

Preservazione dell'Informazione Digitale ed Eredità Scientifica

Riassunto

CNR, INAF, INFN e INGV stanno sviluppando un coordinamento delle attività sul tema della preservazione a lungo termine dei dati digitali. Il primo risultato è stato la costruzione di un progetto che vuole definire e realizzare una piattaforma multidisciplinare per l'archivio dell'informazione digitale e la conservazione dei meccanismi di accesso e di utilizzo dei dati. Le applicazioni scientifiche, di interesse degli enti di ricerca partecipanti, verranno usate per la raccolta dei requisiti e per l'adattamento dei sistemi di preservazione ed accesso del codice e dei dati alle necessità delle varie comunità scientifiche e agli eventuali standard già esistenti. Saranno esaminati quattro casi espliciti: La conservazione dei dati Omics e relativi strumenti software di analisi che ha un volume di dati che cresce esponenzialmente grazie allo sviluppo delle tecnologie genomiche, proteomiche, trascrittomiche e bioinformatiche. La conservazione dei dati del centro INAF-IA2 e relativi sistemi software di data ingestion. Il progetto di conservazione dei dati dell'esperimento CDF, che impone l'adattamento del middleware di trasferimento per i 5 PB di dati e lo sviluppo di una infrastruttura virtuale per il codice di esperimento. Infine la preservazione del patrimonio storico dei sismogrammi custoditi negli archivi degli osservatori distribuiti nel territorio nazionale. Si svilupperanno le funzionalità specifiche per i sistemi studiati seguendo le linee guida dello standard OAIS. Particolare attenzione sarà data alla scalabilità delle soluzioni adottate in vista della complessità che si raggiungerà con i dati degli esperimenti LHC o nel progetto CTA dove il volume dei dati raggiungerà 10 PB all'anno e il software di accesso e analisi dei dati consiste di milioni di linee di codice. Il coordinamento proporrà che gli enti di ricerca presentino questa iniziativa al prossimo bando per i progetti premiali del MIUR. In questo lavoro si vogliono descrivere gli obiettivi posti, la situazione rispetto ai progetti esistenti, le fasi di realizzazione e le strategie per una partecipazione in HORIZON 2020.

Sottomesso a

Workshop GARR

CSD-2012 sulle infrastrutture digitali (e-Infrastructure) per il supporto al calcolo scientifico e alle infrastrutture di ricerca della Roadmap ESFRI. GARR