EMBL-EBI

identifiers.org

Identifiers.org / MIRIAM Registry annotation and cross-referencing framework

Nick Juty, Nicolas Le Novère, Camille Laibe

Introduction

More than ever, the era of *data integration* has highlighted the key requirement to reference specific data in an unambiguous and perennial way, in order to enable community-level sharing, development, exchange and reuse of information. In the field of Systems Biology, which is concerned with creating quantitative models of biological processes, these requirements have directly led to the creation of the Minimal Information Required In the Annotation of Models (MIRIAM, http://biomodels.net/miriam/) **[1]**. MIRIAM provides a specific set of guidelines that can be implemented within any structured modelling format.

We have developed the MIRIAM Registry (http://www.ebi.ac.uk/miriam/) **[2]** to support one requirement of the MIRIAM guidelines: the annotation of a model by identifying all its components. This is achieved by means of *Uniform Resource Identifiers* (URIs).

Identifiers.org is built above the MIRIAM Registry. Created to facilitate the sharing of knowledge beyond the Systems Biology community, Identifiers.org allows the creation of persistent and directly resolvable identifiers, in *Uniform Resource Locator* (URL) form.

Annotations

Annotations are essential for data identification and semantics:
- data understanding and reuse
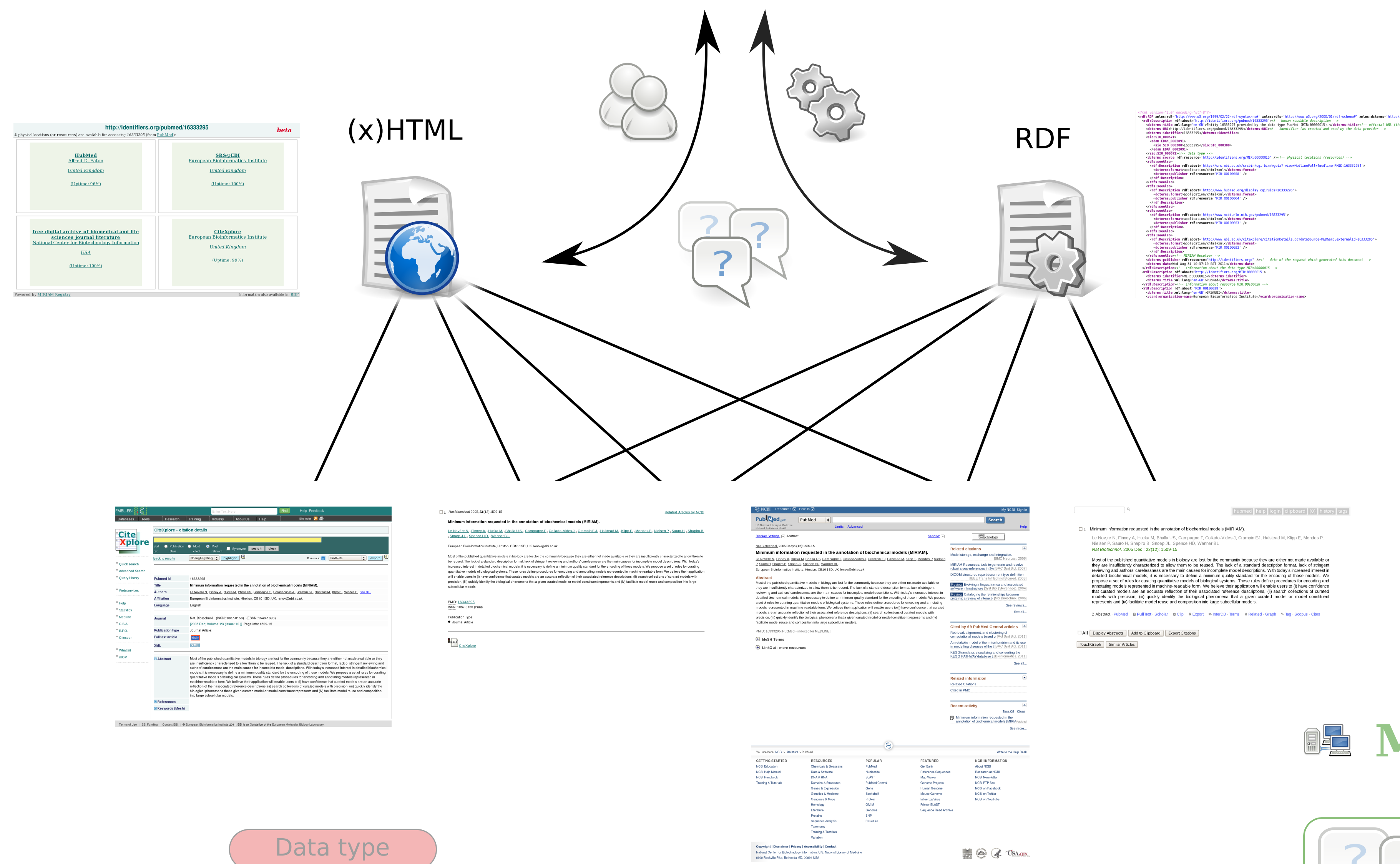- data comparison
- data integration

Cross-references

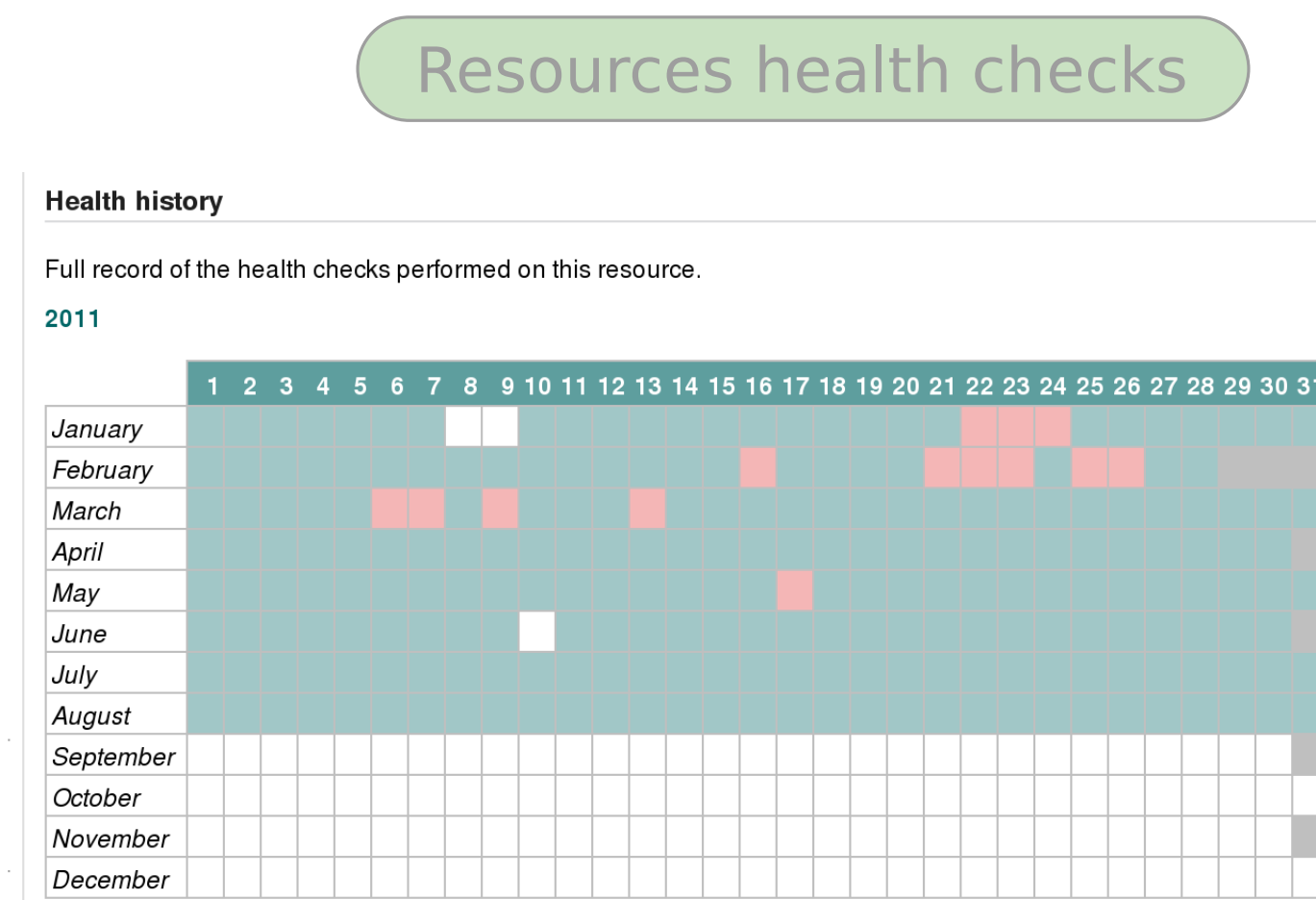Characteristics of a useful identifier for cross-references:
- unique and unambiguous
- perennial
- standard compliant
- resolvable
- free of use

Identifiers.org architecture

http://identifiers.org/pubmed/16333295

SBML URI example

```
[...]
<species metaid="metaid_0000006"
         id="L_EGFR"
         compartment="compartment"
         initialConcentration="0">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22rdfsyntaxns#"
             xmlns:bqbiol="http://biomodels.net/biologyqualifiers/">
      <rdf:Description rdf:about="#metaid_0000006">
        <bqbiol:hasPart>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:uniprot:P07522" />
            <rdf:li rdf:resource="urn:miriam:uniprot:Q9QX70" />
          </rdf:Bag>
        </bqbiol:hasPart>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
[...]
```

MIRIAM URIs

Human calmodulin: P62158 in UniProt
    urn:miriam:uniprot:P62158
    http://identifiers.org/uniprot/P62158

Alcohol dehydrogenase: 1.1.1.1 in EC code
    urn:miriam:ec-code:1.1.1.1
    http://identifiers.org/ec-code/1.1.1.1

MAPKK activation: GO:0000186 in Gene Ontology
    urn:miriam:obo.go:GO%3A0000186
   http://identifiers.org/obo.go/GO:0000186

Resolving services

The **MIRIAM Registry** is a set of on-line services centered around a catalogue of data types. Data types can be ontologies, such as Gene Ontology, or primary data resources available via the Web, such as UniProt and PubMed. Each of these is uniquely identified within the *MIRIAM database*, and information stored regarding the corresponding physical URLs (data resources) through which their associated data can be accessed.

There are two ways to create and resolve annotations arising from data types registered in the database:

i) Web services are available to generate a MIRIAM URI from a data type name and dataset identifier. Resolution services that generate physical locations (URLs) for MIRIAM URIs are provided as SOAP and REST Web Services.

ii) Resolvable URLs using the **Identifiers.org** framework. These URLs resolve directly to an intermediate location, providing previews for the total set of all resources where that information may be resolved.

Data type

Resources health checks

MIRIAM Registry architecture
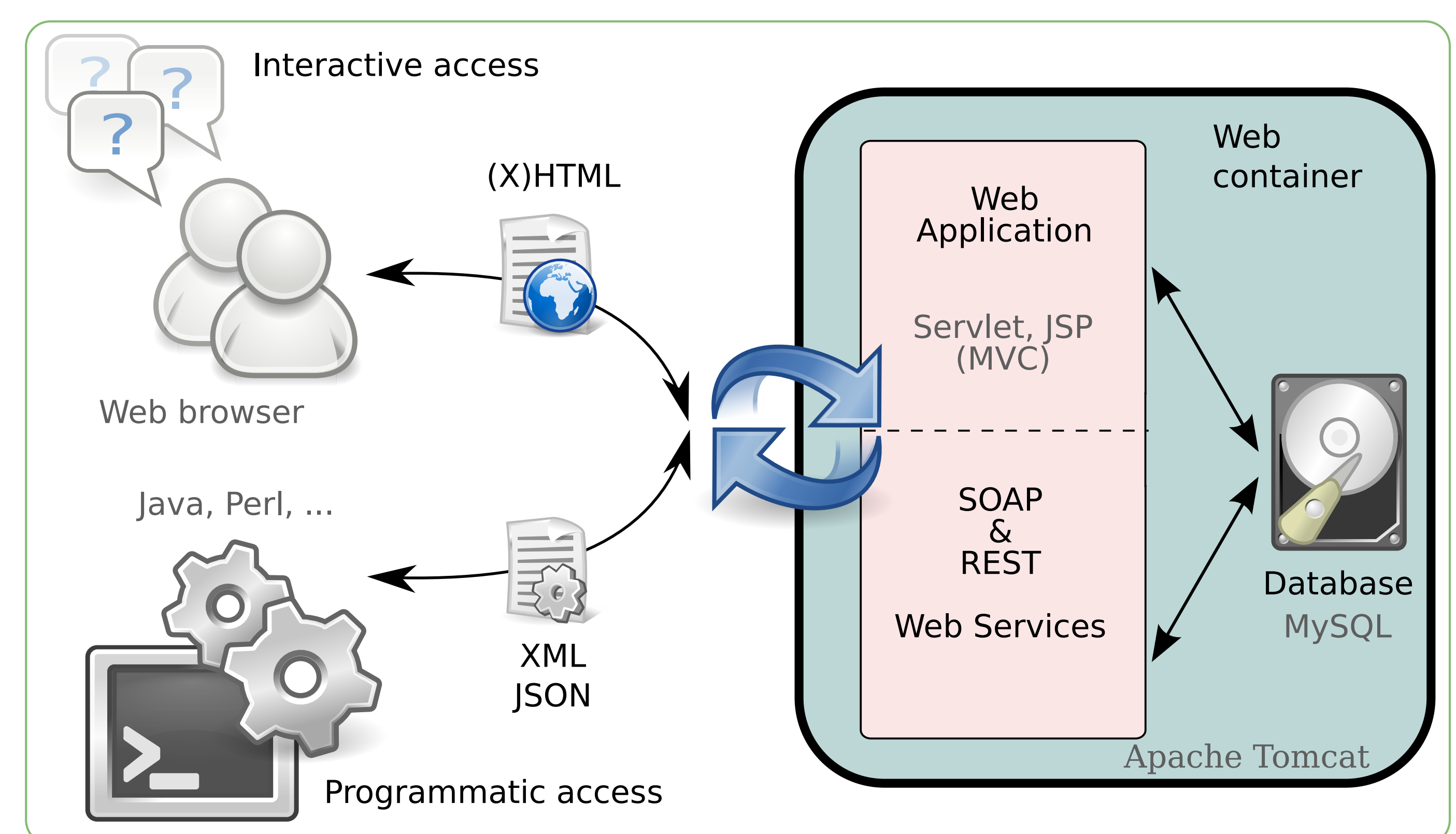
Summary

- MIRIAM annotations are widely accepted, and are being used and supported by various communities such as BioModels.net **[3]**, SBML, CellML, BioPAX and the Proteomics Standards Initiative (PSI).
- In an effort to enable their use by the Semantic Web and Linked Data communities, we have provided an additional, parallel system of identification that provides resolvable URLs, and furthermore allows access to the MIRIAM Registry information in *Resource Description Framework* (RDF) format.
- **MIRIAM URIs** facilitate the identification, exchange, and integration of data in the Life Sciences.

References

[1] N. Le Novère, A. Finney, M. Hucka, U. Bhalla, F. Campagne, J. Collado-Vides, E.J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J.L. Snoep, H.D. Spence, and B.L. Wanner. **Minimum Information Requested In the Annotation of biochemical Models (MIRIAM).** *Nature Biotechnology*, 23(12):1509–1515, 2005.
[2] C. Laibe and N. Le Novère. **MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology.** *BMC Systems Biology*, 1: 58, 2007.
[3] Le Novère N. **BioModels.net, tools and resources to support Computational Systems Biology.** Proceedings of the 4th Workshop on Computation of Biochemical Pathways and Genetic Networks. Logos, Berlin, pp. 69-74, 2005.

laibe@ebi.ac.uk, juty@ebi.ac.uk
lenov@ebi.ac.uk

EMBL - European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton
Cambridge
CB10 1SD
United Kingdom
http://www.ebi.ac.uk/