Tool & Data Services Registry EDAM and SWO

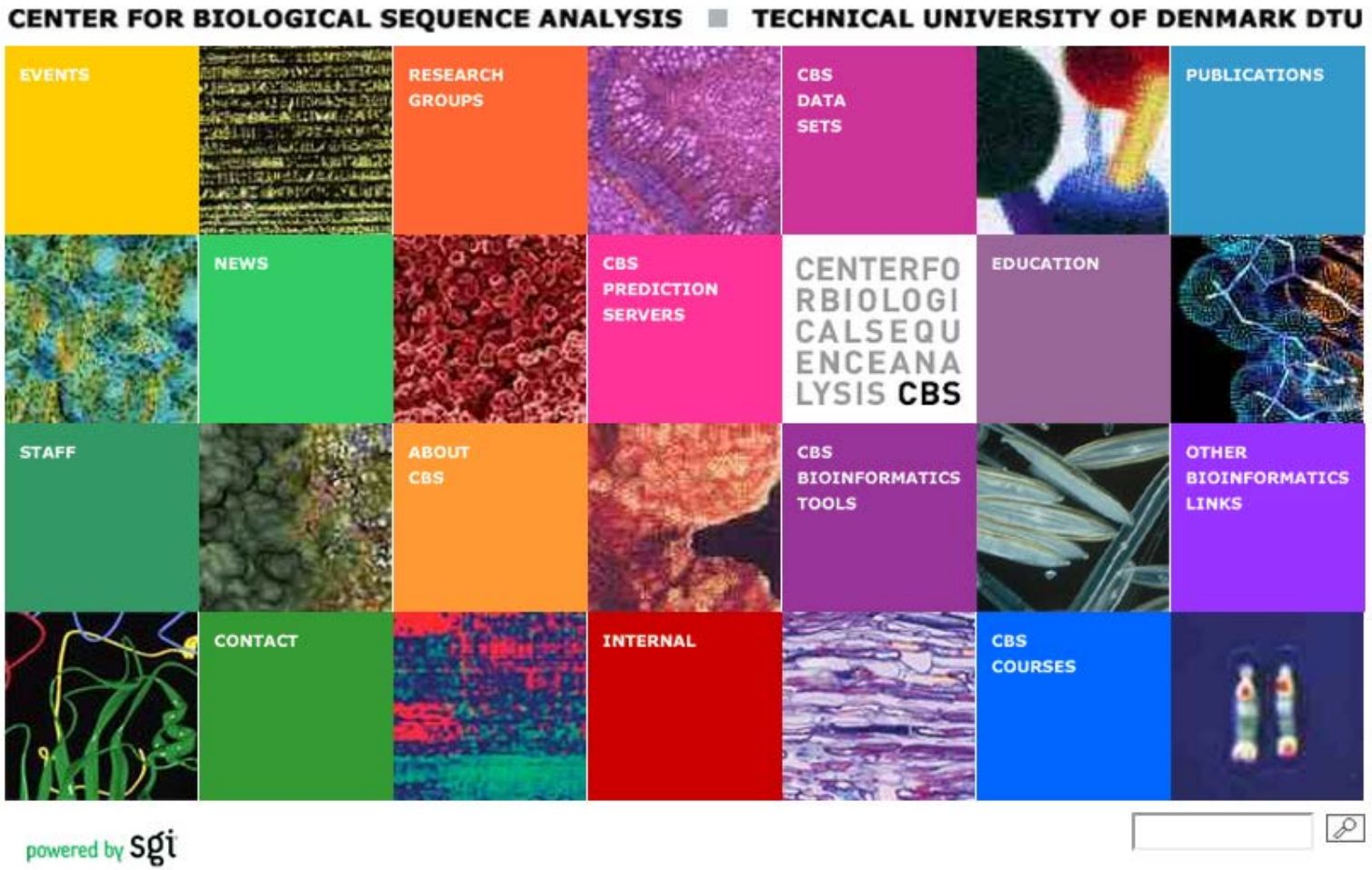
Jon Ison PhD (ELIXIR-DK)

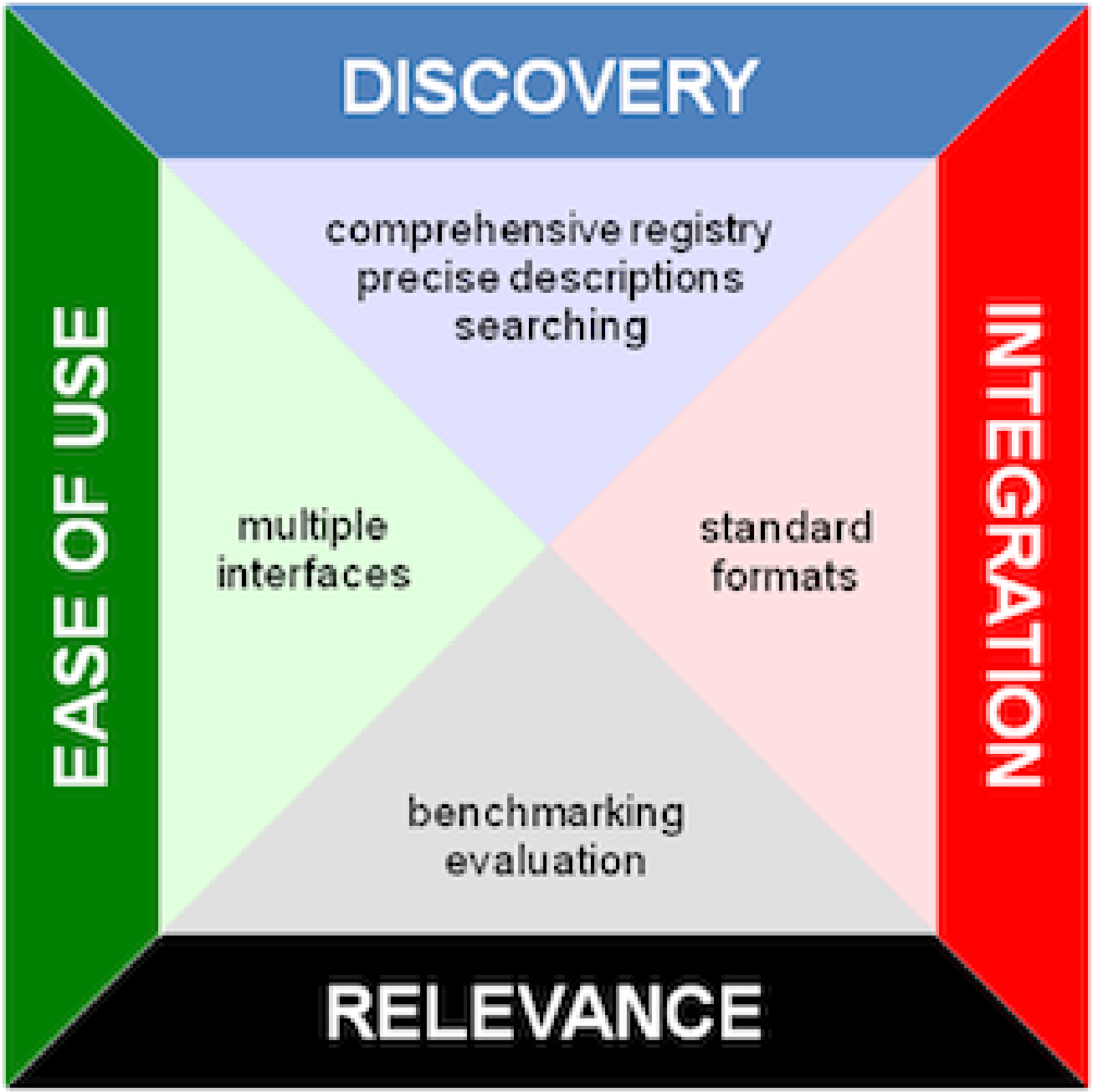
EUON 2014, Amsterdam

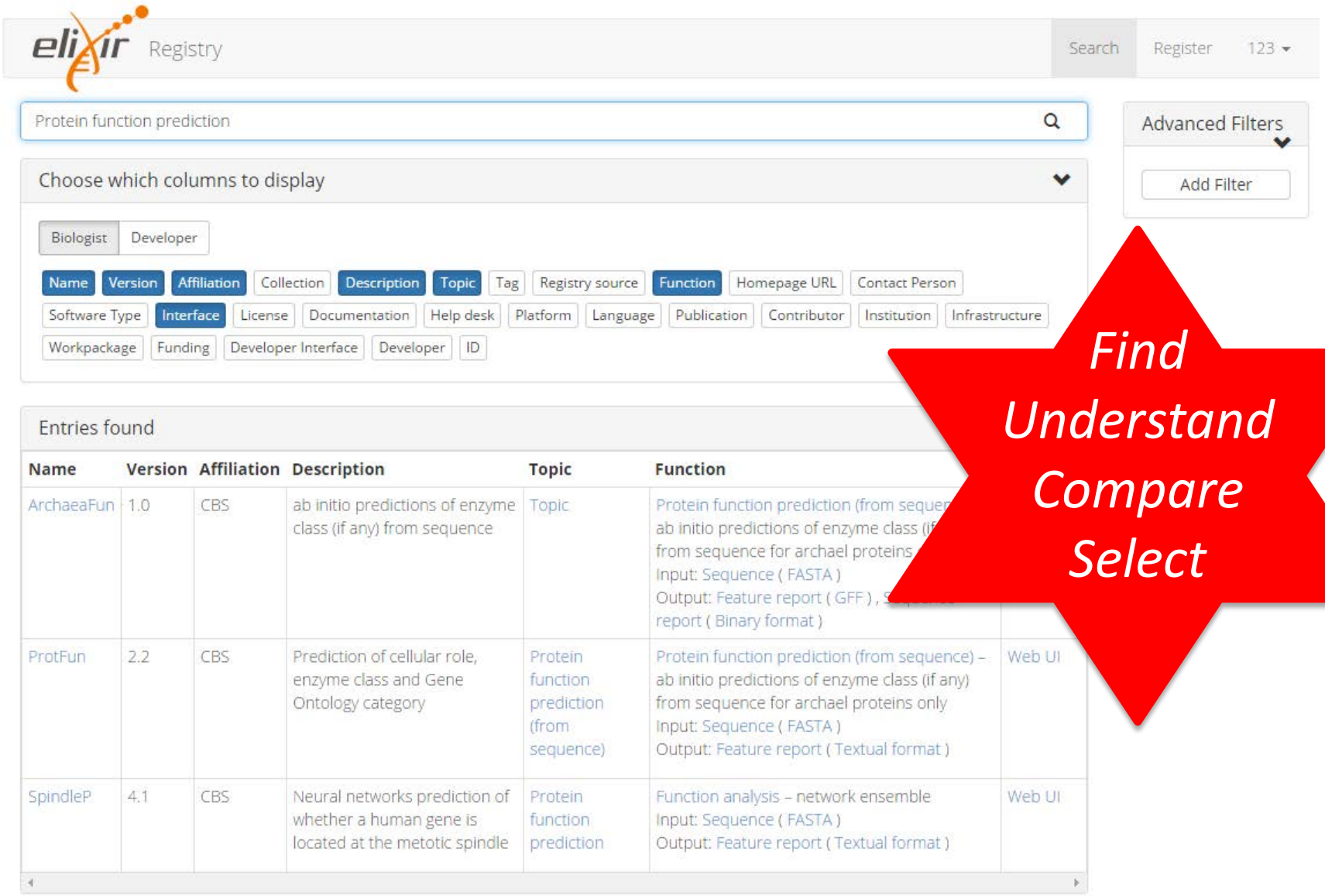

EU-OPENSCREEN
Chemical keys for life's locks
elixir
BBMRI
Biobanking and
Biomolecular
Resources Research
Infrastructure
instruct
Integrating
Biology
erinha
European Research Infrastructure
on Highly Pathogenic Agents
BioMedBridges
EMBRC
EUROPEAN
MARINE
BIOLOGICAL
RESOURCE
CENTRE
ECRIN
EUROPEAN CLINICAL RESEARCH INFRASTRUCTURES NETWORK
EATRIS
European Infrastructure for
Translational Medicine
EURO-BIOIMAGING


EMBL
elixir

ELIXIR-DK : The Tools Node




DISCOVERY
EASE OF USE
INTEGRATION
RELEVANCE
comprehensive registry
precise descriptions
searching
multiple
interfaces
standard
formats
benchmarking
evaluation


elixir Registry
Search
Register
123
Protein function prediction
Advanced Filters
Add Filter
Choose which columns to display
Biologist
Developer
Name
Version
Affiliation
Collection
Description
Topic
Tag
Registry source
Function
Homepage URL
Contact Person
Software Type
Interface
License
Documentation
Help desk
Platform
Language
Publication
Contributor
Institution
Infrastructure
Workpackage
Funding
Developer Interface
Developer
ID
Find
Understand
Compare
Select
Entries found
Name
Version
Affiliation
Description
Topic
Function
ArchaeaFun
1.0
CBS
ab initio predictions of enzyme class (if any) from sequence
Topic
Protein function prediction (from sequen
ab initio predictions of enzyme class (if
from sequence for archael proteins
Input: Sequence (FASTA)
Output: Feature report (GFF) ,
report (Binary format)
ProtFun
2.2
CBS
Prediction of cellular role, enzyme class and Gene Ontology category
Protein function prediction (from sequence)
Protein function prediction (from sequence) – ab initio predictions of enzyme class (if any) from sequence for archael proteins only
Input: Sequence (FASTA)
Output: Feature report (Textual format)
Web UI
SpindleP
4.1
CBS
Neural networks prediction of whether a human gene is located at the metotic spindle
Protein function prediction
Function analysis – network ensemble
Input: Sequence (FASTA)
Output: Feature report (Textual format)
Web UI

Rationale

- Software provision is highly distributed
  - large service organisations to individual scientists
- Little coordination of scientific scope, description, use or interoperability
  - within / between research infrastructures
- Software landscape is fragmented !
  - diverse offerings
  - scientists must trawl Web or ask colleagues

Rationale

- Software provision is hi istributed
  - large service organisa dividual scientists
- Little coor ription, use or interoper
  - within a rese
  - between R
- Software
  - diverse offerings
  - scientists must trawl t or ask colleagues to comprehend it

Comprehensive, consistent and searchable registry

!! Boost tool utilization !!!

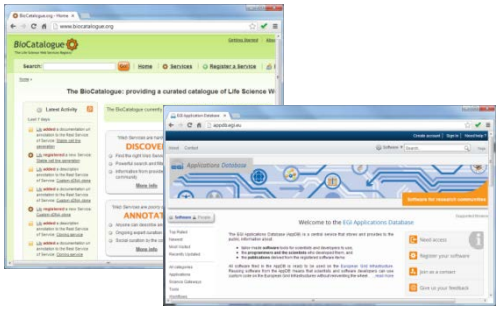
Institutional providers

EBI, SIB, DTU …

SIB Swiss Institute of Bioinformatics

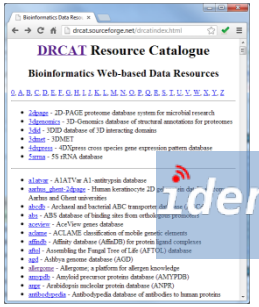
Existing registries

AppDB, BioCatalogue …



Databases

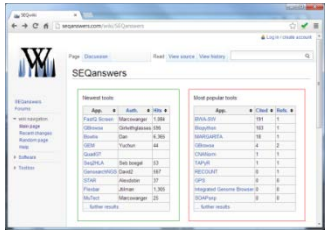
EBI, Identifiers.org, DRCAT …



identifiers.org

WIKIs

SeqWiki …



Software suites

CCP4, EMBOSS …

CCP4

emboss

Other catalogues & lists

CCISB, AllBio …

instruct Integrating Biology

ALLBIO BIOINFORMATICS

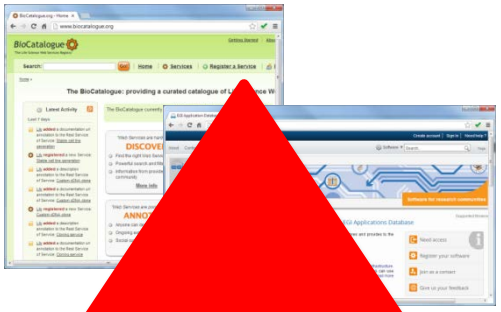
Institutional providers

EBI, SIB, DTU …

SIB Swiss Institute of Bioinformatics

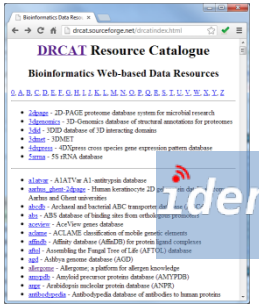
Existing registries

AppDB, BioCatalogue …



Databases

EBI, Identifiers.org, DRCAT …

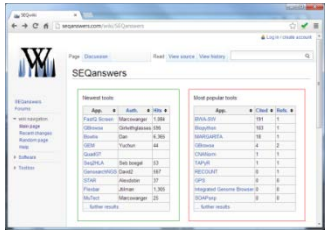


identifiers.org

Federated curation model

WIKIs

SeqWiki …



Software suites

CCP4, EMBOSS …

CCP4

emboss

Other ca… …ts

CCISB, AllBio …

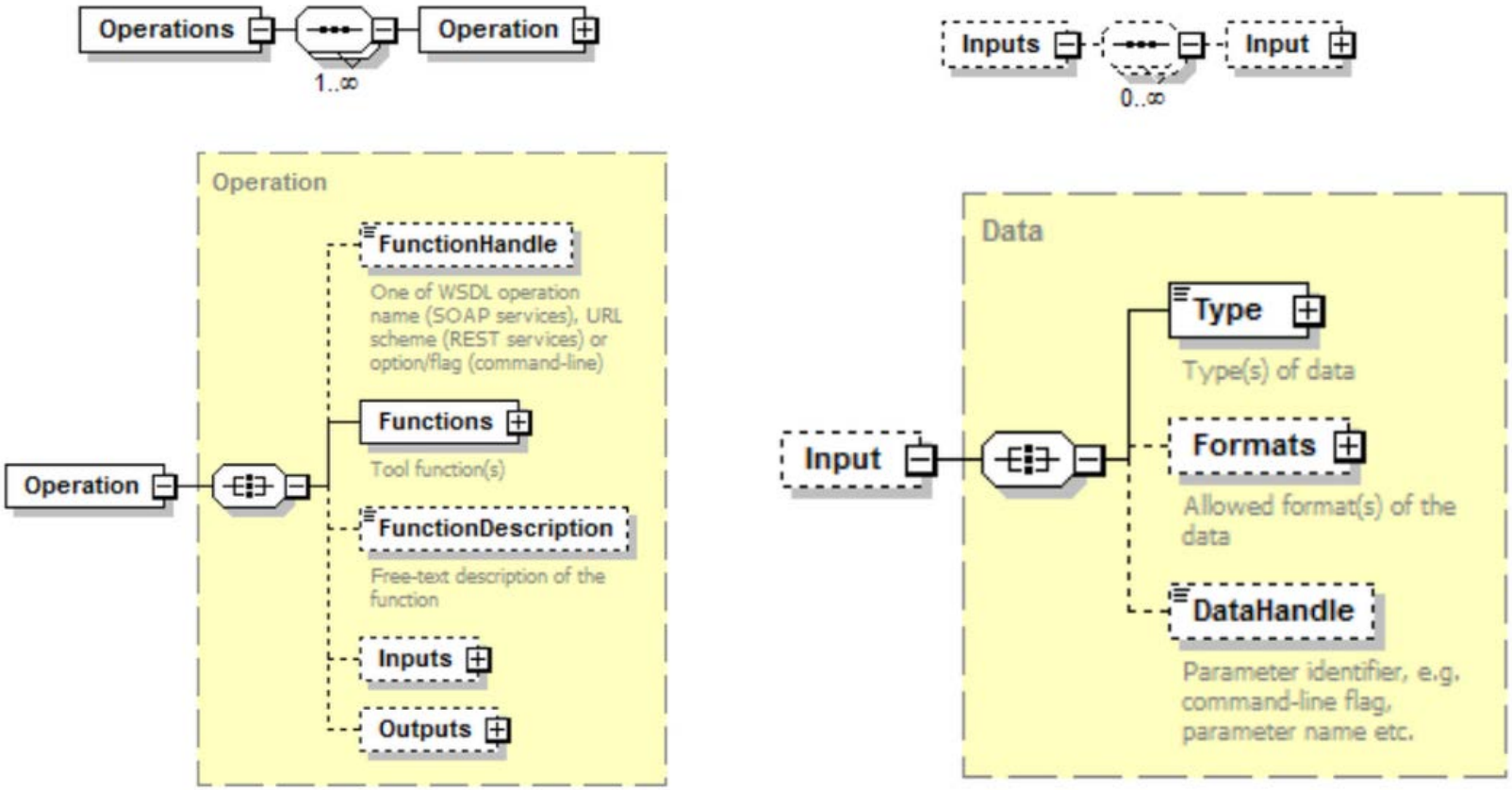
instruct Integrating Biology

ALLBIO BIOINFORMATICS

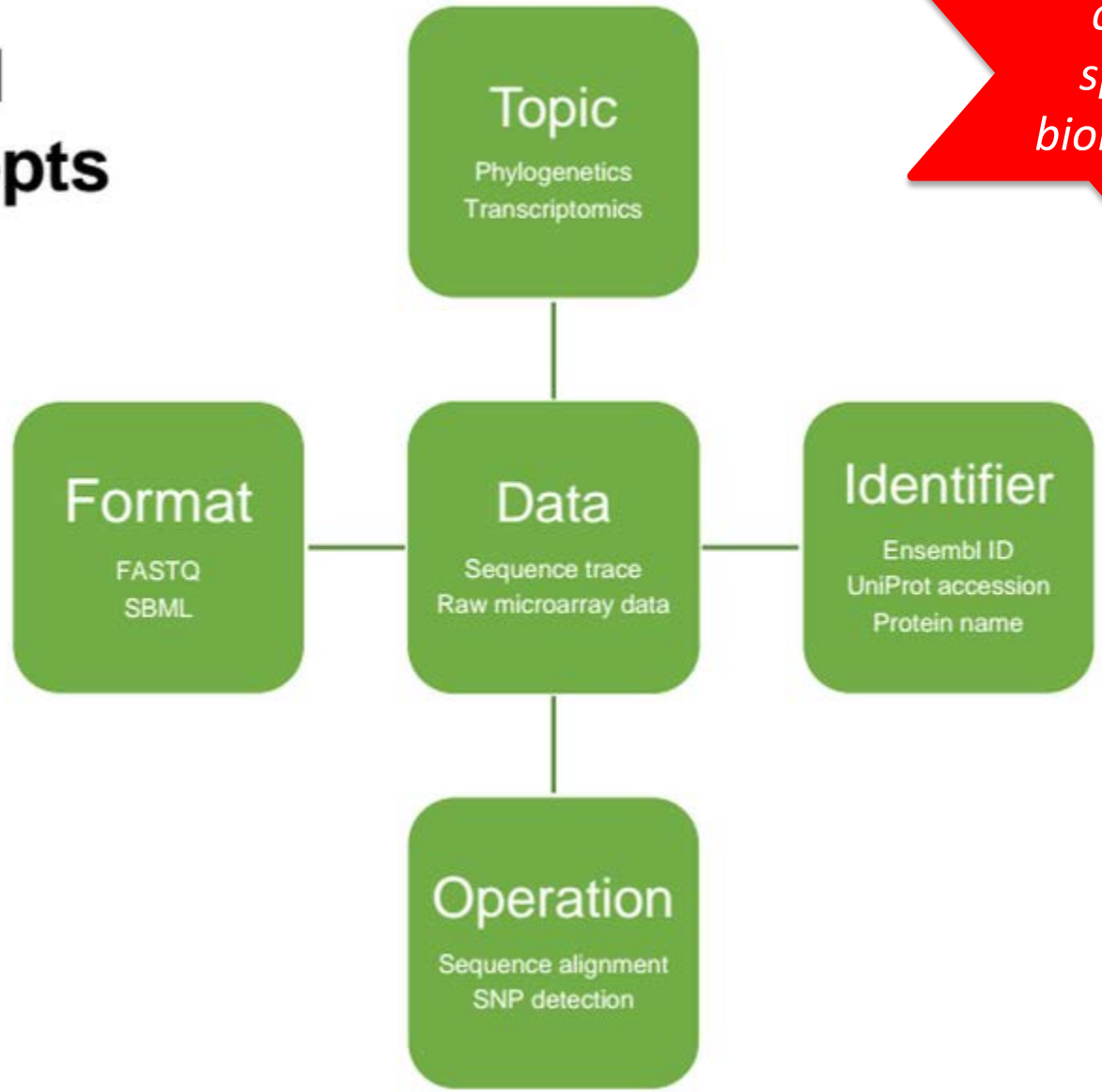
Role of ontologies

- Concise, precise and ***consistent*** information
  - support resource discovery
- Practical scientific, technical and administrative details
  - what a tool does
  - how to access and use it
- Highly streamlined & usable UI
  - specify exactly what you want
  - get a tailored list of suitable resources

Resource description model

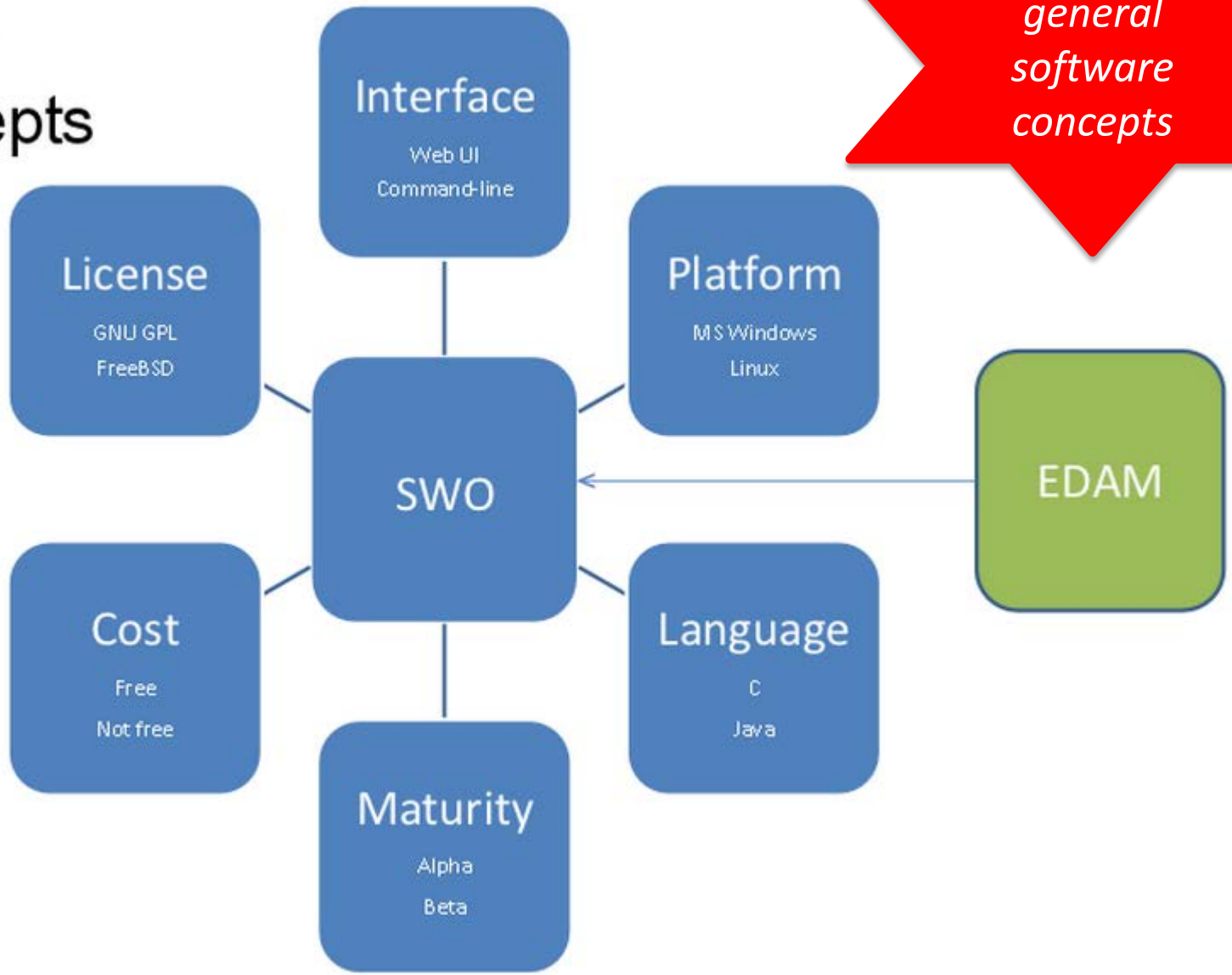


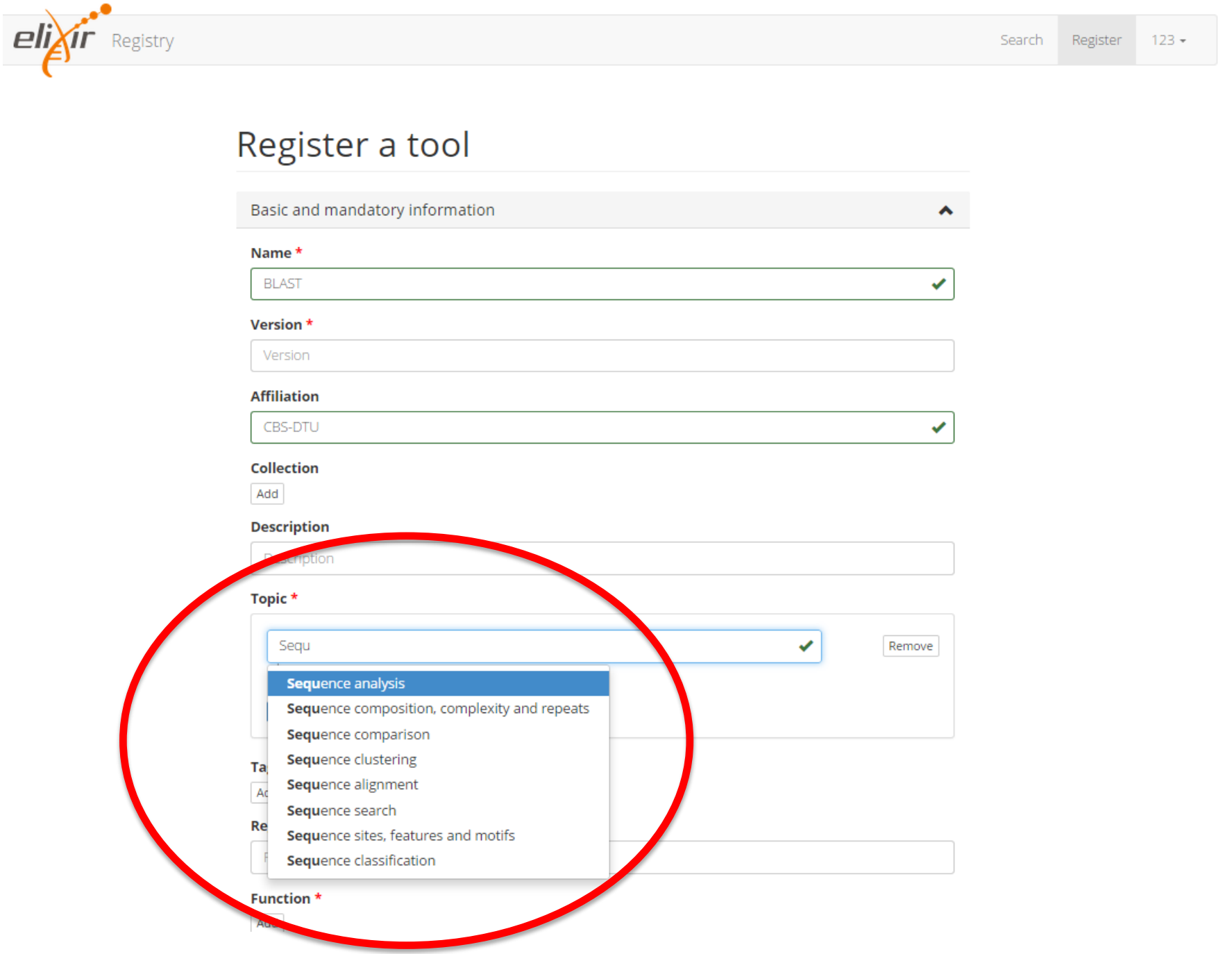
EDAM concepts



concepts specific to bioinformatics

SWO concepts




elixir Registry
Search
Register
123
Register a tool
Basic and mandatory information
Name *
BLAST
Version *
Version
Affiliation
CBS-DTU
Collection
Add
Description
Topic *
Sequ
Remove
Sequence analysis
Sequence composition, complexity and repeats
Sequence comparison
Sequence clustering
Sequence alignment
Sequence search
Sequence sites, features and motifs
Sequence classification
Function *

tinyurl.com/TopicsOutline

BioMedBridges

HOME AIMS NEWS WORK PACKAGES CONTACT LOG IN

Search

Workshop: A common vocabulary to classify resources in the life science domain

What

A one day workshop organised by ELIXIR, BioMedBridges and RDA to kick-start agreement on a common topics vocabulary to classify resources in the life sciences domain, including databases, tools, courses, training materials, meetings, jobs and publications.

Registration is now closed - if you would still like to attend, please contact Joy Friesner.

Why

Service providers and registries use different vocabularies to classify resources. From the user point of view, a common topics vocabulary would enhance the ability to find related resources of interest. From the provider point of view, a common vocabulary would be an important step to facilitate the interoperability of existing resources.

Who

We welcome representatives of societies, networks, institutes, organisations, research infrastructures and projects interested in using a common vocabulary to classify resources in the life science domain. We will involve and welcome domain specialists and ontologists to assist the discussion.

tinyurl.com/TopicsOutline

When and where

Thursday, 16 October 2014: one day before the collocated RDA Europe Workshop about Data Sharing and Interoperability.

Strasbourg + Luxembourg rooms, Science 14 Atrium, Rue de la Science, 14b, Brussels

Hotel: For those who are receiving support from ELIXIR to cover accommodation for the night of 16th October, accommodation will be arranged at: IBIS Brussels Centre Ste Catherine (Joseph Plateaustraat 2, 1000 Brussels).

Outcomes

During the workshop we will:

- *Agree high-level terms for general topics: the foundation of a common vocabulary in biological and biomedical science and bioinformatics*
- *Discuss use cases and how this vocabulary will make an impact in our community*
- *Discuss a publication on the common vocabulary with potential authors*

Further, we will seek to assemble a coalition of resources, to seek funding to adopt and develop this vocabulary in support of resource integration and discovery, with concrete commitments to:

- *Promote the use of this vocabulary within your infrastructure, institute, project etc., and beyond,*

and, subject to resource constraints:

- *Adopt the vocabulary within your own resources, e.g. for purposes of annotation, search / browsing and sharing data*
- *Develop the vocabulary going forward, building on the foundation agreed during the workshop.*