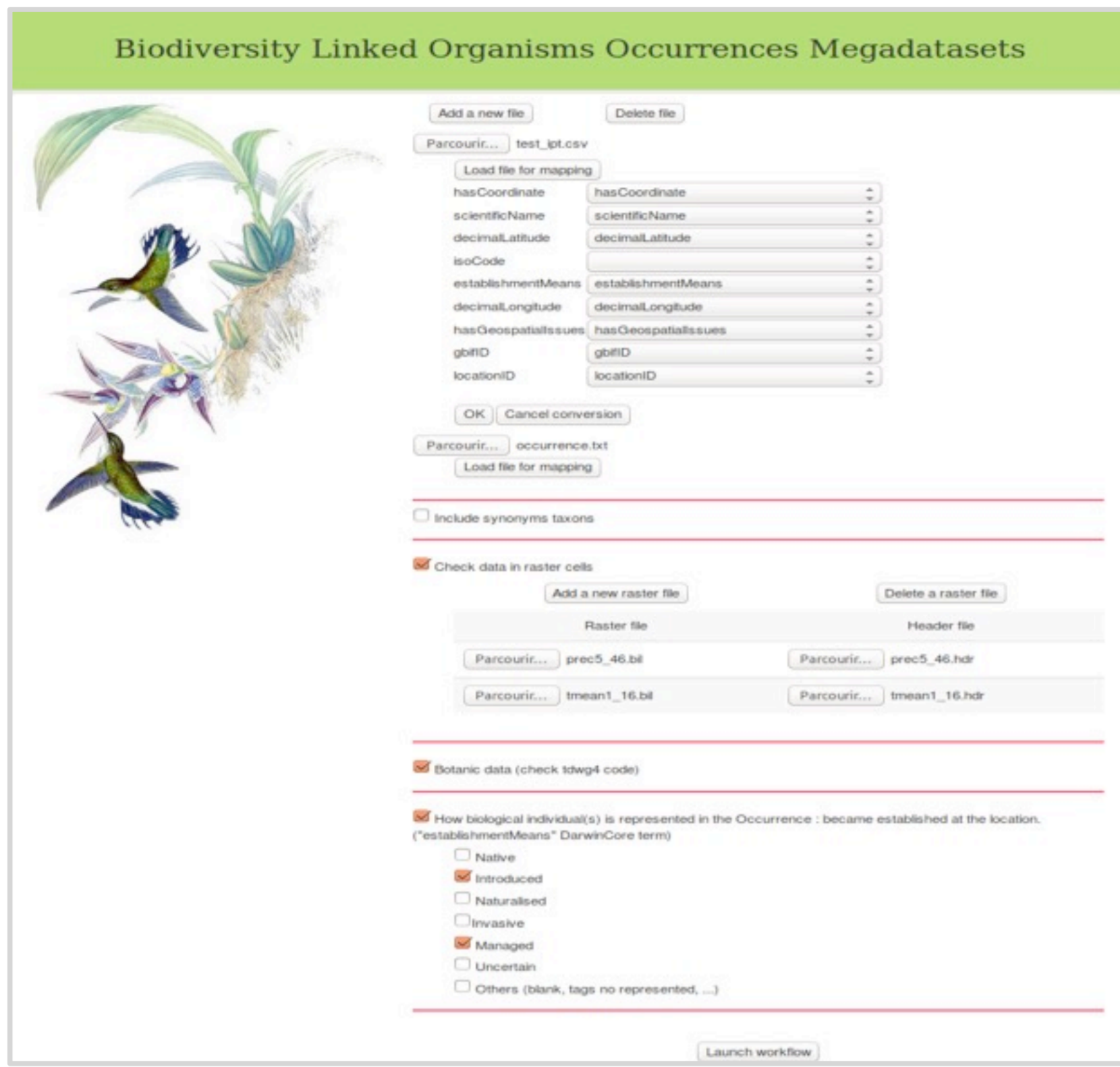


Towards an automatic curation process of Biodiversity data : the BLOOM project



Summary: Biodiversity research institutions foster ever-growing amounts of data that are essential to answer fundamental questions about life on our planet and have a high scientific, economic and societal value. Over the years European and global initiatives resulted in advanced information infrastructures for aggregating and managing these data, resulting in domain-specific (e.g. specimen data, observations, molecular data) huge amounts of complex and heterogeneous big data. When combined, these datasets give a picture of the world's past and present biodiversity, highlighting baseline data and tools to predict our future biosphere.

The BLOOM project objectives focus on the data curation process, providing a versatile and appealing workflow for any protagonists engaged in Conservation, Ecology or other cross-domains of research and expertise. We designed it as an open tool fully compliant with existing platforms for Biodiversity implementing data standards, which will improve biological forecasting by validating data quality as a first step of analysis. The project suits present-day global issues with regards to monitoring the biosphere toward detecting next ecological systems shift and evaluating biological irreversible thresholds which we believe will certainly opening up new research tracks and consolidate current knowledge on Biodiversity. This tool will be also useful for macroecologists using massive datasets such as the ones sourced from the digitization of entire herbaria or natural history collections.

DarwinCore: body of standards created by the Biodiversity Information Standards (TDWG, <http://www.tdwg.org>). It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.

Biodiversity Information Standards TDWG

Interoperability mapping to DarwinCore

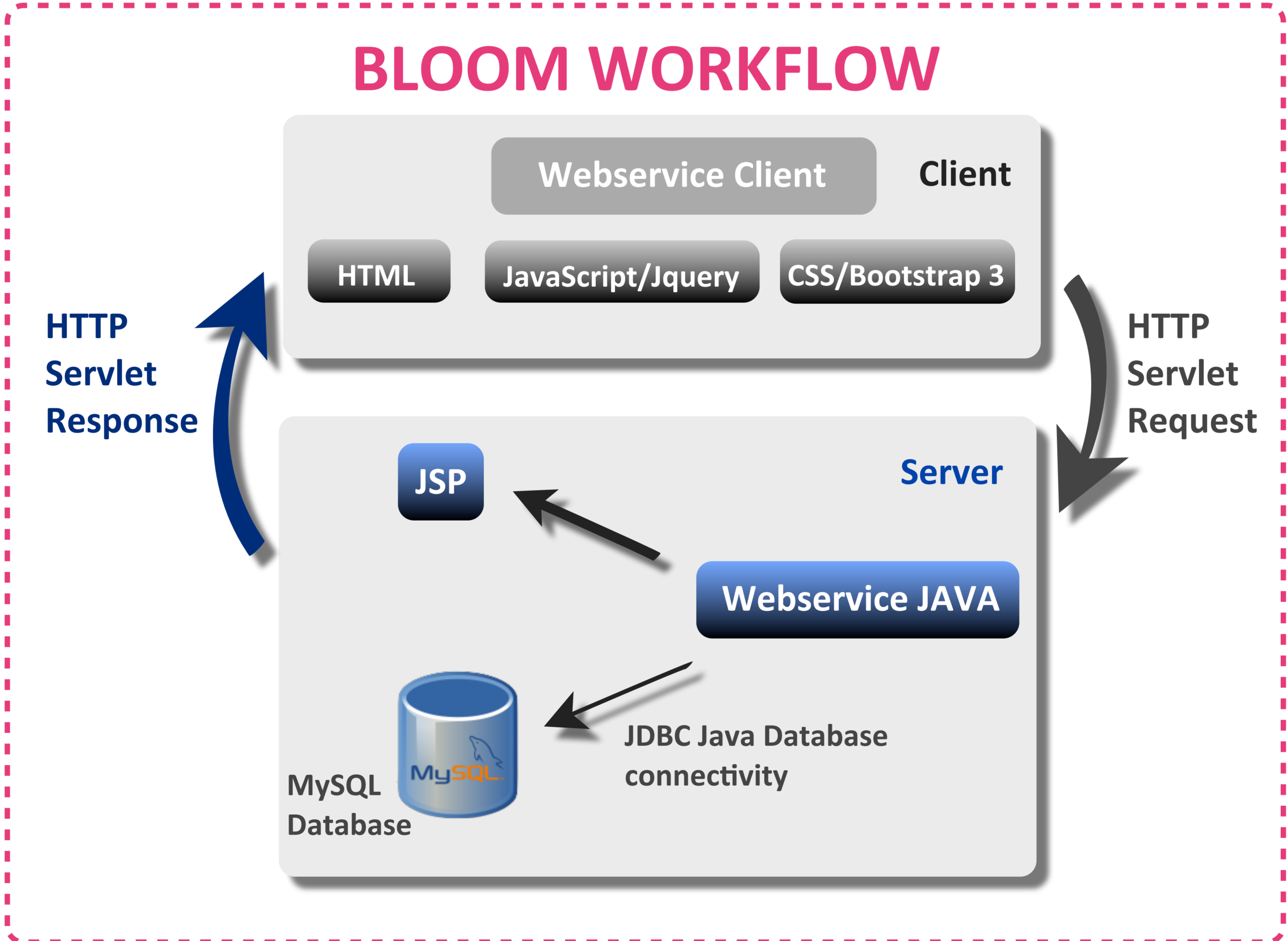


Georeferencement validation tool
checking latitude/longitude consistency (null, out of range), iso2 coding (country codes, iso3166)



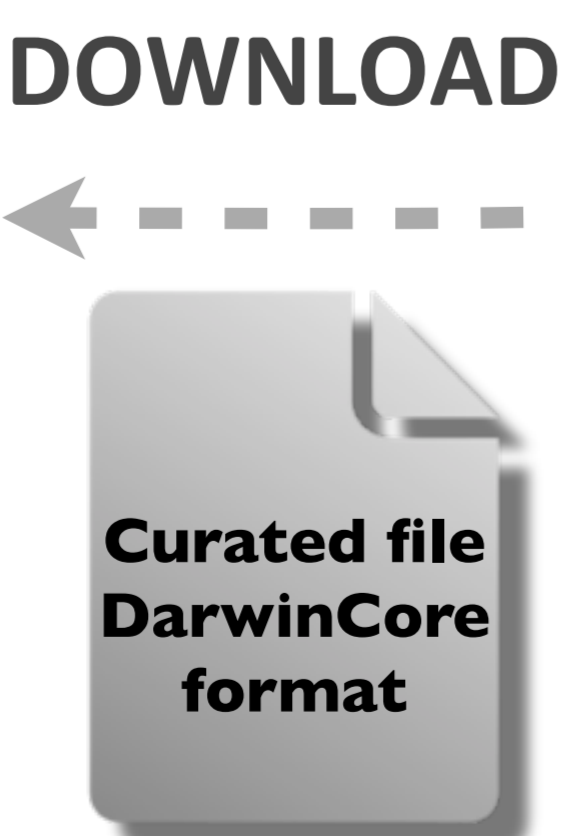
- data integration in MySQL database through a JDBC Java API database connection
- UUID (Universal Unique Identifier) guaranty persistence

errors are stored in specific files



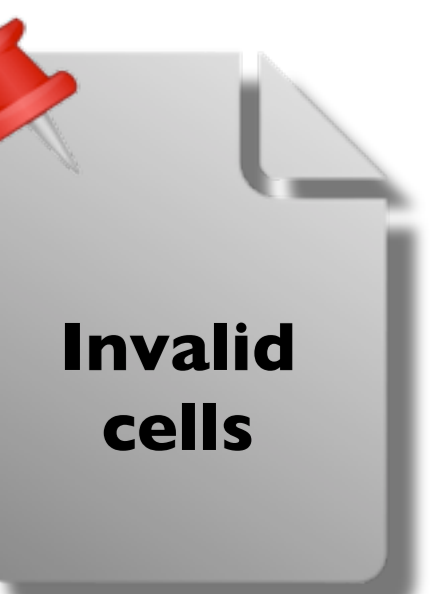
Taxonomic validation tool
checking taxonomic names with use of reconciliation service (API Kew, <http://www.kew.org>) or other reconciliation API

errors are stored in specific files



Masks validation tool
checking raster data (projection system), bioclimatic data (WorldClim)

errors are stored in specific files



Biodiversity analysis
Ecosystems monitoring

Macroecology