**Titolo:**
Sistemi di modellazione concettuale per oggetti digitali eterogenei

**Progetto di ricerca**:

Scopo dell’assegno è di lavorare per il CRR-MM (Centro di Risorse per la Ricerca - Multimedia) dell’Università di Bologna (<http://www.unibo.it/it/ricerca/progetti-e-iniziative/centro-di-risorse-per-la-ricerca-multimediale-crr-mm>).

Scopo prioritario del Centro è di creare una piattaforma per l’archiviazione di grandi collezioni di dati eterogenei (corpora, documenti di testo, fotografie, video, audio, etc.) nel solco dei repository delle semantic digital libraries (cfr. Kruck and McDaniel 2009).

Tali collezioni, in prima battuta provenienti dal lavoro di ricercatori presso l’Università di Bologna, sono costituite da oggetti digitali realizzati attraverso tecnologie diverse e diversi linguaggi: basi di dati (da Access a SQL, etc.), documenti in formato solo testo, documenti semi-strutturati (come HTML, XHTML, XML), e documenti audio e video. Collezioni che vanno quindi preservate, sia dal punto di vista fisico che logico (cfr. digital preservation), e rese accessibili attraverso ambienti di accesso e query Web (es. Exist e linguaggio XQuery; SQL/PHP).

L’infrastruttura che ospiterà tali dati dovrà essere realizzata attraverso l’analisi dei modelli concettuali esistenti (OAIS, DELOS), stabilendo sistemi ad accesso aperto per l’harvesting dei dati (OAI-PMH) e costruendo un ambiente per l’accesso agevolato lato operatore (CMS).

Tale infrastruttura dovrà consentire poi di utilizzare le collezioni in una dimensione sociale, partecipata e collaborativa nel solco dei principi del Web 2.0 (collaborative tagging anche folksonomico).

Sarà poi opportuna una definizione di ontologie atte a gestire l’intero repository rispetto ai diversi livelli della gestione (descrittivo, amministrativo-gestionale, strutturale) ma anche le singole collezioni nell’ottica del Web Semantico (XML/RDF(S)/OWL).

Allo scopo di garantire massima visibilità dei dati scopo del CRR è anche di elaborare linee guida per la disseminazione dei dati in forma LOD (Linked Open Data).

L’infrastruttura dovrà garantire l’integrazione di servizi ad hoc (tools) per le specifiche esigenze dei progetti gestiti.

**Piano di attività:**

Per ottemperare a questi scopi le attività previste riguardano:

* analisi delle piattaforme esistenti (es. Fedora) e dei modelli concettuali relativi;
* analisi dei servizi per la gestione dei dati, verificando i progetti sw in essere (es. tool integrati in TextGrid, Collate, NINES, etc. o tool autonomi es. Pundit);
* analisi delle esistenti ontologie per il cultural heritage (es. CIDOC-CRM, EAD/EAC, FRBR) e dei vocabolari per la metadatazione (es. DC, TEI, METS, MODS) nonché dei vocabolari controllati in essere (es. Getty Thesaurus, WordNet, Geonames, Dbpedia, etc..) e dei sistemi di soggettazione e classificazione;
* analisi dei CMS esistenti (es. Joomla, Drupal, Plone, etc.);
* studio dei formati audio e video per il Web;
* impiego dei linguaggi di markup (XML, HTML) e dei relativi linguaggi di layout e manipolazione (XSLT e CSS);
* produzione di basi di dati: modelli concettuali, metodi implementativi (modelli logici), costruzione di interfacce d’accesso (es. faccette), strumenti di query Web;
* realizzazione di ontologie ad hoc per singolo progetto e di un’ontologia generale per il repository;
* creazione di LOD su singoli progetti.