The SLING project has supported the European exploitation of biomolecular information in three ways:

1) Providing high-quality electronic services allowing access to comprehensive, accurate and up-to-date information

SLING funding has contributed to the EBI effort in collecting, organizing and making available a wide range of biomolecular data, and providing services allowing scientists to exploit those data (including genes and genomes, protein sequences and structures, molecular behaviour and interactions). During SLING, EBI web hits have more than doubled; the monthly unique users have increased by over a half; individual EBI resources have genereally seen large increases in access and numbers of users.

SLING funding has contributed to the SIB protein annotation enhancements for UniProtKB/SwissProt (a high-quality curated resource of protein sequences and functional annotation). During SLING, 124,000 curated sequence records were added, and 7,800 corrections were made, 85,000 Gene Ontology manual annotations were added, 724 new complete EC numbers were created, and 100,000 individual PTMs were curated. SLING funded ongoing curation of structural information from publications - 5,000 references, 40,000 cross-references, in 4,000 UniProtKB/Swiss-Prot records.

2) A programme of joint research activities which will enhance these services and develop then in response to changing science

Before SLING there was little or no exploitation of Patent Literature in basic science. SLING has now contributed to significant advances in the availability of full-text resources that support computational approaches to literature/data integration. SLING work has established 1) a full-text article repository (An Open Access article FTP site, plus release of a public web service) and 2) a full-text patent repository (containing 1.3 million full-text patents, plus a basic web service). In addition, the SLING work has informed possible future developments of full-text resources; in the case of journal articles, the value of re-using full-text has been demonstrated by the development of a text-mining based tool that supports curation work.

Before SLING, databases were poorly-adapted to include next-generation sequence data storage and interpretation. After SLING, those data have been collected, integrated and organised through such activities as the development of 1) an epigenomics checklist of minimal submitter information, 2) a meta- and full-data exchange pipeline between NCBI and EBI, 3) submission tools with automated receipt and status reporting, and re-launch of loading jobs, 4) metadata text-search and link to API that allows users to provide text that may appear in the metadata layer and return all of those records where the text is present, 5) an Ensembl 'Regulatory Build' analsysis pipeline, and 6) a portal between Ensembl and the sequence archives.

Before SLING, there was patchy information attached to genomes. SLING has now provided the infrastructure, methods, and curatorial expertise to acquire, annotate, integrate and disseminate Next Generation Sequencing based transcriptomics data across the EBI, between SLING partners, in Europe and Internationally. To achieve these aims, SLING funding has contributed to the development of 1) a community-generated and adopted format (MINSEQE); 2) a data-exchange agreement between the world major repositories (USA NCBI Gene Expression Omnibus and the European ArrayExpress); 3) Open Source software supporting the data exchange agreement with conversion scripts, and curator support tools; 4) an Open Source submission tool for submitting NGST data; 5) training courses and resources; 6) an enhanced ArrayExpress GUI with support for search and display of NGST data; 7) the ArrayExpressHTS R package for processing NSGT data for inclusion in the Gene Expression Atlas; 8) a pipeline to link the

ENA short read component, ArrayExpress and Ensembl; and 9) an ontology supporting the description of NGST experiments, and aiding curation.

Before SLING, the richest quantitative proteomics information was poorly provided. After SLING, the PRIDE database has now been enhanced and refined to represent quantitative protein expression data, and to provide both tabular download and external analysis capabilities for such data. Work has been done on a simplified data model and representation for information that supports the main quantification approaches. A new Open Source tool (PRIDE Inspector) has been developed for visualizing and performing an initial assessment of the quality of mass-spec proteomics data. The tool is particularly useful for journal editors and reviewers, as well as facilitating connection to the external Ensembl and Reactome tools. Through SLING, initial steps have been taken to develop PRIDE from a mass-spec specialist resource into a much wider protein expression resource for the molecular biology community in Europe.

Before SLING, dynamic interaction networks were scarcely represented in databases. After SLING, the IntAct database has now been enhanced to allow efficient deposition, curation, display, and analysis of dynamic interaction data, along with integration into a systems biology context. To integrate 3rd party data, a standard interface (PSICQUIC), has been developed (and adopted by all 25 major resource providers of 150 million interactions) to query multiple interaction data resources using the same query. Before SLING, the connection between molecular interaction data and supporting mass spectrometry data was difficult to make, because they were located in two unconnected databases: IntAct (molecular interactions), and PRIDE (mass spectrometry). After SLING, there is now a better connection between the two types of data, allowing the IntAct data to directly access the supporting mass-spec evidence in PRIDE via the existing DAS standard.

Before SLING, the standardisation of protein annotation was limiting inter-database connections. After SLING, a set of protein-naming guidelines has now been developed to facilitate inter-database connections. The naming-standard guidelines have been agreed upon and adopted by a number of the major sequencing centres, data providers, and nomenclature committees. The on-going resolution of existing discrepancies (in protein nomenclature between resources) will facilitate further data integration and interpretation by users. In addition, SLING has contributed to the enhancement of annotation standards of UniProtKB binary protein interactions via a tool developed to conform to the IMEx Consortium MIMIx standard. During SLING, 10 curators have used this tool to curate 2,815 MIMIx-level binary interactions from 1,407 experiments described in 409 publications.

Before SLING, enzyme information lacked connectivity and richness. After SLING, the BRENDA enzyme database has now been substantially enhanced for applications in systems biology and medicine, with emphases on 1) completing the manually-annotated data with full sets of enzyme data by text-mining methods; 2) provision of automated access to the manually-annotated data; and 3) the enlargement of the fields covered by BRENDA (including new enzymes). Text-mining methods have been developed to extract kinetic enzyme data from literature abstracts, to add to the existing manually-extracted data. New output functionality has been developed to allow automatic generation of a single SBML file containing the kinetic data of enzyme-catalyzed reactions of an organism. The widely-used BRENDA tissue ontology standard has been expanded by adding branches and nodes to the tree, and compiling new terms including their definitions; 1,345 new single terms and 1,179 new definitions have been added under SLING. BRENDA now gives access to the PDB enzyme 3D structures and to the UniProt enzyme-specific protein sequences. SLING has also enhanced the naming and classification of enzymes by funding the submission of 336 new EC numbers.

Before SLING, there was poor information for chemicals in biology, and a dependency on proprietary code. During SLING, to understand the needs of multiple user communities and to custom-tailor further ChEBI development, a User Survey was commissioned. Consequent to the survey results, the curation tool was enhanced to link to the CiteXplore text-mining infrastructure to text-mine a given citation; along with the increased new citations, the text-mining enabled addition of extra biological and chemical role data to the ChEBI ontology. Before SLING, ChEBI relied on a number of proprietary chemoinformatics modules (for chemical structure searching, display/editing of 2D chemical structure diagrams) which prevented the free dissemination of the ChEBI technology to the scientific community. Under SLING, ChEBI has now, and is being re-engineered to implement Open Source alternatives to all but one of the proprietary libraries, towards the ultimate aim of making ChEBI source code accessible and re-distributable without licensing restrictions. After SLING, the number of monthly unique visitors to the ChEBI website has increased from 15,150 to 24,100; programmatic access to the ChEBI data has increased from a monthly average of 418,000 hits to 3,365,000 hits.

3) Extensive pan-European user-training to facilitate exploitation of the information

Before SLING, the demand for bioinformatics training far exceeded supply. To meet that demand, SLING has now played a crucial role by funding a training programme of 33 Roadshows that have reached c. 1,100 experimental researchers in the molecular life-sciences community throughout Europe, particularly the new EU Member States. The Roadshows have also provided Europe with a lasting legacy of online training courses that have been accessed by 24,000 researchers in the 1st year of operation. In addition, SLING funding has enabled European bioinformatics trainers to meet annually to exchange and develop a comprehensive set of best-practice trainer guidelines.