BIOLOG

Interoperability and networking: common aim, diverse approaches

Walter G. Berendsohn
Dept. for Biodiversity Informatics
Botanic Garden and Botanical Museum Berlin-Dahlem

BIOLOG

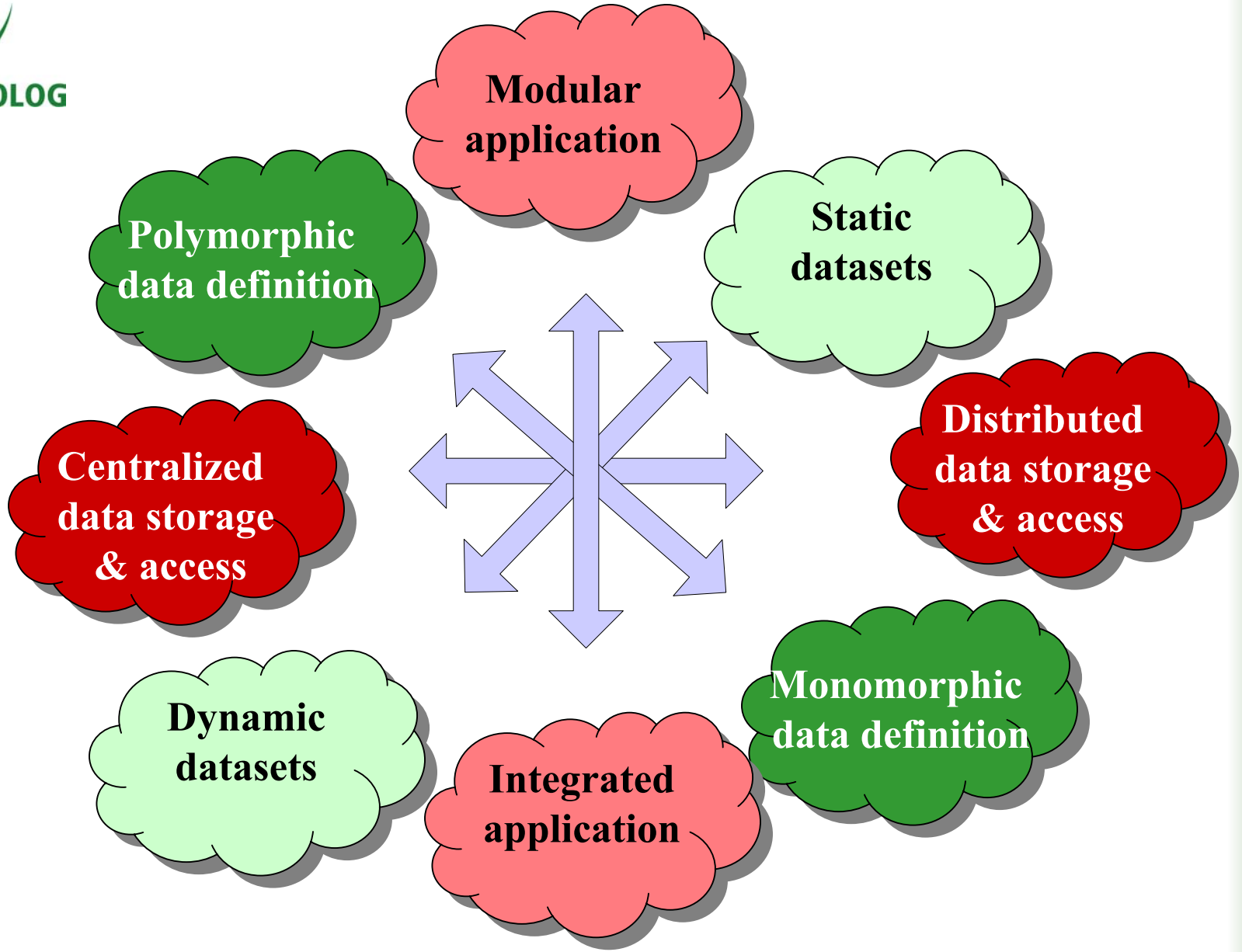
Informatics-diversity

- No uniform approach within BIOLOG
- Great advantage in a rapidly evolving environment (technical, sociological, state of knowledge)
- Diversity of approaches allows for an evolutionary process of adaptation

---

- Incorporation of successful features into an interoperable environment

BIOLOG


Modular application
Polymorphic data definition
Static datasets
Centralized data storage & access
Distributed data storage & access
Dynamic datasets
Integrated application
Monomorphic data definition

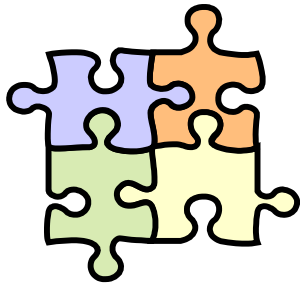
BIOLOG

EDIS / SysTax

- Pragmatic approach:
  - EDIS project databases and SysTax are integrated applications
  - Centralized storage and WWW access
  - Static datasets are imported from dynamic systems (time slice / edition) using a monomorphic import format
- Remote data input possible; centralized SysTax system envisioned to replace EDIS project databases

BIOLOG

Modular application design

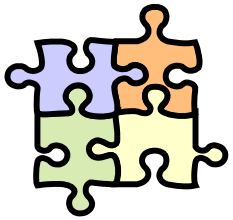


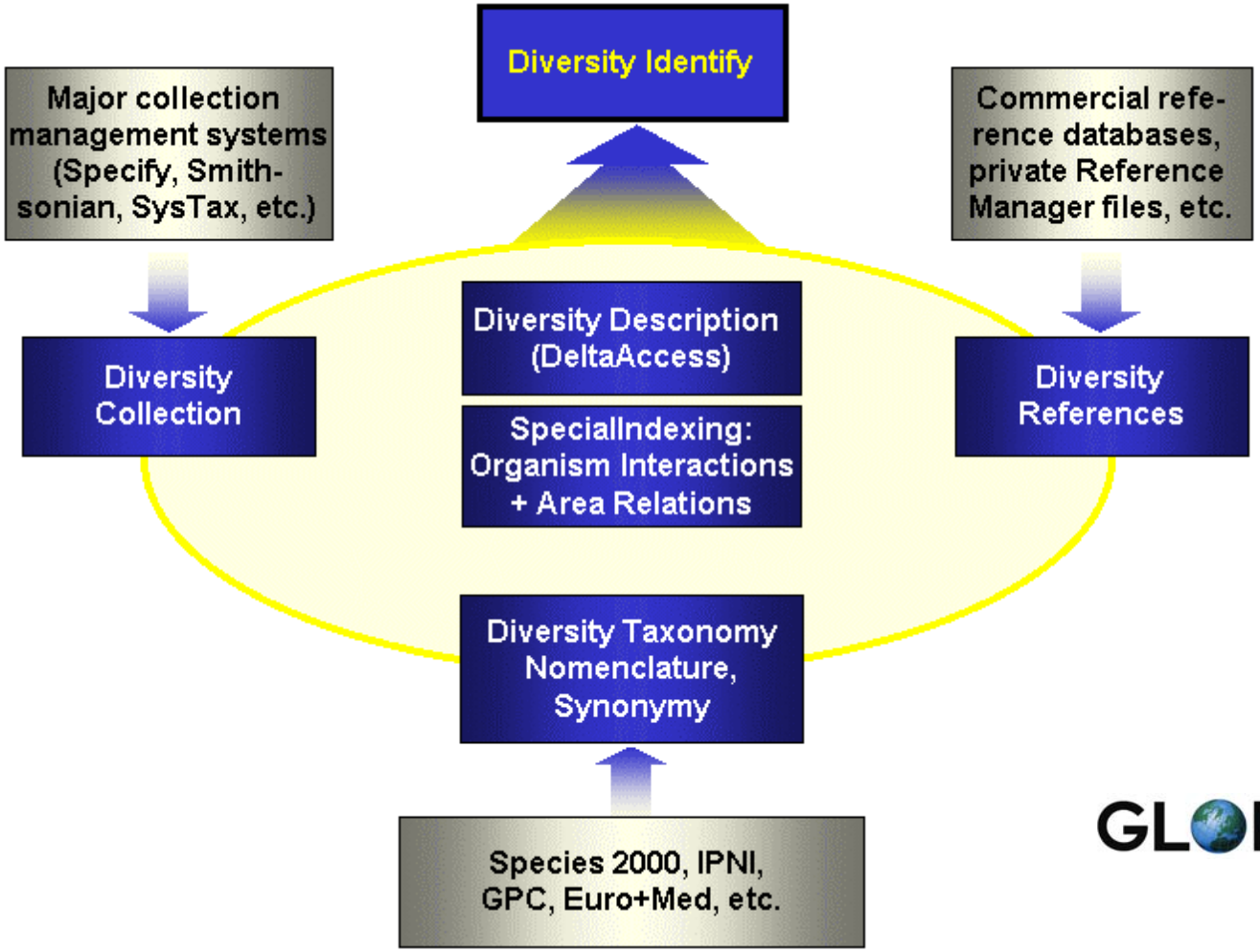
Example in BIOLOG:

DiversityWorkbench (GLOPP-IT)

BIOLOG

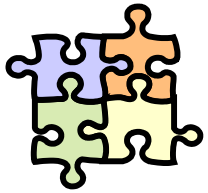
DiversityWorkbench

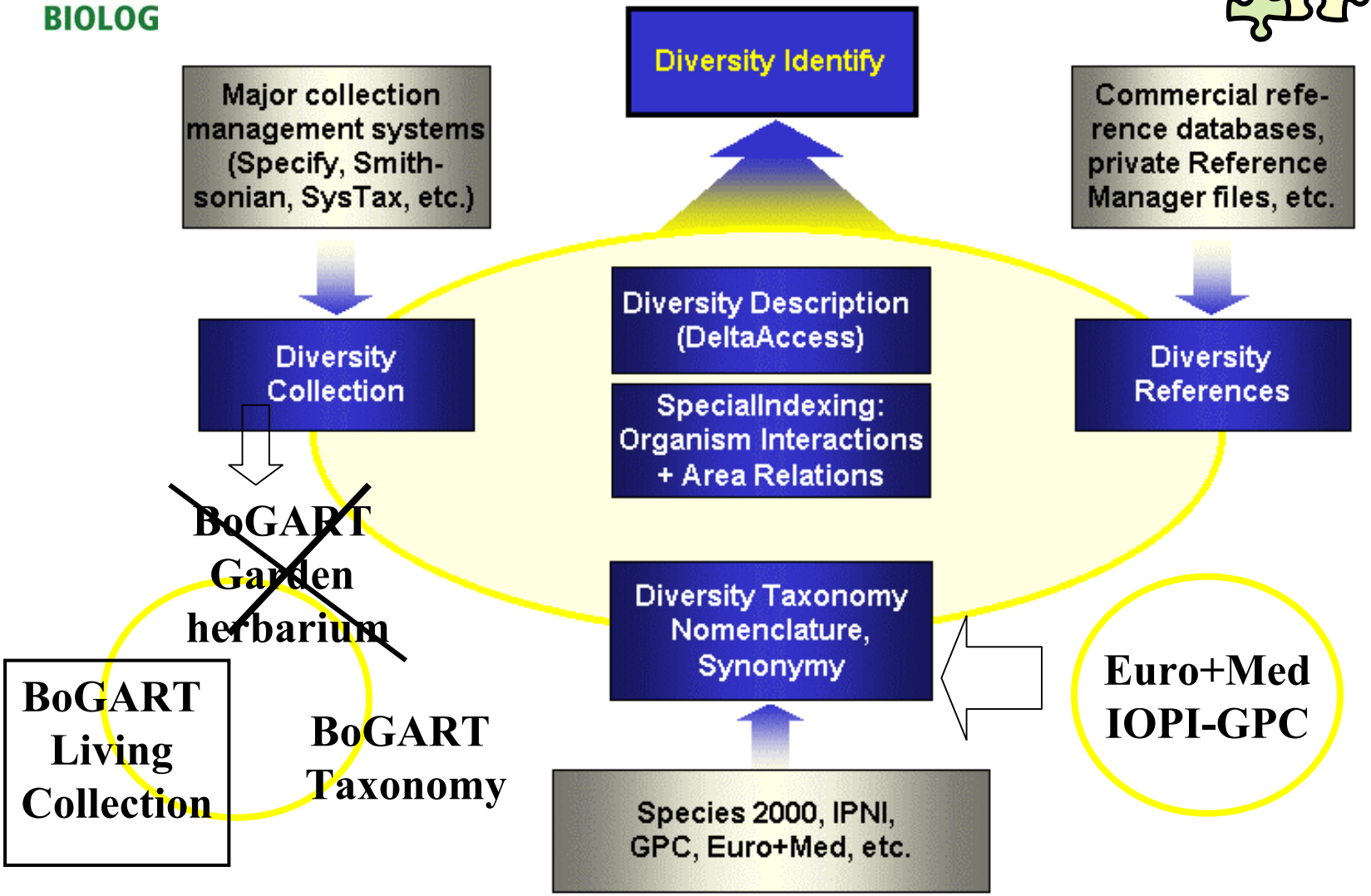




GLOPP

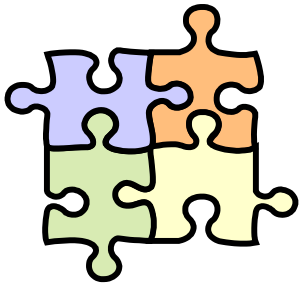
BIOLOG




Diversity Identify
Major collection management systems (Specify, Smithsonian, SysTax, etc.)
Commercial reference databases, private Reference Manager files, etc.
Diversity Description (DeltaAccess)
SpecialIndexing: Organism Interactions + Area Relations
Diversity Collection
Diversity References
BoGART Garden herbarium
BoGART Living Collection
BoGART Taxonomy
Diversity Taxonomy Nomenclature, Synonymy
Euro+Med IOPI-GPC
Species 2000, IPNI, GPC, Euro+Med, etc.

BIOLOG

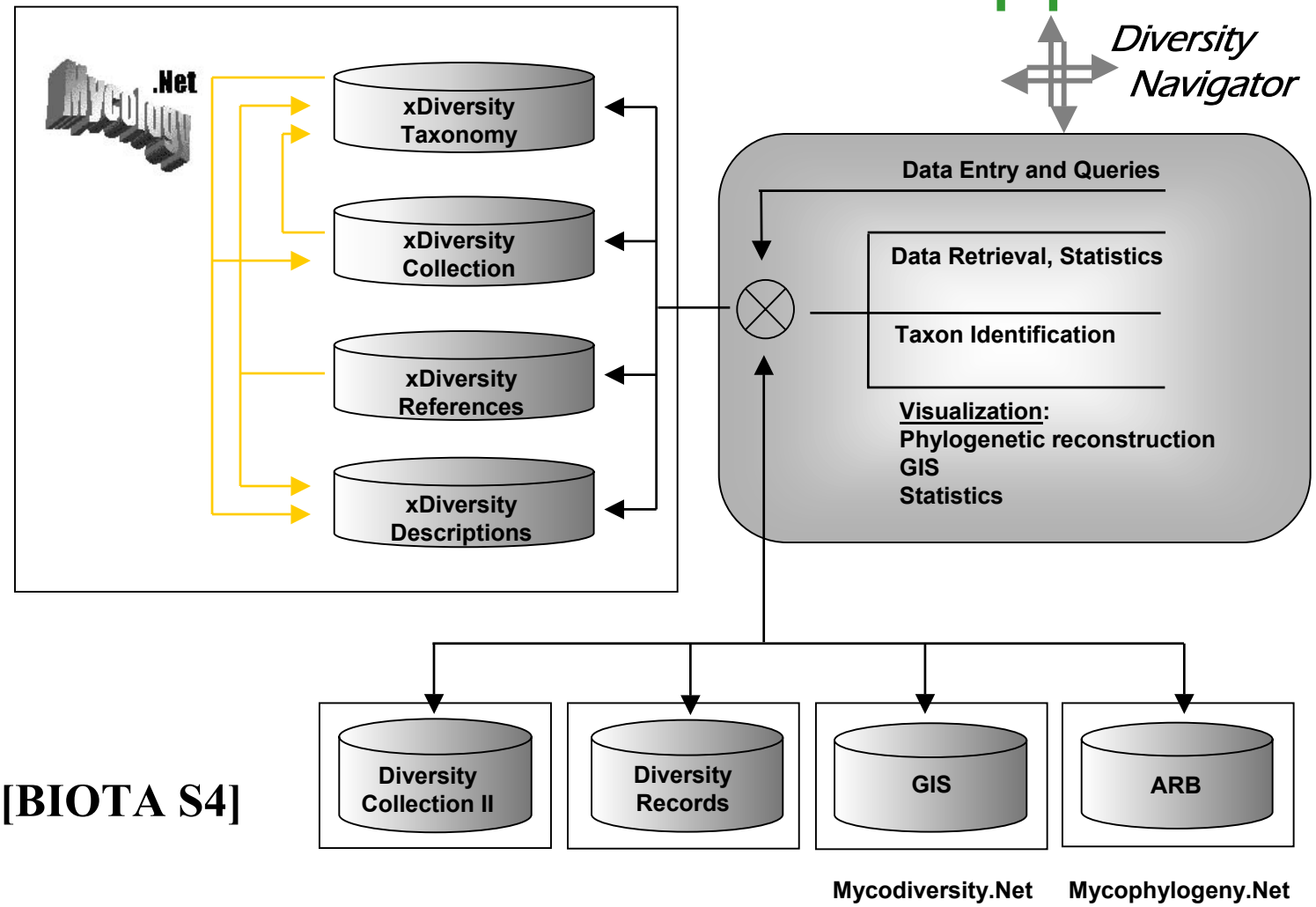
Modular application design



- Interoperability layers in local applications
  - Complementary database modules
  - Complementary database modules with “business logic”
  - Complementary modules with user interface
- Distributed applications
  - Web services
  - SOAP etc.

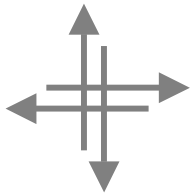
BIOLOG

Modular design -> distributed applications



[BIOTA S4]

BIOLOG



Diversity Navigator

- Java interface to biodiversity data
- Data model: DiversityWorkbench; OS: Linux; DBMS: Postgres
- Web Services: Soap (simple object access protocol) and remote procedure calls
- Monomorphic data definition; distributed

[Biota S4: U. Bayreuth, G. Rambold]

Common access as a first step towards interoperability

Example: biological collections

- BIOLOG project: ZEFOD
- EU projects:
  - ENHSIN
  - BioCASE
  - ENBI
- International collaboration:
  - TDWG/CODATA working group on biological collection data access connects several initiatives

BIOLOG

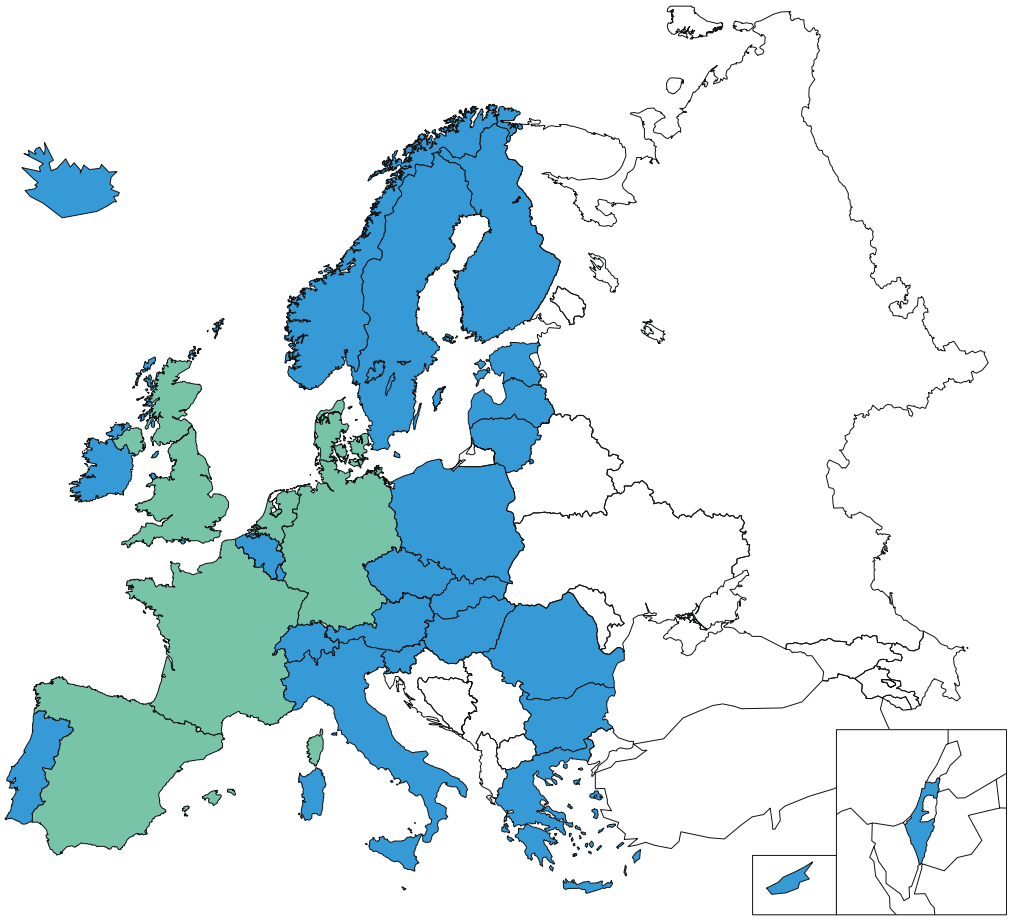
Access to European collections

ENHSIN

European Natural History Specimen Information Network

BioCASE

Biological collection Access Service for Europe



BIOLOG

ENHSIN/BioCASE concepts

- Distributed data storage
- Monomorphic metadata
- Modular applications
- Common distributed thesauri
- Static and dynamic datasets
- Polymorphic unit data
- Conforms to TDWG/CODATA standard

BIOLOG

Interoperability standard

- TDWG/CODATA working group on biological collection data access
- Two workshops in 2001, with participants from BIOLOG projects
- Content definition subgroup
- Protocol development subgroup
- www.bgbm.org/TDWG/CODATA/

BIOLOG

Interoperability standard

Content definition subgroup

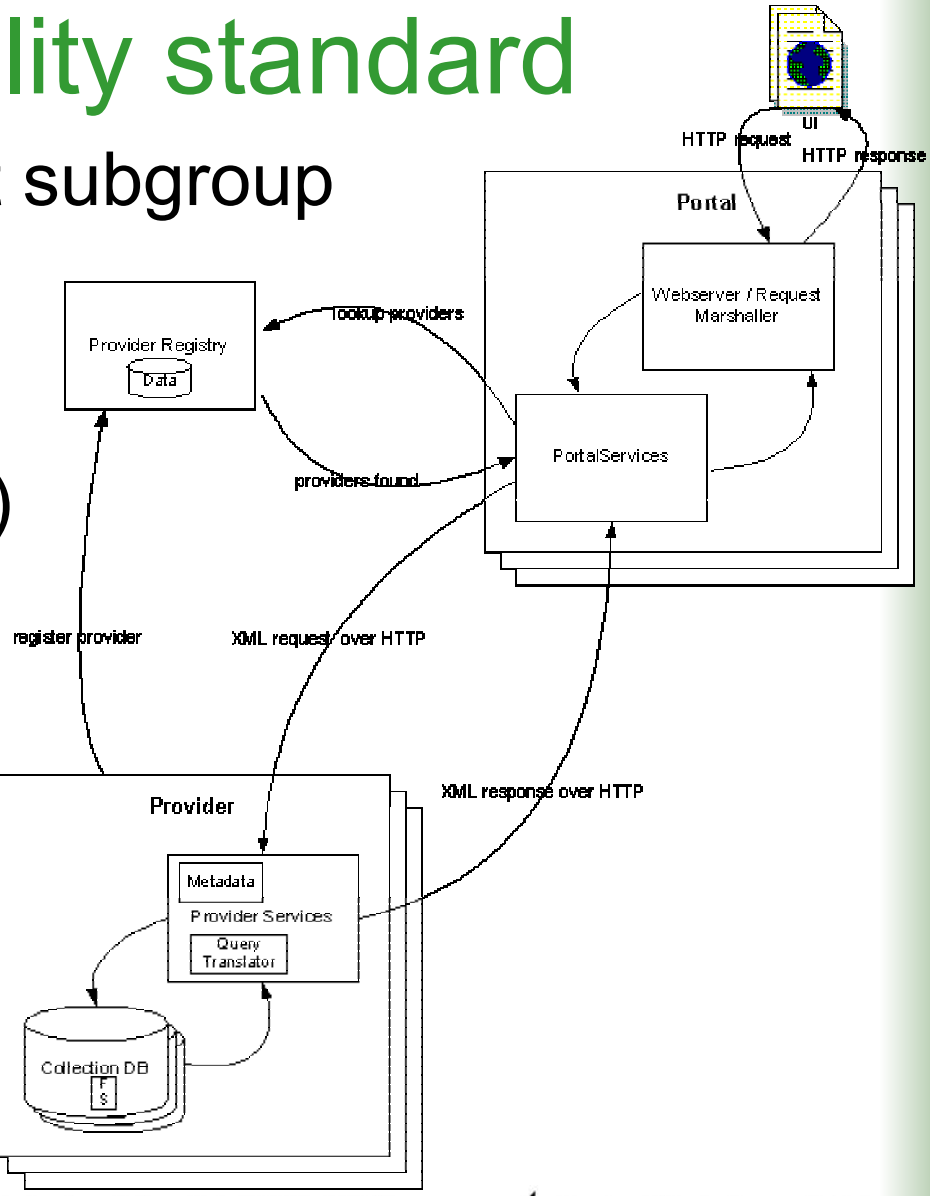
- Develop a “federation scheme”, i. e. an XML schema where data elements are described (a.k.a. data dictionary)
- Based on published standards and information models used (HISPID, ITF, ASC, BioCISE, UK-Recorder, ..)
- RFC (request for comment) starting now

BIOLOG

Interoperability standard

Protocol development subgroup

- DiGIR Project (Distributed Generic Information Retrieval)
- Aim:
  to define a protocol for retrieving structured data from multiple, heterogeneous databases



BIOLOG

Interoperability standard

Basic architecture of protocol

- Query and results based on federation schema
- Wrapper on provider's side with metadata and query translator; understands XML query and produces XML output
- Providers registered at a web service
- Portal services receive queries (from user interface or other services), checks provider registry, checks metadata, and sends information request to appropriate providers

BIOLOG

The international context

- Individual projects well connected internationally
- Many use internationally available information resources (e.g. lists of taxa)
- Several directly contribute to emerging global information systems
- Projects are building resources for GBIF