

## WP7 - Work package description

Work package number	WP7
Type of activity	RTD
Work package title	EBI Pipeline, Cross-map, and GSD Piping
Start month	1
End month	36
Lead beneficiary number	3

### Objectives:

To prototype, make available for test, and implement components from EBI for ongoing usage in collaboration with WP 2. These components will be designed in collaboration with WP2, where there will be integration with the rest of the project, scoping, testing and acceptance for service. The EBI components are:

1) a taxon-centric portal, both web and webservice, into molecular data held in the European Nucleotide Archive (ENA); this would provide summary details of sequences and selected annotations available for a given taxon in the NCBI Taxonomy and onward access into the molecular records themselves.

2) a protocol and service for supplying a particular instance of the NCBI Taxonomy in use at EBI to the CoL cross-mapping tool for the purpose of running a cross-map against a particular instance of the CoL.

3) a service for the capture of the CoL synonymic species checklist and its taxonomic hierarchy into the ENA informatics platform at an agreed frequency appropriate to the dynamic nature of the CoL Taxonomy.

4) a service to transmit batches of taxon records, based on CoL or NCBI taxonomies, that following cross-mapping remain unmapped to the Catalogue of Life; this service would present data to the Catalogue of Life reception buffer, for subsequent processing. The input to the CoL reception buffer will harmonise with the input requirements for the CoL GSD-piping tool being created in WP12, using protocols agreed and tested in WP2. Those 'successfully' processed by the Sp2000 GSD Taxonomic Expert Array, will appear within later instances of CoL.

5) an extension to the taxon-centric portal described in component (i) of this workpackage and drawing on the cross-mapping made available at EBI by component (iii) of this workpackage.

### **Description of work and role of partners:**

All tasks in this work package undertaken by Partner 2 in collaboration with WP2.

#### **Task 7.1**

Create a taxon centric portal to the ENA for EBI, both web and web-service, indexed by the NCBI Taxonomy, accessing summary details of sequences and selected annotations, with onward access to the molecular sequences themselves.

#### **Task 7.2**

Service for supply of NCBI Taxonomy to CoL cross-mapping: a protocol and service for supplying a particular instance of the NCBI Taxonomy in use at EBI to the CoL cross-mapping tool for the purpose of running a cross-map against a particular instance of the CoL.

#### **Task 7.3**

Capture of CoL checklist and hierarchy, including cross-mappings: a workflow for the capture of the CoL synonymic species checklist and its taxonomic hierarchy, and a cross-mapping with the NCBI Taxonomy into the ENA informatics platform at an agreed frequency appropriate to the dynamic nature of the CoL Taxonomy and meeting CoL standard credit requirements.

#### **Task 7.4**

Supply of unmapped taxa to the CoL reception buffer: a service to transmit batches of taxon records, based on NCBI taxonomies, that following cross-mapping remain unmapped to the Catalogue of Life; this service would present data to the Catalogue of Life reception buffer, for subsequent processing.

#### **Task 7.5**

Extend the taxon-centric portal described in component (i) of this workpackage and, drawing on the cross-mapping made available at EBI by component (iii) of this workpackage, to integrate the taxon-centric view with searching by both the NCBI and CoL Taxonomies.

### Person-Months per Participant:

Participant number	Participant short name	Person-months per participant
3	EMBL	13.00
	Total	13.00

### List of deliverables:

Deliverable Number	Deliverable Title	Lead Beneficiary number	Estimated indicative person-months	Nature	Dissemination level	Delivery date
D7.1	Taxon-centric portal at EBI	3	13.00	D	PU	36
		Total	13.00			

### Description of deliverables:

D7.1)

Taxon-centric portal at EBI: Extension of taxon-centric portal at EBI to support the cross-mapped CoL Taxonomy.

[month 36]

### Schedule of relevant Milestones:

Milestone number	Milestone name	Lead Beneficiary number	Delivery date from Annex I	Comments
MS6	Taxon-centric portal available at EBI	3	6	
MS23	Service for supply of NCBI Taxonomy to CoL for cross-mapping	3	18	
MS28	Initial implementations linking CoL services reported to WP2	3	20	
MS35	Capture of the CoL checklist and hierarchy, with cross-mappings to the NCBI Taxonomy	3	23	
MS39	Supply of unmapped taxa to the CoL Reception Buffer	3	24	