



IBM Première 2007

Dove le idee diventano realtà.



L'informazione al servizio della ricerca


Pietro Leo


Executive IT Architect


IBM Global Business Services – IBM Innovation Lab


23 Ottobre 2007 – Milano Area Kitchen

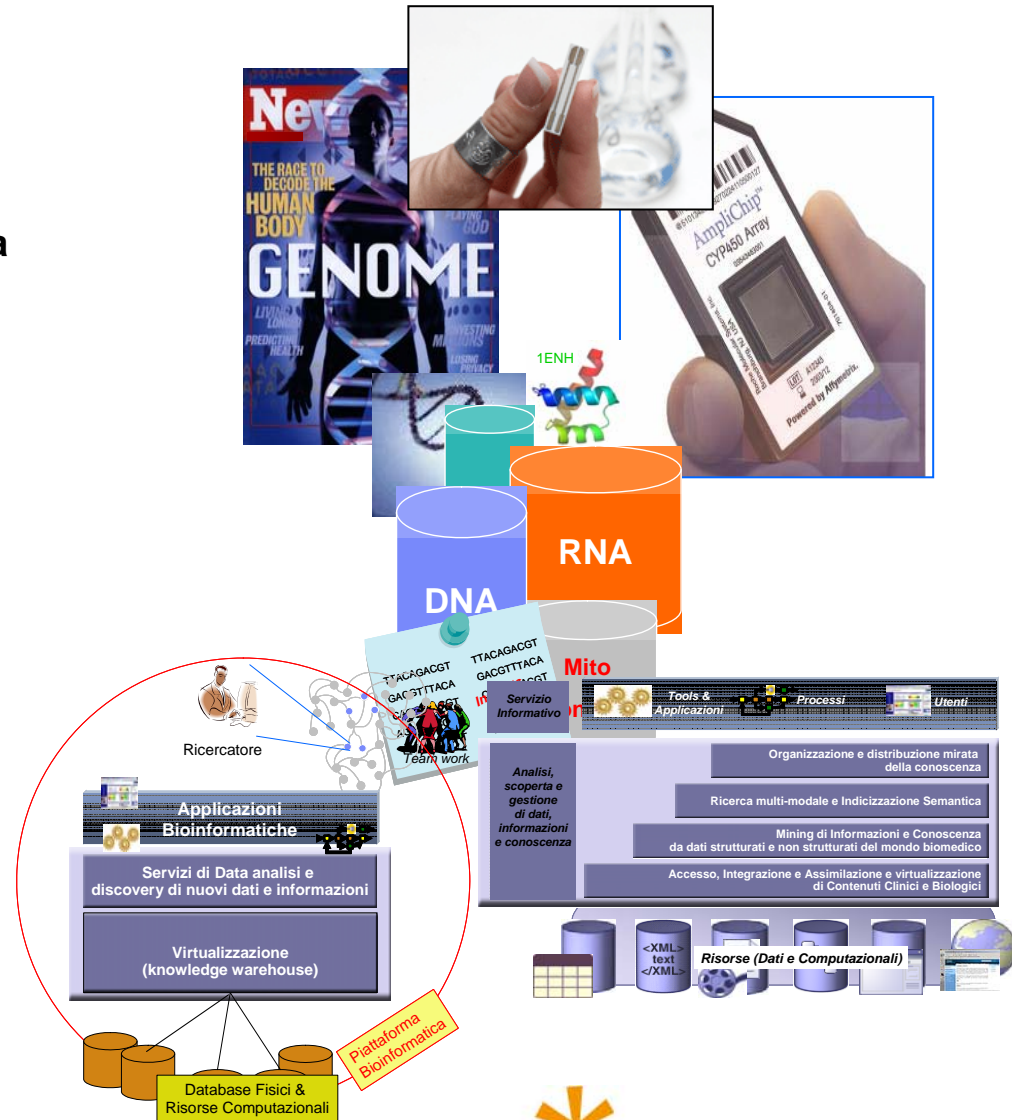
Temi

- 
 Le nuove sfide e i problemi affrontati dalla biologia moderna con l'avvento delle **informazioni genetiche**

- 
 La bioinformatica, che studia **come analizzare e gestire informazioni genetiche** per trarre nuova conoscenza sul mondo biologico

- 
 Anatomia di una **piattaforma applicativa** per supportare il lavoro dei bioinformatici

- 
 Il **progetto LIBI**: un esempio di piattaforma bioinformatica complessa



Io ho un problema da risolvere.....



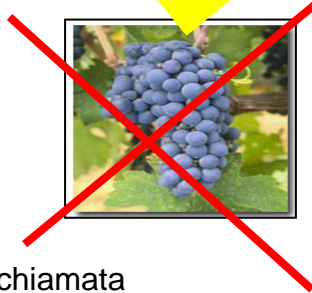
Ovviamente non è quello che state immaginando!

...il mio problema è un'altro: utilizzare le **informazioni genetiche** per analizzare la qualità del bicchiere di vino che sto bevendo!

Io vivo qui, a Bari, in Puglia

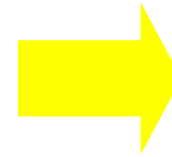


Questa varietà d'uva è chiamata **PRIMITIVO** ed è coltivata in questa zona di Puglia. Dai tempi dei romani è usata per produrre il **vino PRIMITIVO**

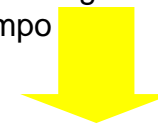


.. per far questo potremmo usare il "**DNA Barcode**" del PRIMITIVO. Una sorta di codice a barre genetico che individua la specifica varietà d'uva usata per produrre il vino!

Quest'altro grappolo è di una varietà chiamata **ZINFANDEL** e cresce sulle colline californiane. Da questa varietà si produce un vino molto simile al PRIMITIVO tant'è che i produttori californiani affermano che anche'essi producono il vino PRIMITIVO!



Il DNA Barcode del PRIMITIVO potrà quindi essere incapsulato all'interno di dispositivi di analisi a bassissimo costo chiamati **DNA chip**, che saranno in grado di effettuare analisi anche in tempo



Nell'**IBM Innovation Lab**, quando benviamo il PRIMITIVO, saremo finalmente sicuri di bere il nostro PRIMITIVO e non quello fatto con il vitigno Zinfandel!



I primi strumenti basati su test genetici cominciano a comparire: è il caso del CYP Test di Roche in grado di determinare la capacità di metabolismo di farmaci da parte di un individuo in base ad esami genetici condotti sui geni 2D6 e 2C19 del citocromo p450 implicati nel metabolismo di numerosi farmaci di larga prescrizione

Profili genetici analizzati

- CYP2D6: antidepressivi, antiemetici, antipsicotici, antiaritmici, beta-bloccanti e oppiacei
- CYP2C19: anticoagulanti, anticonvulsivi, anti-malarici, benzodiazepine, inibitori della pompa protonica.



*Categorie in cui viene
classificato il paziente in
base alla sua capacità
metabolica:*

1. Ultra-rapid
2. Extensive
3. Intermediate
4. Poor Metabolizer

http://www.roche-diagnostics.it/prodotti/laboratorio/Biologia_Molecolare/AmpliC.htm



Il progetto Genographic

Una partnership di 5 anni per ricostruire i principali flussi migratori dell'umanità

La ricerca riguarda sia popolazioni indigene che generici individui

E' la piu' ambiziosa iniziativa di antropologia genetica realizzata sino a questo momento

Decine di migliaia di Kit raccolta campioni e analisi genetiche effettuate



Kit Raccolta Campioni



www.genographic.com



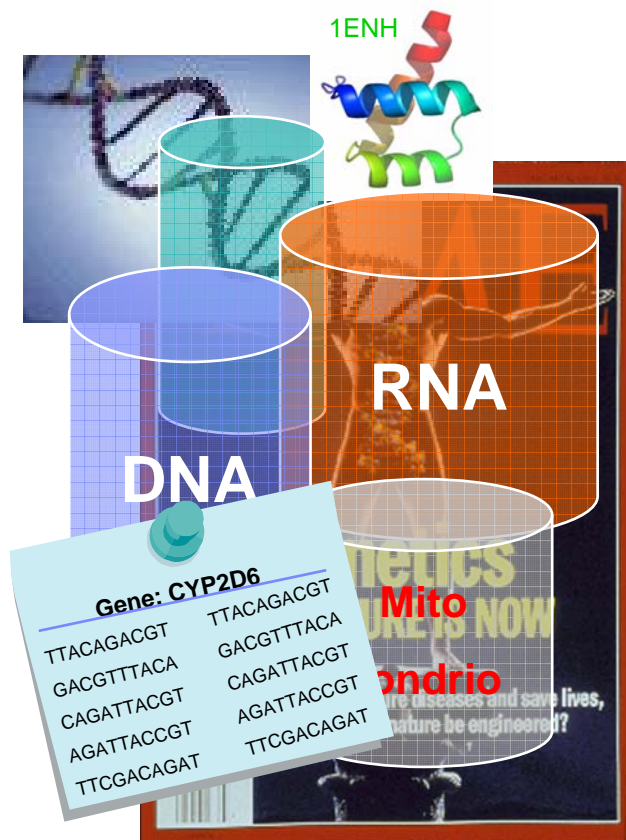
Evidentemente si pongono nuove frontiere etiche per tutti noi e nuove informazioni e dati su cui riflettere: il caso Deborah Lindner

Deborah Lindner è una donna di 33 anni che ha effettuato un test genetico ed ha scoperto di avere il un'alta probabilità di ammalarsi di cancro al seno nel corso della sua vita. Ha deciso di effettuare **una mastectomia preventiva** per ridurre il rischio di contrarre la malattia del 90%



Deborah Lindner al computer (foto dal New York Times)

Queste ed altre esperienze hanno avuto inizio dal sequenziamento del genoma umano che comincia ad influenzare la pratica medica e sarà sempre più influente nei prossimi anni

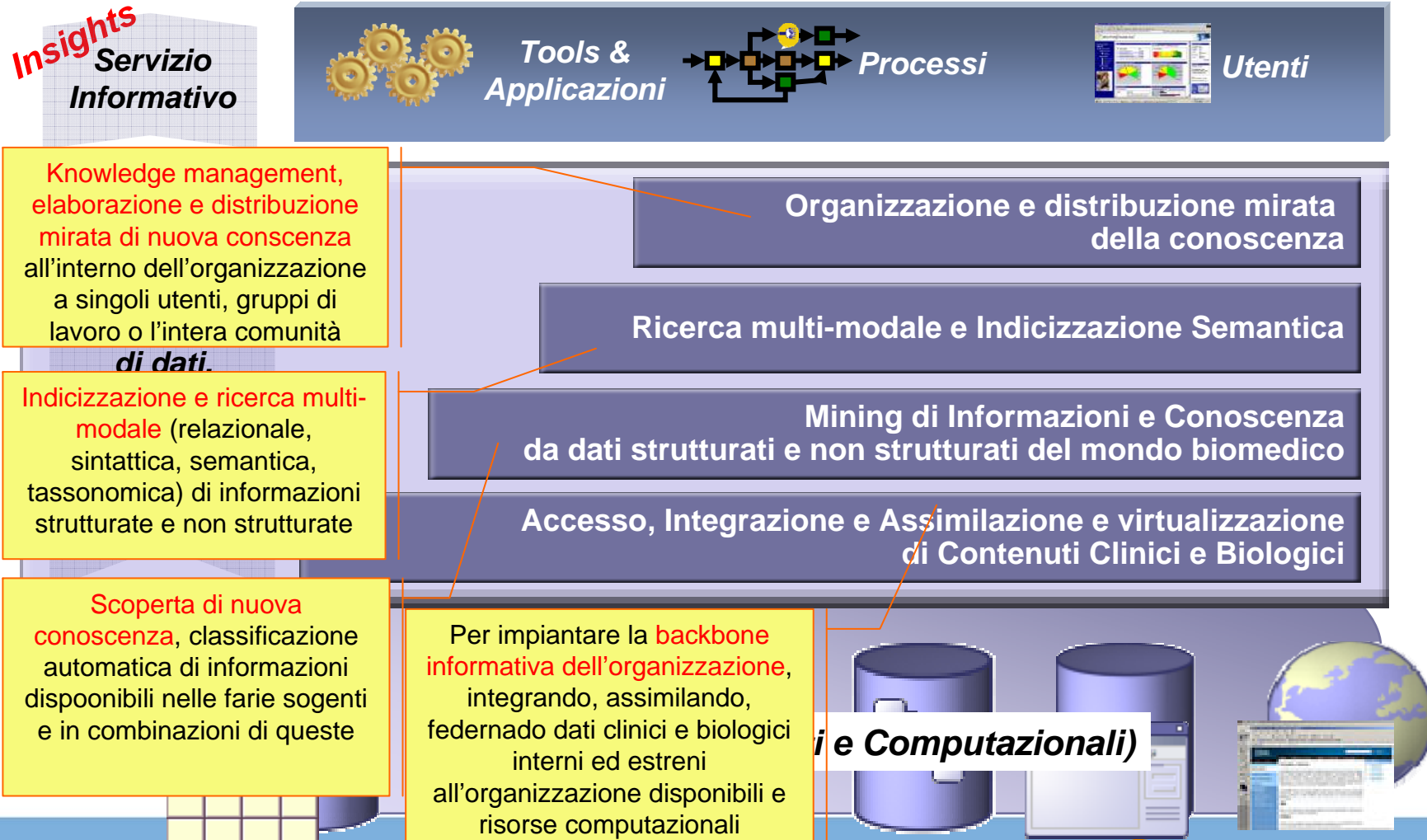


Uno dei più importanti motori dei sistemi biologici, e quindi della vita sulla terra, è **l'informazione**. Questa è contenuta, codificata e trasmessa in svariati modi.

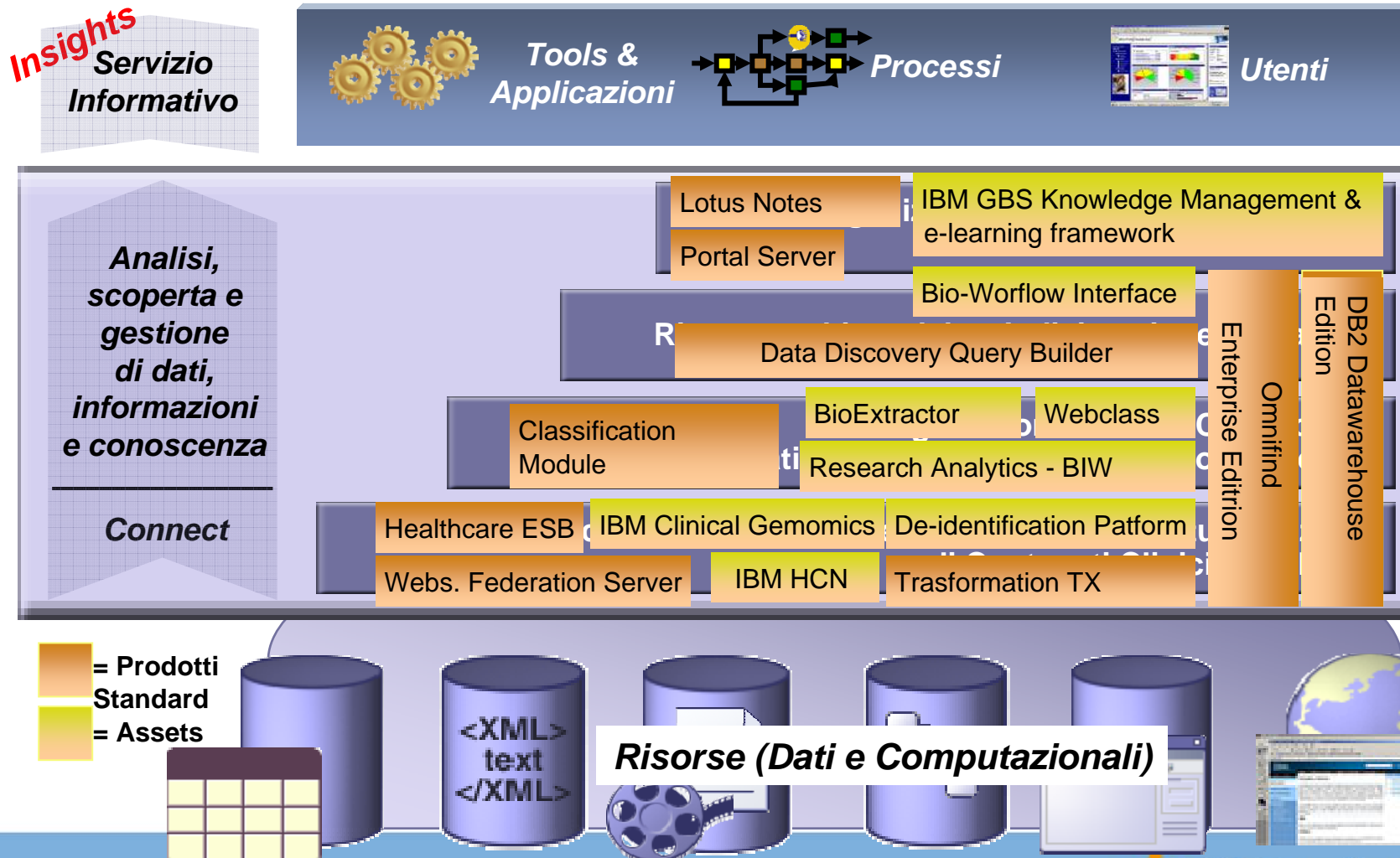
L'informazione è contenuta in molti **"database"**, il più importante è il **DNA**, replicato in ogni cellula, contiene le istruzioni base (note come Geni) che governano la struttura e il funzionamento di ogni organismo.

Ogni specie ha delle peculiarità che si connotano in una **"diversità molecolare"**, vale a dire possiede contenuti caratteristici nel proprio **"database"** che possono essere usati per caratterizzare e distinguere le specie fra loro. La diversità molecolare è uno dei più potenti strumenti usati dalla bioinformatica per realizzare metodi di analisi

Dal punto di vista IT supportare il lavoro dei bioinformatici corrisponde alla realizzazione di complesse piattaforme applicative in grado di trattare e fornire informazioni genetiche a diversi livelli e per differenti scopi di utilizzo











La realizzazione di piattaforme bioinformatiche è un'attività complessa dal punto di vista dell'ingegneria informatica. Occorre non solo disporre di innumerevoli middleware e prodotti ma molte componenti occorre svilupparle ad hoc



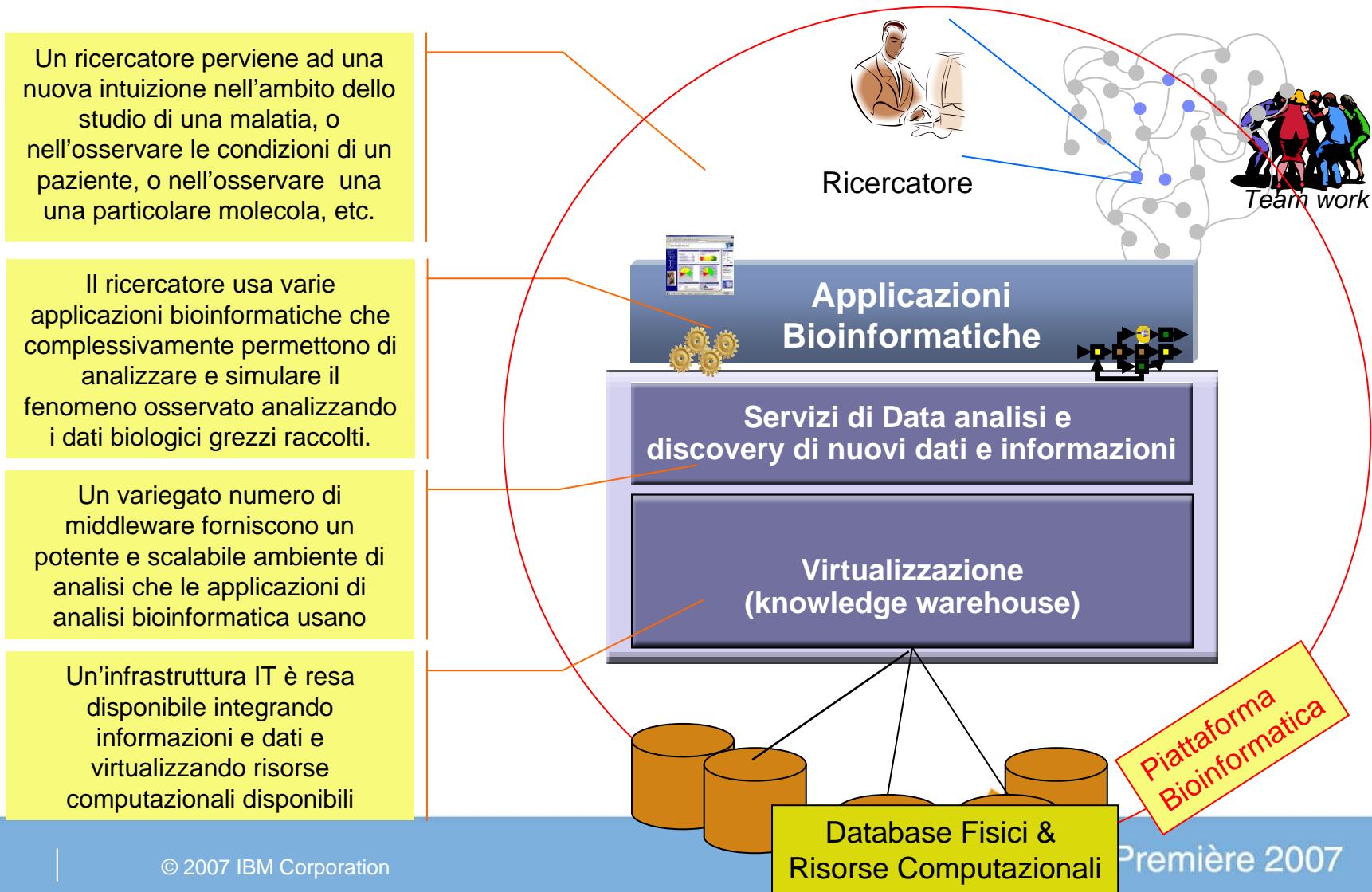
Progetto LIBI: Laboratorio Internazionale di Bioinformatica

- **LIBI è un laboratorio di Bioinformatica e Biologia Computazionale** rivolto sia ad aspetti di ricerca di base che applicata.
- **Il LIBI è concepito come Laboratorio, "without walls"; costituisce uno spazio di lavoro virtuale** per i partner che forniscono le proprie infrastrutture e utilizzano gli strumenti ed i macchinari del laboratorio, usufruendo delle capacità e competenze messe a disposizione dagli altri partner.
- **Il LIBI si propone come una unità organizzativa in grado di concentrare e potenziare competenze multidisciplinari** e piattaforme tecnologiche adeguate e offrire un insieme di servizi specializzati alla comunità bioinformatica nazionale e internazionale
- **LIBI è un progetto FIRB**, co-finanziato dal Ministero dell'Università e della Ricerca

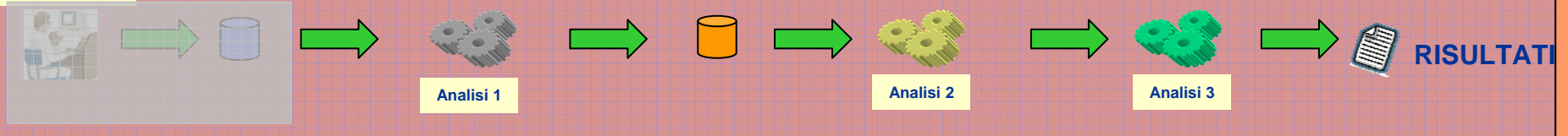


	▪ Istituto di Tecnologie Biomediche-Sezione di Bari, (CNR-ITB) – Coordinatore
	▪ IBM –Innovation Lab - Bari
	▪ Istituto di Nazionale di Fisica Nucleare, (INFN)
	▪ CINECA
	• Università di Lecce, (SPACI)
	• Università di Trieste (CBMTS)
	• Università di Milano – Dipartimento di biologia, (UNIMI-BI)
	• Dipartimento di Biologia Molecolare, Università degli Studi di Bologna (UNIBO-BM)

Il progetto LIBI si pone l'obiettivo di realizzare una piattaforma di lavoro virtuale rivolta alla comunità di ricercatori bioinformatici



WORKFLOW



e-workplace – Piattaforma LIBI

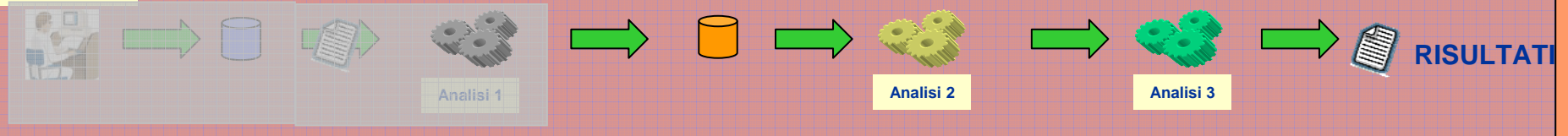


Un biologo avvia un complesso **workflow di analisi bioinformatica** operando con il Portale LIBI fornendo i dati di input e attendendo per ottenere i risultati. L'analisi richiesta consiste di 3 complessi step che verranno realizzati in modo distribuito sulla piattaforma LIBI.

Il primo passo è una **una query per reperire i dati biologici oggetto dello studio dal nodo LIBI Data Grid**, gestito dall'IBM Innovation Lab di Bari e che federa innumerevoli banche dati biologiche.



WORKFLOW



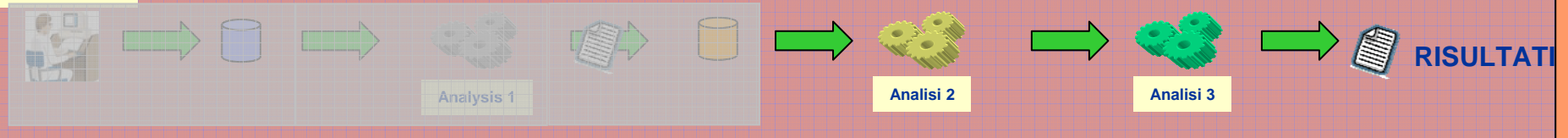
e-workplace – Piattaforma LIBI



Il primo step di analisi richiede l'esecuzione di specifici algoritmi bioinformatici che opereranno su un **cluster standard consistente di 154 G4 bi-processor**i, ospitato presso la server farm dell'istituto SPACI (Southern Partnership for Advanced Computational Infrastructures) di Lecce.



WORKFLOW



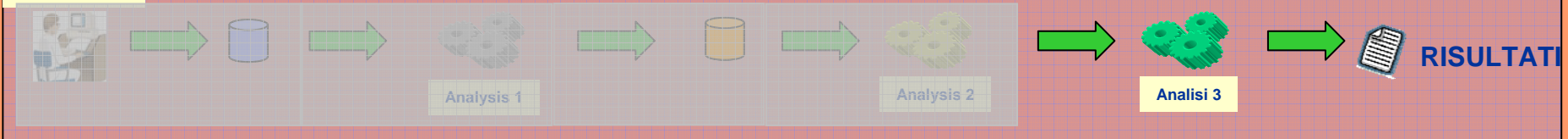
e-workplace – Piattaforma LIBI



Parte dei risultati prodotti dal primo step di analisi sono quindi utilizzati per effettuare **una nuova query al nodo LIBI Data Grid DB**, ospitato presso la server farm dell'IBM Innovation Lab di Bari. Ulteriori dati biologici sono estratti dalle banche dati federate e verranno analizzati durante gli step di analisi successivi



WORKFLOW



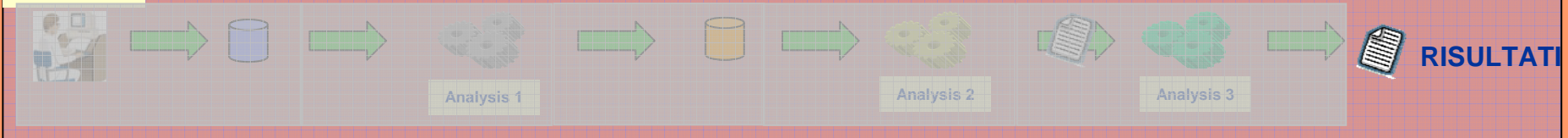
e-workplace – Piattaforma LIBI



Lo step di analisi 2 richiede un'infrastruttura di **High Performance Computing**. Questo task viene eseguito da un cluster HPC (IBM SP Cluster 1600 consisting of 512 CPUs) ospitato presso la server farm del CINECA (Interuniversity Consortium, the largest Italian computing centre) a Bologna



WORKFLOW



e-workplace – Piattaforma LIBI



Il terzo step di analisi prevede l'esecuzione di programmi lanciati su un ambiente di **grid computing**, in questo caso viene usata la griglia EGEE, gestita dall'INFN (Italian National Institute for Nuclear Physics).

Il risultato del workflow sono a questo punto disponibili e vengono proposti al biologo attraverso il portale web LIBI



Conclusioni

- **L'IT pervade permette di affrontare adeguatamente le moderne sfide di gestione di informazioni e conoscenza che complessi domini applicativi pongono, come quello della biologia e della medicina moderna**
- **La Bioinformatica è un esempio di dominio applicativo complesso in cui sono richieste sofisticate tecnologie per gestire e trattare informazioni genetiche**
- **La realizzazione di piattaforme bioinformatiche richiede l'uso di prodotti software comuni a tanti domini applicativi combinati con asset e sviluppi ad-hoc**
- **Sono stati brevemente richiamati gli obiettivi del progetto LIBI che punta a realizzare una complessa piattaforma bioinformatica. Progetto FIRB co-finanziato dal MUR e guidato dal CNR che coinvolge i più importanti gruppi di ricerca italiani in bioinformatica**



