Presentation Viktor BioSyst 2017 Gothenburg Sweden

1. Slaid 1: Title, logos
2. Slaid 2: What is OpenBiodiv – definition – the 3 elements of the brand:
   1. OpenBiodiv Ontology
   2. OpenBiodiv Linked Open Data set
   3. OpenBiodiv knowledge system (website, application)
   4. Extended title of OpenBiodiv: OpenBiodive Knowledge Management System (OBKMS)
3. Why it is needed – from the viewpoint of data handling
   1. No way to make biodiv interoperable so that to generate new knowledge
   2. No way to make this data interoperable with otherf domains (e.g. genomics, biomedicine, geography, etc.)
   3. No way to utilize data included in the narrative of the published papers
   4. No way to put all this data within a common, logically consistent, conceptual model
   5. No way to handle data of various type and various classes via a common UI
   6. No way to atomize small data facts together with their provenance., e.g. sources of citation (published articles)
4. Why it is needed – from the viewpoint of taxonomists and biodiversity scientists
   1. New technical and technological method to approach the long-standing issues in taxonomy and biodiversity science (GNA goals, reconciliation of names, use of taxonomic concepts, dark taxa, etc.
   2. No way to put together various classes biodiv data from dispersed sources (articles, journals)
   3. Need in open democratic system of biodiversity data free of predefined hierarchical pre-requisites (e.g. a taxonomic backbone used as an “architectural backbone”; several taxonomic trees are possible and can be used in such a system)
   4. Formalization of the taxonomic concepts and their use following a common data model
   5. Joint formal management of LKinnean names and OTUs
5. Where it is rooted?
   1. Semantic publishing of biodiversity papers (XML tagging and TaxPub)
   2. Textmining from legacy literature using same model as prospectively published papers
   3. Common model of terms, data classes and relationships between these
   4. Rapid development of RDF, OWL and LOD technologies in all domains
6. How it works?
   1. What is a RDF triple (perhaps continuing the XML example of the species name from the previous slide and translate it into an UNIVERSAL language (RDF)
   2. Extracting data and creating triples
   3. Linking triples via a common ontology model
   4. Using triples to provide search, browsing and extracting
   5. Using triples to make new inferences and create new knowledge from a potentially huge dataset of various classes and types of data
7. How it looks like in reality
   1. Answering pre-defined questions, for example:
      1. Which species have been described in family “XX-ceae or YY-idea” from Afrika between 1850 and 2010
      2. Which specimens from collection XYZ have been published in the literature in certain period, and/or by certain taxonomist and so on.
   2. Creating queries via user-friendly interface (similar to advanced search functionality but much wider ins scope and topics
   3. Creating queries via SPARQL end-point – unlimited flexibility to work with large data from a particular or several federated linked open datasets
   4. Rapid alerting services for any interested users on newly published names, nomenclatural acts, occurrences , facts, checklists, etc.
   5. Automated profiles of taxon names and concepts
   6. Extracting and *de novo* publishing of nano-publications from OpenBiodiv
      1. New taxa and names
      2. New species hypotheses or OTUs
      3. Nomenclatural changes
      4. Refining taxonomic concepts (annotations on published taxonomic concepts – taxonomic concept alighments)
      5. New occurrences
      6. New biological facts
      7. New traits
      8. Nanopublications in the form of annotations on other nanopublications (e.g. nao-reviews)
   7. Data re-use via specially developed apps linked to the system
   8. Use by machines via Application Programming Interfaces (APIs)
8. Sample use cases and users
   1. For taxonomists
   2. For collection managers
   3. For biodiversity scientists (in general)
   4. For data scientists and data aggregators in biodiversity
   5. For scientists and data aggregators from other domains
9. Timeline and implementation schedule
   1. Collating all data from all xisting Pensoft journals and Plazi TreatmentBank via automated continuous update mechanism
   2. Exploring new methods for creating new inferences via advanced search and machine-learning (collaboration with Swedish Natural History Museum)
   3. Implementation of the website and user interface
   4. OpenBiodiv fully operational and open for anyone to contribute, use and collaborate

**Nanopublications**

1. Based on TaxonA -> TaxonB relationships

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TaxonConcept A** | | | | | **Predicate** | **TaxonB** | | | | |
| Genus or other rank | Species | Subspecies | Author | Taxon concept  sensu… |  | Genus or other rank | Species | Subspecies | Author | Taxon concept  sensu… | |
| **Taxon concept relationship** | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Nomenclatural relationships** | | | | | | | | | | |
|  |  |  |  |  | IsSynonymOf |  |  |  |  |  | |
|  |  |  |  |  | RemovedFromSynonymy |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Biological relationships** | | | | | | | | | | |
|  |  |  |  |  | IsParasiteOf |  |  |  |  |  | |
|  |  |  |  |  | IsHostOf |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |

Provenance: Published in: DOI, or Unpublished

Publishing record: Registered person + time stamp

**Entering taxon names in the interface:**

1. Autocomplete from OpenBiodiv system (based on the GBIF backbone and their/our identifiers
   1. TaxonName + Authority, Year
   2. TaxonConceptLabel (if different from the original?)
2. Create new name (if absent in the autocomplete)