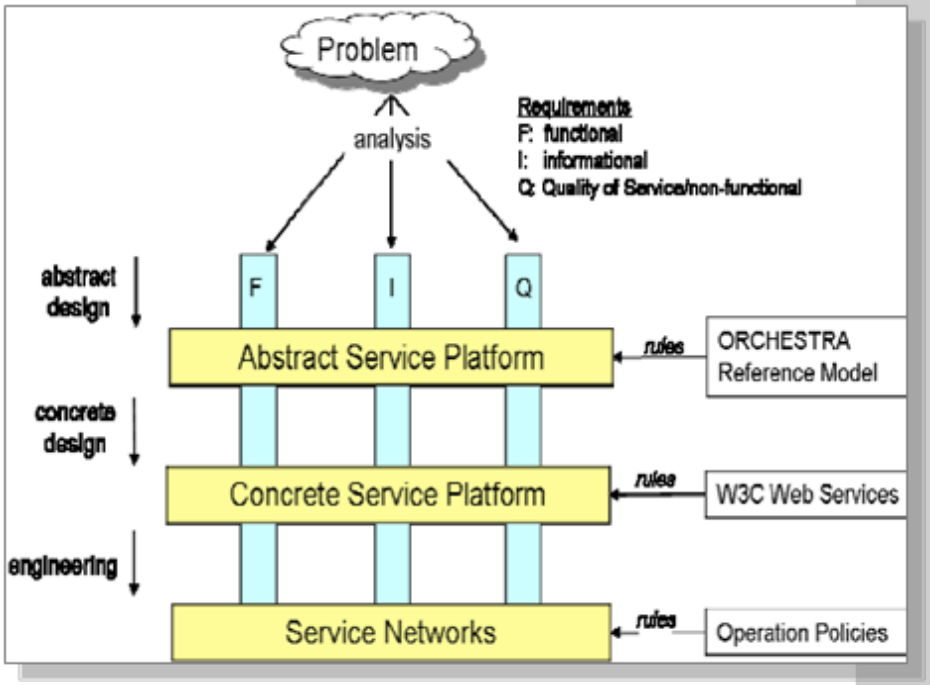
- LifeWatch Reference Model
  - Basis of technical strategy
- Standards-based
  - ORCHESTRA RM
  - OGC RM
  - RM for ODP (ISO/IEC 10746)
- Viewpoints
  - Enterprise, Information, Service
  - Engineering, Technology



The LifeWatch Reference Model ('LifeWatch-RM')

Gives 3 freedoms:

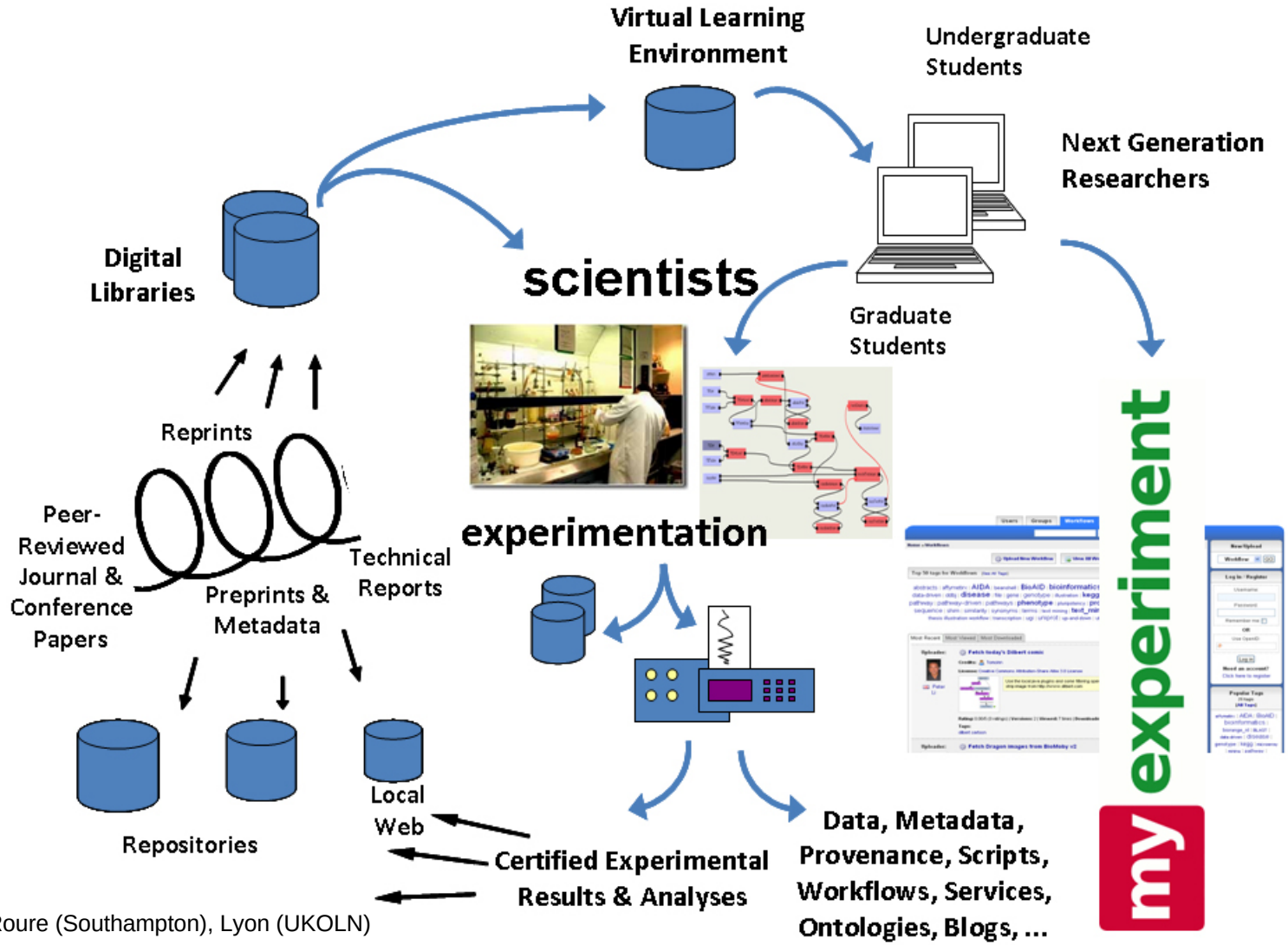
- Technology independence
- Ability to extend technical capabilities
  - Functionalities expressed as services
  - Applications as networks of service instances
- Support for thematic extensions



5 challenges (and 5+ solutions)

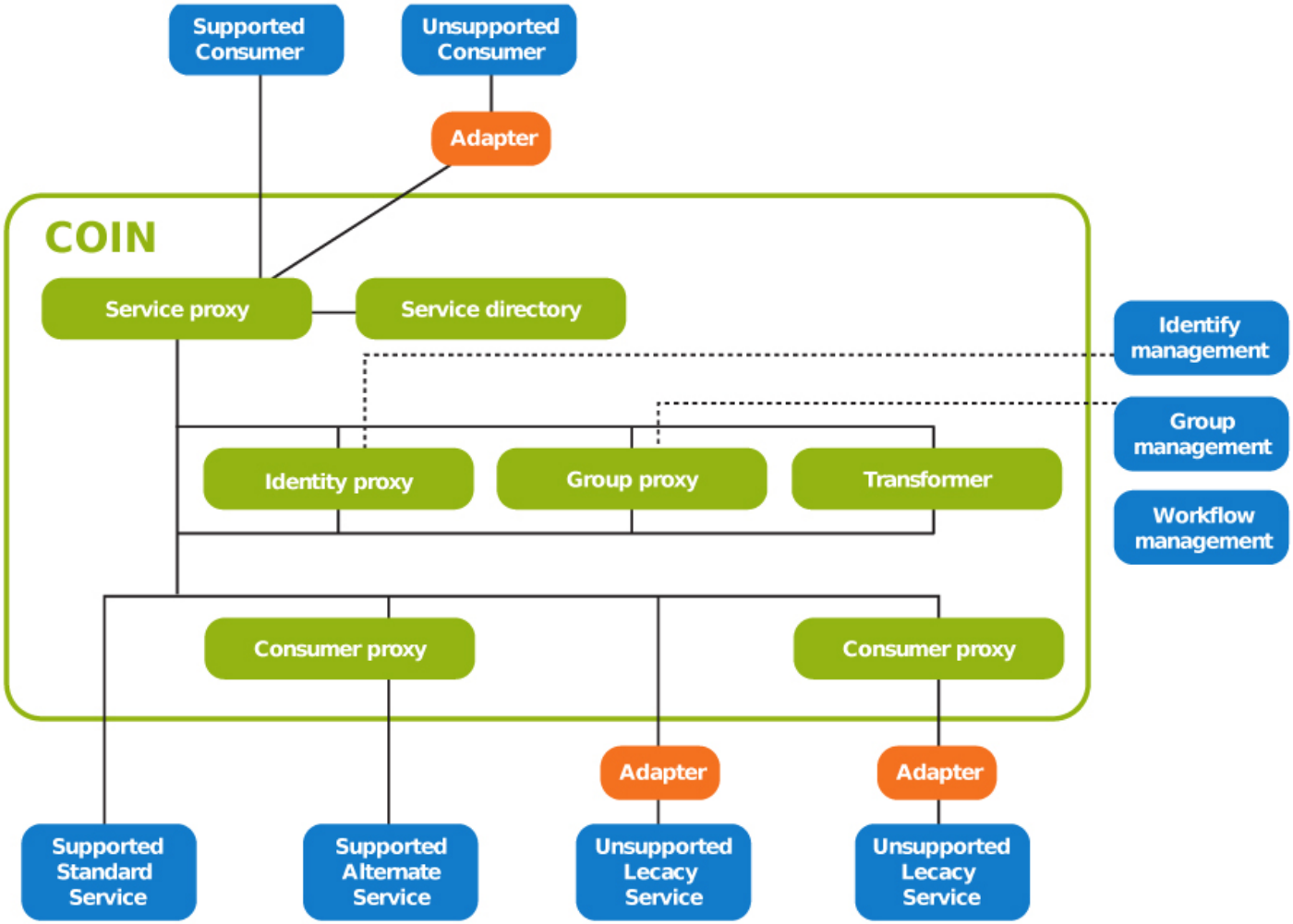
- HETEROGENEITY of the community's requirements, its data resources and tools
- GAP between current practice and future vision
- SCALE of implementation of a pan-European infrastructure, €375m, >25,000 users
- PACE of innovation in ICTs
- FIT with mainstream industry and Higher Education / Research sector directions for ICT service

Challenge of FIT: e-Research lifecycle, Science 2.0



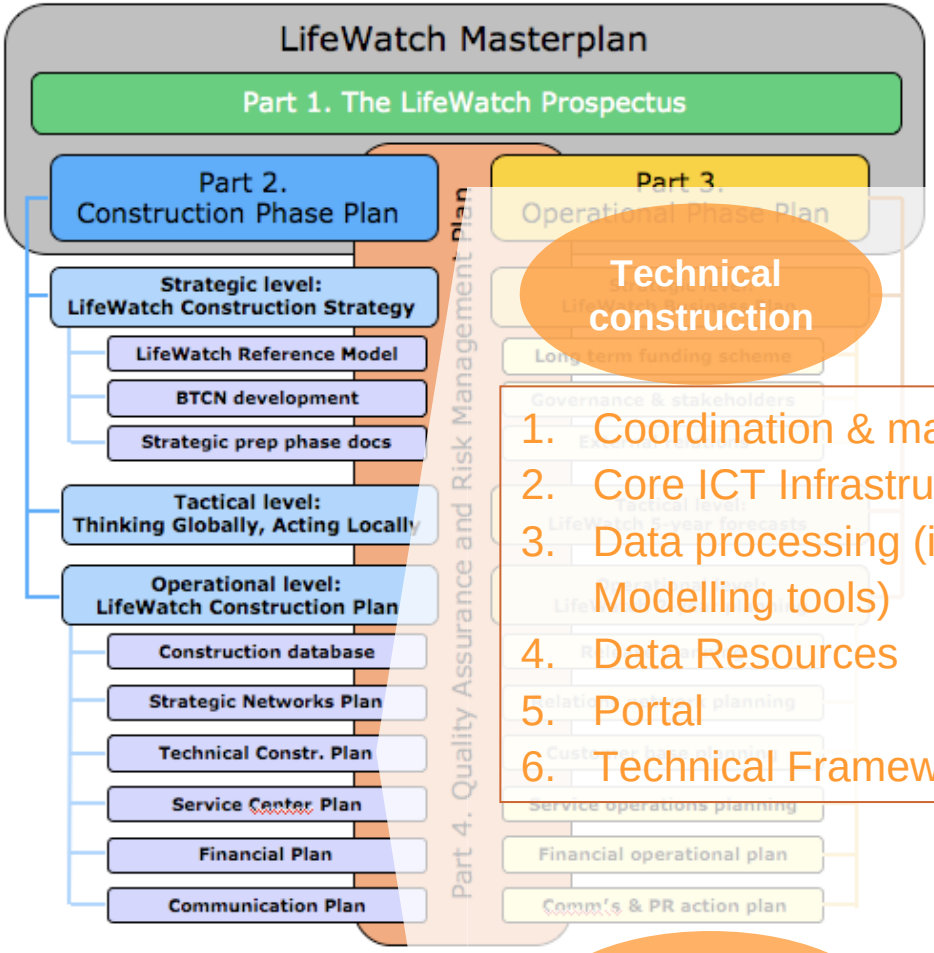
Source: De Roure (Southampton), Lyon (UKOLN)

Challenge of FIT: Collaboration infrastructure



Source: Niels van Dijk, SURFnet, Netherlands

FIT solution: A clear blueprint



"…shows sufficient ambition, but also realism for the next few years"

External reviewers, June 2010

Technical construction

1. Coordination & management; central staff
2. Core ICT Infrastructure
3. Data processing (incl. Analytical & Modelling tools)
4. Data Resources
5. Portal
6. Technical Framework and Architecture

Enabling accelerated and targeted data generation

1. External Data Facilities
2. Marine sites
3. Sensor data resources and
4. Systematics collections
5. Taxonomic backbone

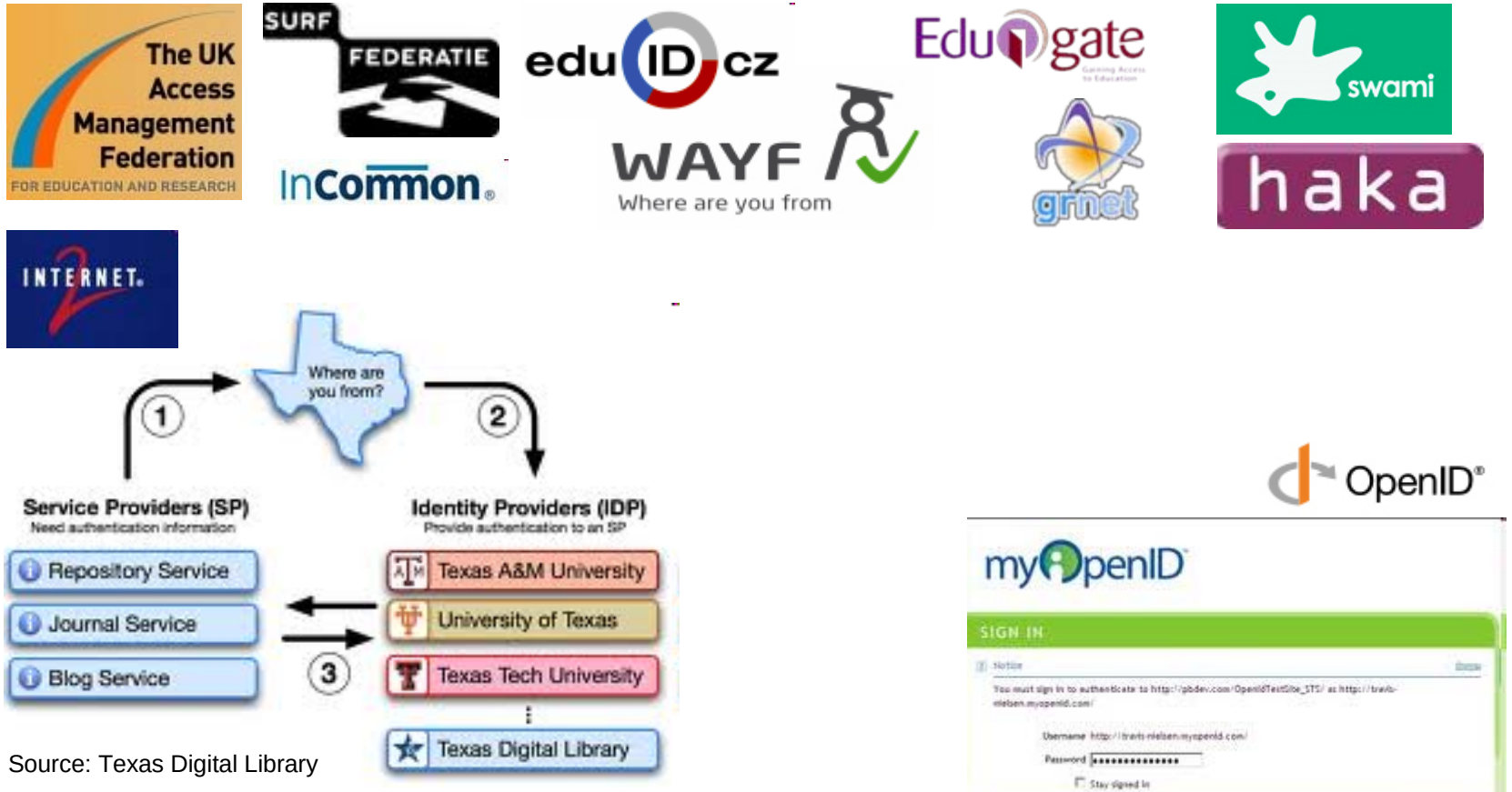
Innovation Lab

Appli ser

1. Core
2. Supp
3. Supp
4. Them

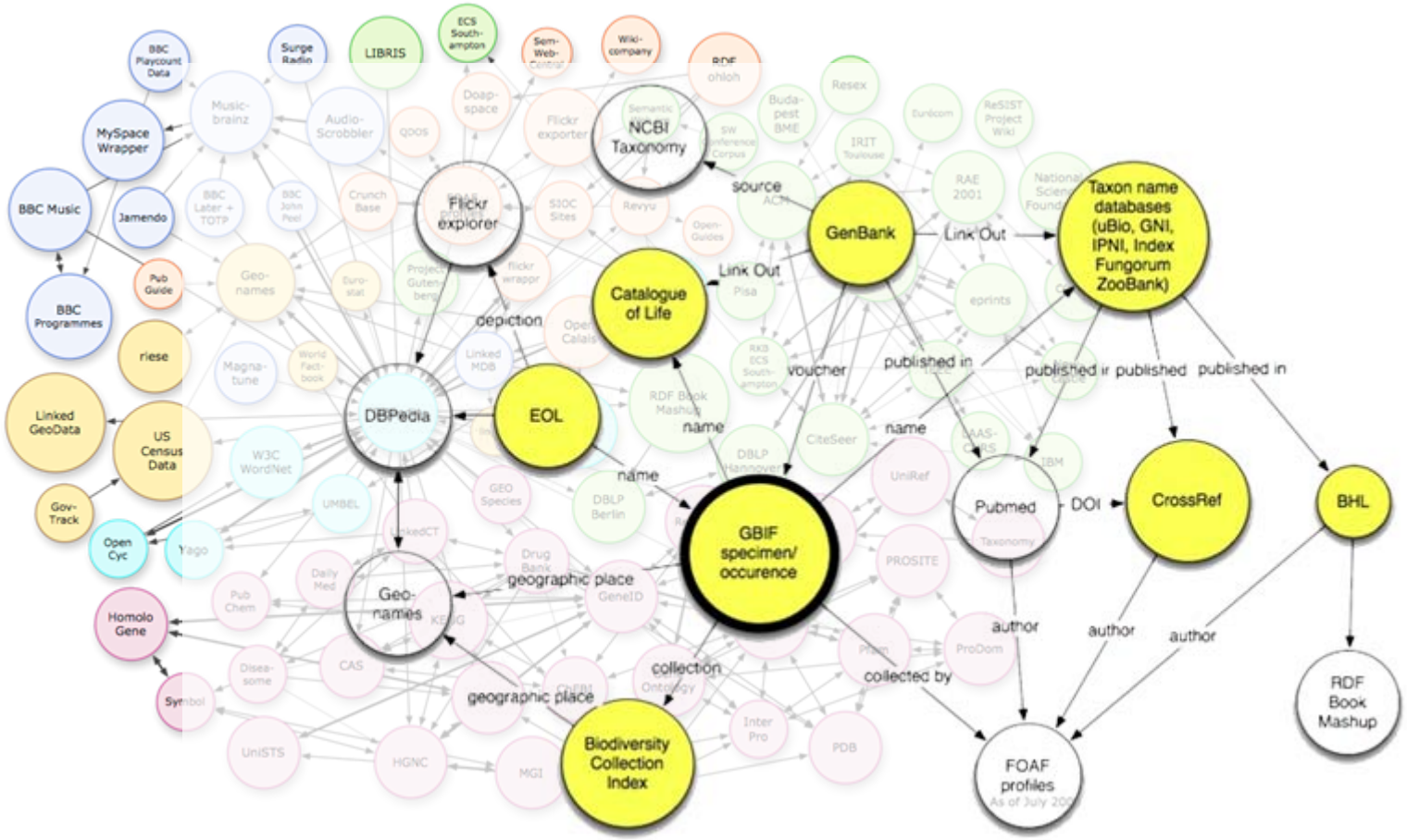
FIT solution: Solutions for authentication

Shibboleth and OpenID, not X.509!



Source: Texas Digital Library

FIT solution: Linked Data



Source: Prof. Rod Page, Glasgow University

August 2010: >19billion triples

http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets/Statistics

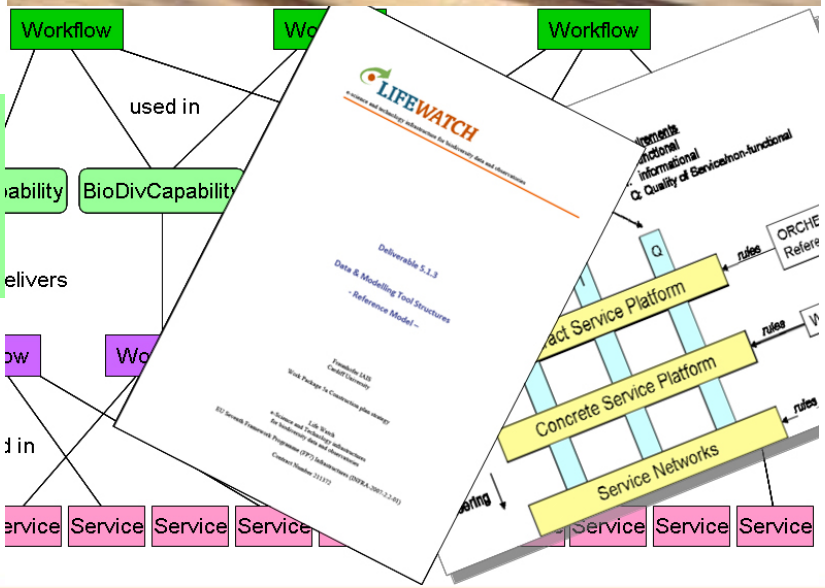
In conclusion

Thinking globally, Acting locally

The mechanism to address the socio-technical challenge of bringing communities together and uniting them behind common technical approaches

Reference model, open standards, composable capabilities

Leads to interoperability and flexibility to accommodate novelty



Thank you

- Questions?

- Acknowledgements
  - LifeWatch colleagues, in particular:
    - Axel Poigné and Vera Hernandez-Ernst, Fraunhofer IAIS, Germany for much of the Reference Model
    - Herbert Schentz, Umweltbundesamt GmbH, Austria for assistance and thinking on semantic interoperability