E-BioSci

E-BioSci

Semantic networks of biological information

Les Grivell
European Molecular Biology Organisation

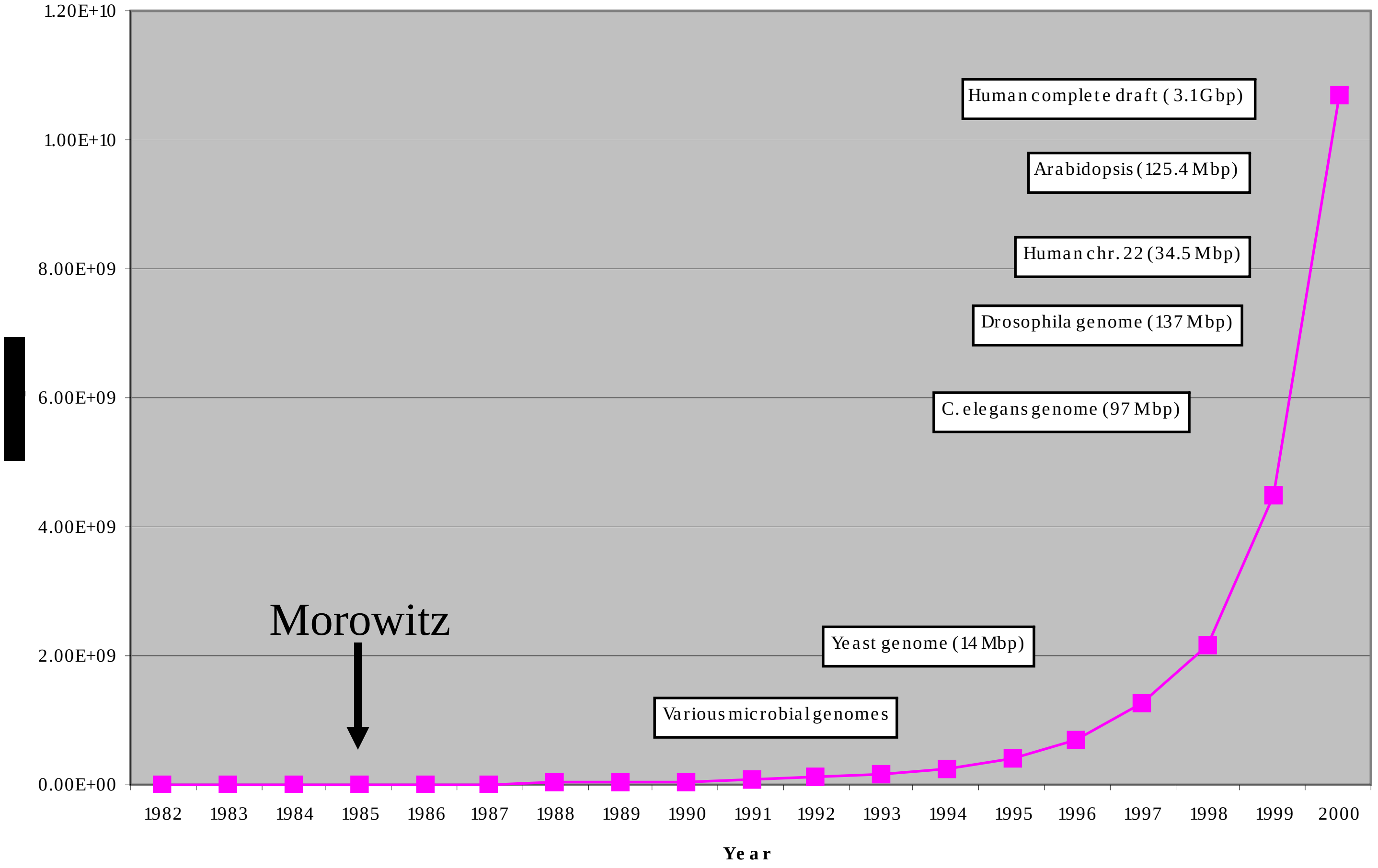
EMBC

EMBO

“Biological research has reached a point where new generalizations and higher order biological laws are being approached, but may be obscured by the simple mass of data”

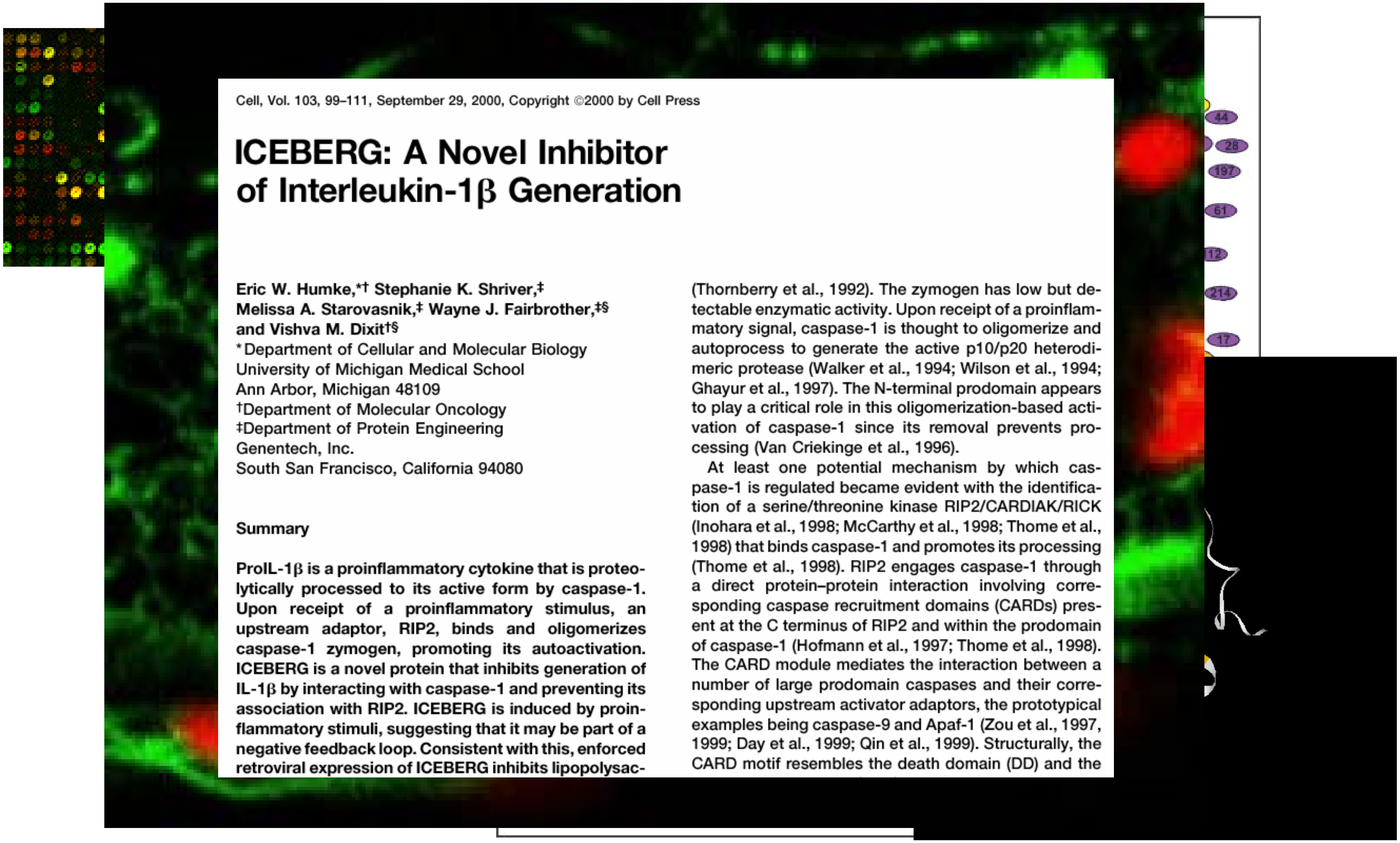
Harold Morowitz, 1985
Report to the U.S. National Academy of Sciences

One part of the information explosion ….



[redacted]

Sequences are not the only form of digital information



Cell, Vol. 103, 99–111, September 29, 2000, Copyright ©2000 by Cell Press

ICEBERG: A Novel Inhibitor of Interleukin-1β Generation

Eric W. Humke,*† Stephanie K. Shriver,‡
Melissa A. Starovasnik,‡ Wayne J. Fairbrother,‡§
and Vishva M. Dixit†§
*Department of Cellular and Molecular Biology
University of Michigan Medical School
Ann Arbor, Michigan 48109
†Department of Molecular Oncology
‡Department of Protein Engineering
Genentech, Inc.
South San Francisco, California 94080

Summary

ProIL-1β is a proinflammatory cytokine that is proteolytically processed to its active form by caspase-1. Upon receipt of a proinflammatory stimulus, an upstream adaptor, RIP2, binds and oligomerizes caspase-1 zymogen, promoting its autoactivation. ICEBERG is a novel protein that inhibits generation of IL-1β by interacting with caspase-1 and preventing its association with RIP2. ICEBERG is induced by proinflammatory stimuli, suggesting that it may be part of a negative feedback loop. Consistent with this, enforced retroviral expression of ICEBERG inhibits lipopolysac-

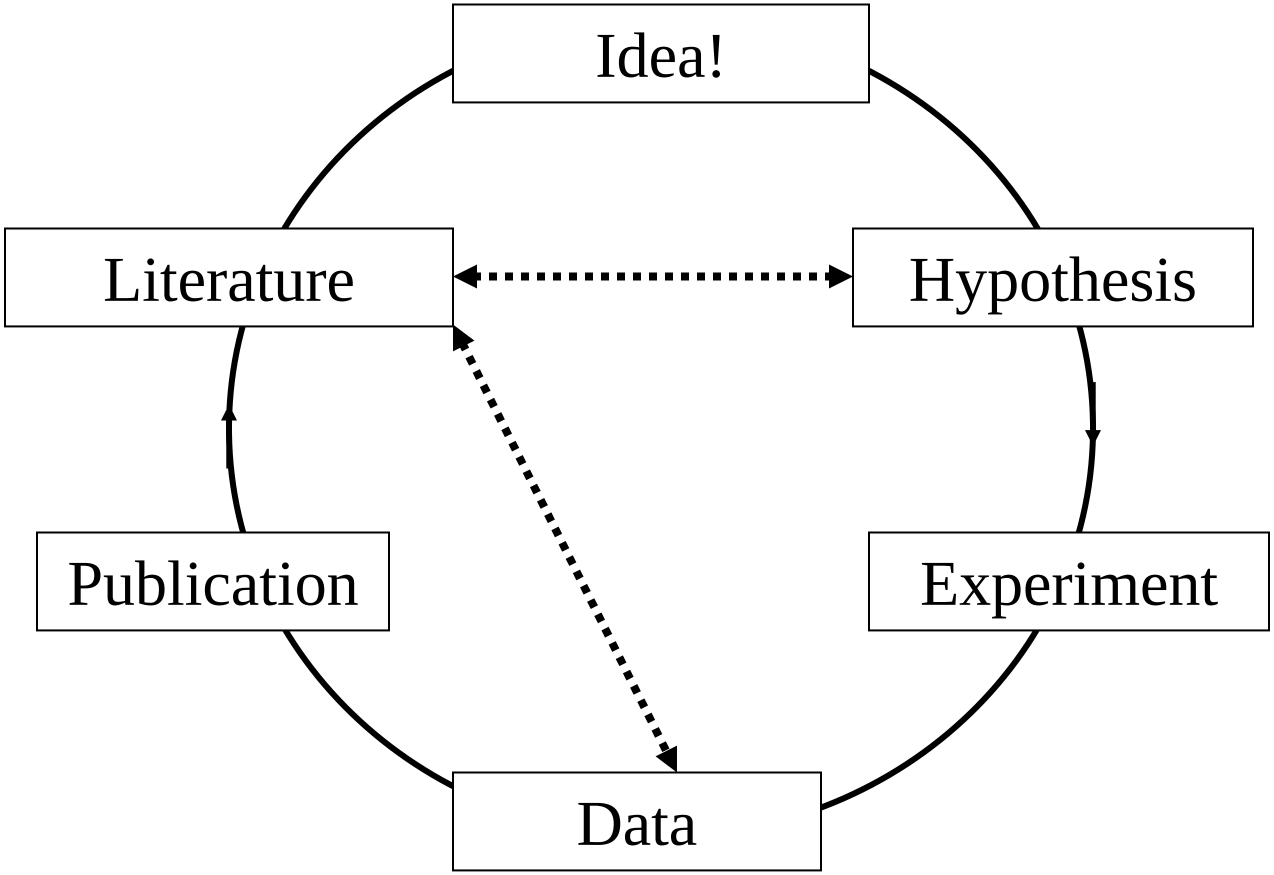
(Thornberry et al., 1992). The zymogen has low but detectable enzymatic activity. Upon receipt of a proinflammatory signal, caspase-1 is thought to oligomerize and autoprocess to generate the active p10/p20 heterodimeric protease (Walker et al., 1994; Wilson et al., 1994; Ghayur et al., 1997). The N-terminal prodomain appears to play a critical role in this oligomerization-based activation of caspase-1 since its removal prevents processing (Van Criekinge et al., 1996).

At least one potential mechanism by which caspase-1 is regulated became evident with the identification of a serine/threonine kinase RIP2/CARDIAK/RICK (Inohara et al., 1998; McCarthy et al., 1998; Thome et al., 1998) that binds caspase-1 and promotes its processing (Thome et al., 1998). RIP2 engages caspase-1 through a direct protein–protein interaction involving corresponding caspase recruitment domains (CARDs) present at the C terminus of RIP2 and within the prodomain of caspase-1 (Hofmann et al., 1997; Thome et al., 1998). The CARD module mediates the interaction between a number of large prodomain caspases and their corresponding upstream activator adaptors, the prototypical examples being caspase-9 and Apaf-1 (Zou et al., 1997, 1999; Day et al., 1999; Qin et al., 1999). Structurally, the CARD motif resembles the death domain (DD) and the

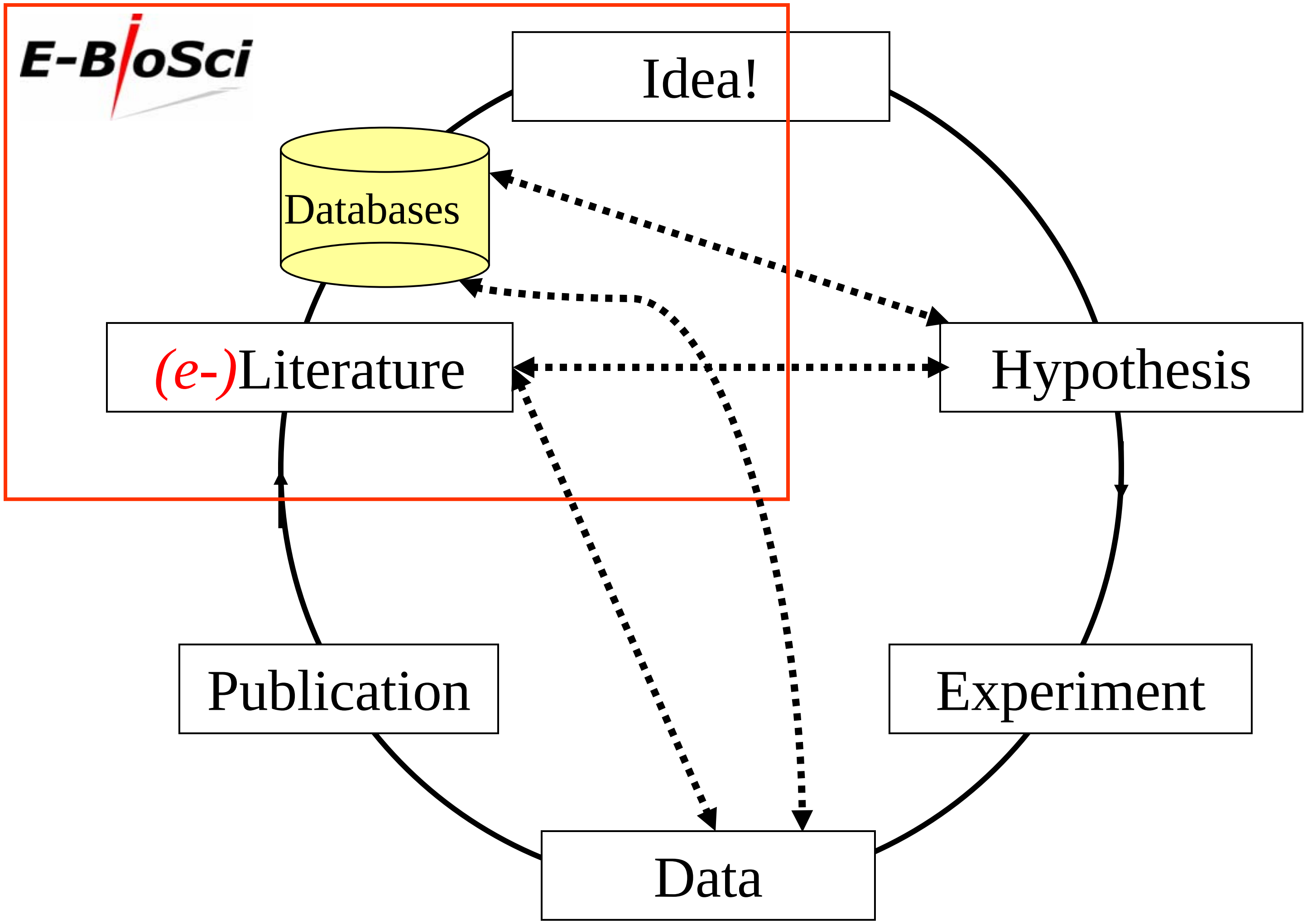
The digital revolution is ….

- driving dramatic changes in the way research is conducted and in the way knowledge is being generated
- exposing limitations in current mechanisms for disseminating and sharing information
- driving changes in publication strategies

The knowledge cycle (traditional)



The knowledge cycle (extended)



E-BioSci

Idea!

Databases

(e-)Literature

Hypothesis

Publication

Experiment

Data

Biological information: current reality …..

- Hundreds of different databases, many in flat-file format
- Non-uniform or lack of external identifiers
- Lack of interoperability at the level of syntax and semantics
- Knowledge scattered across the literature in many thousands of non-computer readable journal articles

Information retrieval from text

- There is increasing need for the use of literature as a computer-readable resource (intelligent search / retrieval)
- There is a need to apply computational methods to text as well as data analysis (mining / analysis; literature as discovery resource)
- There is a need for new methods of integration and visualization of this information
- There is a need for a scale-up in the rate of *curated* database growth through integration with the literature

Dealing with the data deluge: what do researchers need?

- Intuitive interfaces to web resources
- Advanced search / retrieval facilities
- Smooth navigation from one resource to another
- Immediate availability of authoritative information; free of charge at the point-of-use
- Information that can be integrated / manipulated / visualized /output in another form

E-BioSci

A new information service for the life sciences that will interlink factual and image data repositories with the research literature

EU Quality of Life research infrastructure:
platform under construction

Closely linked to ***ORIEL****, an EU-funded research project in information technology*

EMBO

The current E-BioSci - ORIEL partnership

- Distributed network of information resources
- Europe-based; world-wide role

The platform

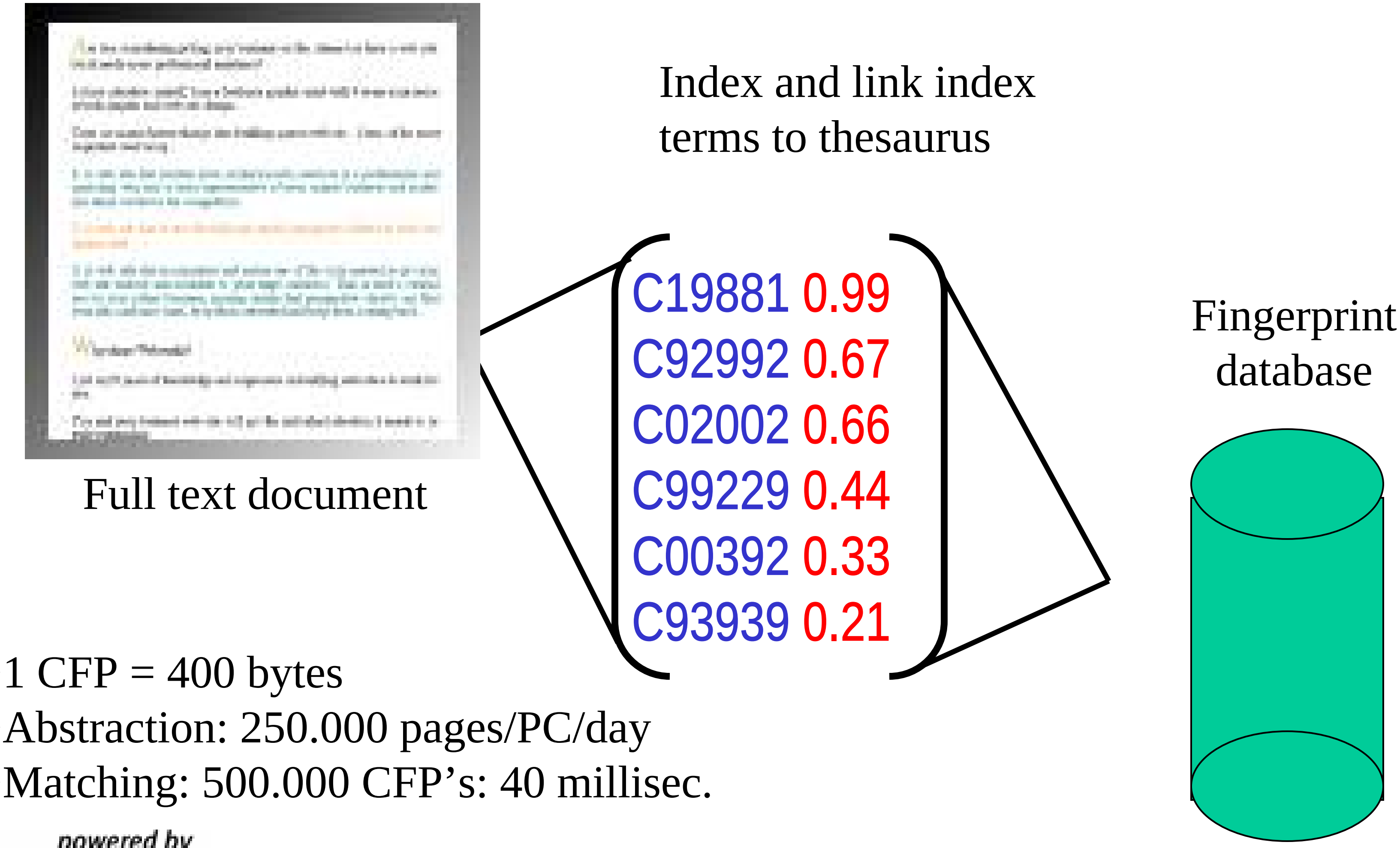
- Services freely accessible to the academic research community
- Set of *distributed* resources (literature, sequence- and image- databases)
- Full-text search
    - across document repositories
    - using cross-language queries (e.g. English – French, - German etc)
    - 2-way navigation links between literature and molecular datasets via gene symbol recognition

Features implemented via conceptual fingerprinting

A discovery tool

powered by Collexis®

Conceptual fingerprints



Full text document

Index and link index terms to thesaurus

C19881 0.99
C92992 0.67
C02002 0.66
C99229 0.44
C00392 0.33
C93939 0.21

Fingerprint database

- 1 CFP = 400 bytes
- Abstraction: 250.000 pages/PC/day
- Matching: 500.000 CFP’s: 40 millisec.

powered by Collexis®

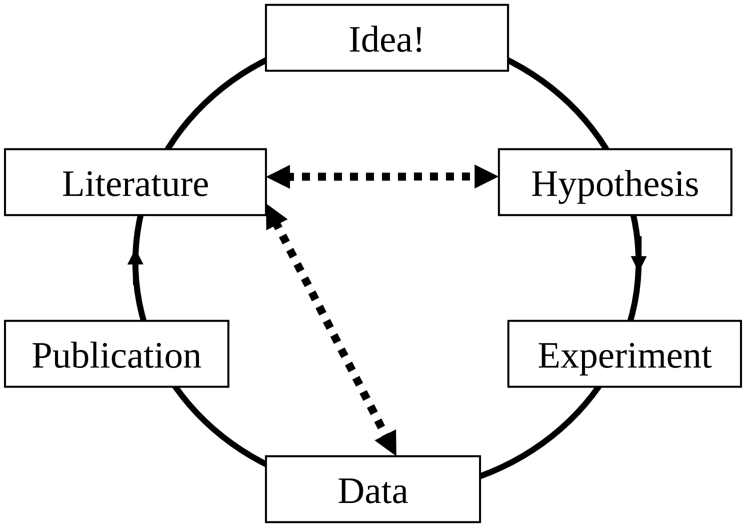
E-BioSci – distinguishing features

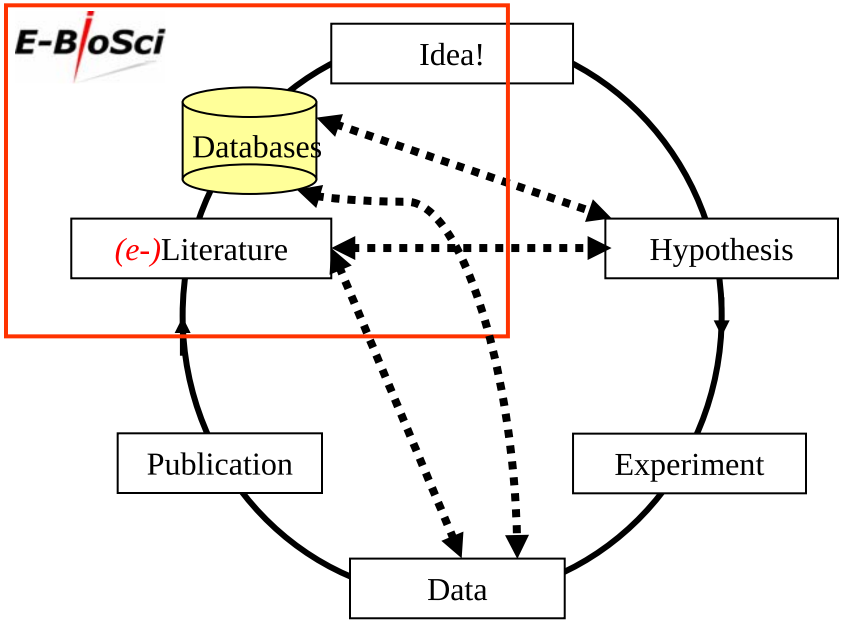
- Is a free academic access to services
- Places strong emphasis on concept searching of full text – a discovery tool
- Uses conceptual fingerprints to semantically link text with different data types (in particular genomic and image data)
- Links only to refereed material that meets criteria of editorial control
- Welcomes principles of free access, but respects existing restrictions of (commercial) content providers

Community-driven *e*-science (1)

- The growing importance of interactive databases

Traditional → Extended

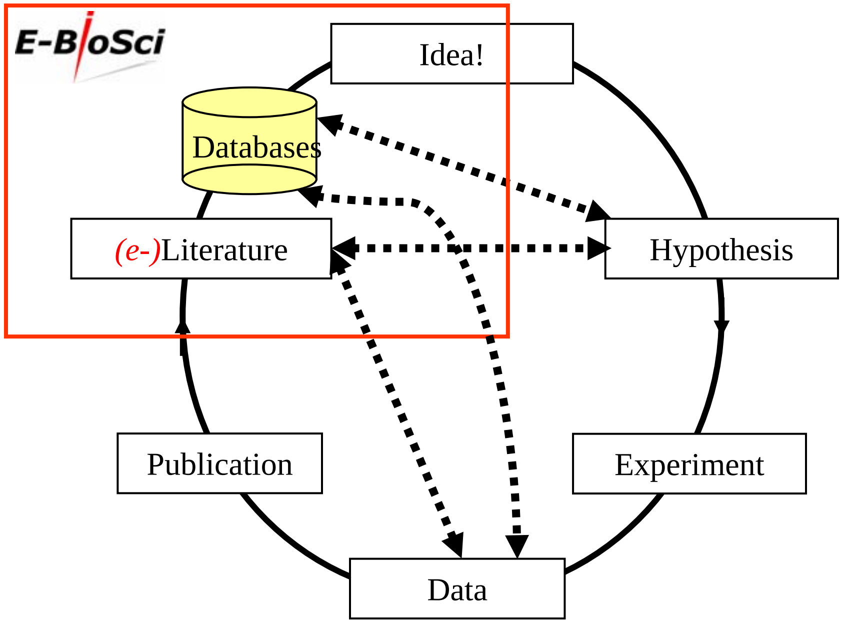




Community-driven *e*-science (2)

- Large-scale projects will drive further changes in communication and publishing practice

‘Traditional’ → Community ‘data-centric’



E-BioSci

Idea!

Databases

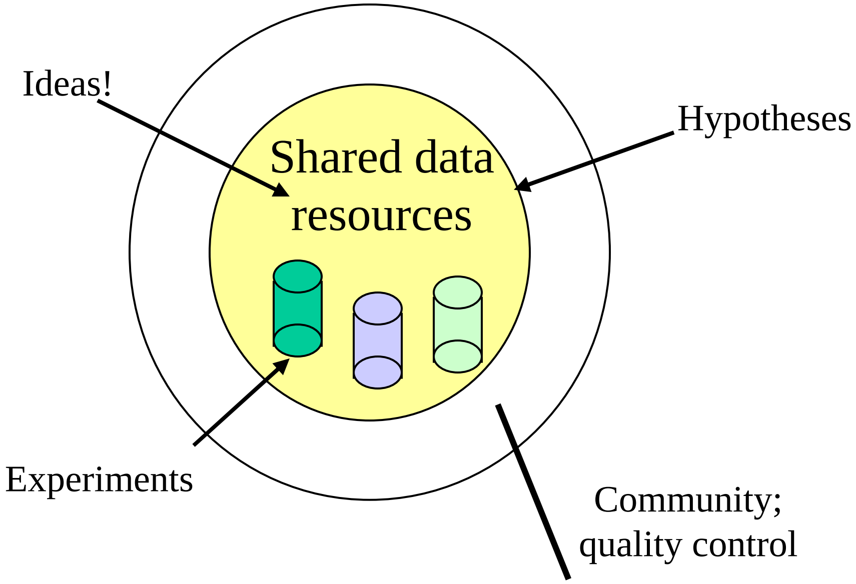
*(e-)*Literature

Hypothesis

Publication

Experiment

Data



Ideas!

Hypotheses

Shared data resources

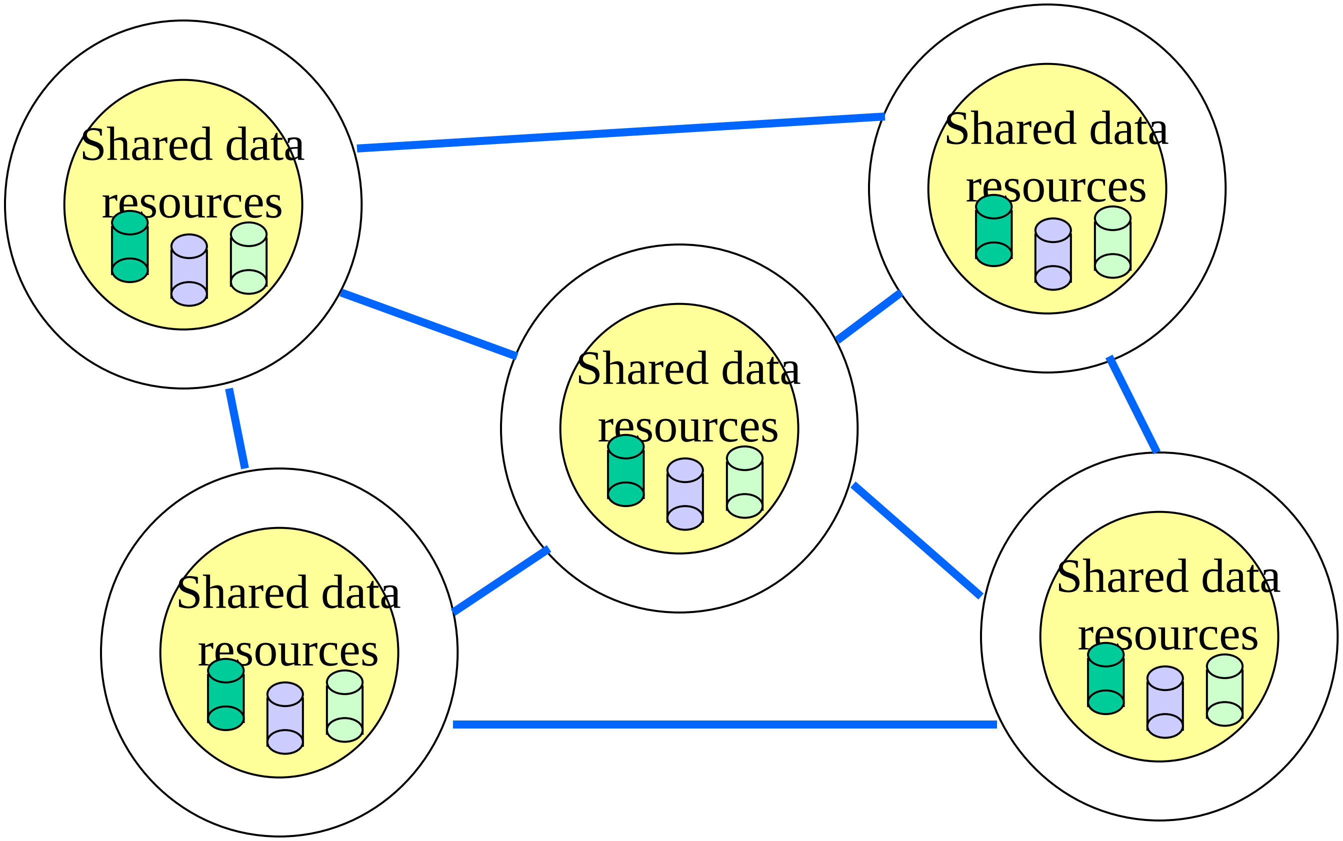
Experiments

Community; quality control

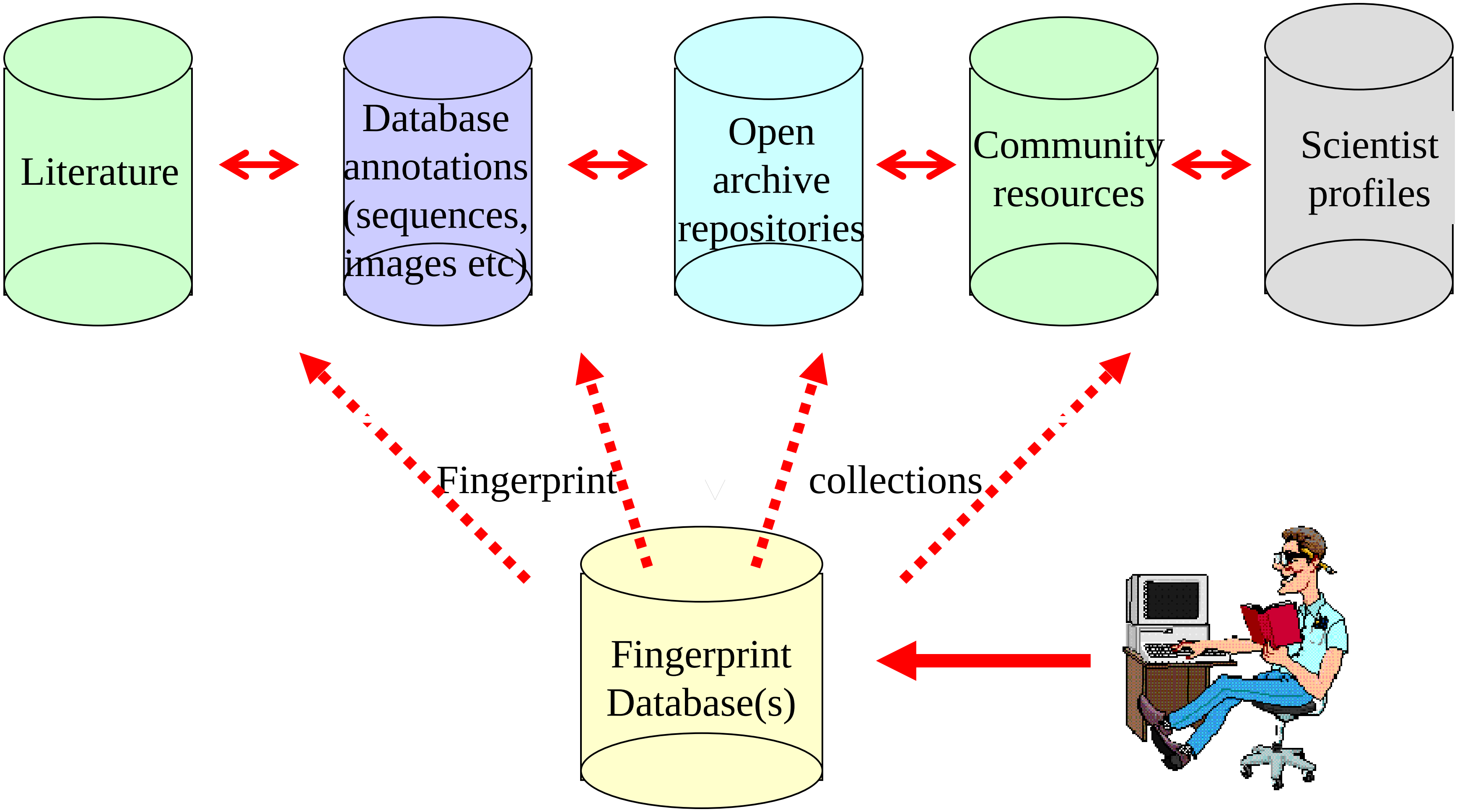
Community-driven *e*-science (3)

- Proliferation of community knowledge networks raises new challenges of semantic interconnectivity

E-BioSci



E-BioSci and semantic interconnection of searchable resources



Literature

Database annotations (sequences, images etc)

Open archive repositories

Community resources

Scientist profiles

Fingerprint collections

Fingerprint Database(s)

Acknowledgements

- Frank Gannon, Executive Director EMBO
- … and many others who contributed ideas to the concept of E-BioSci
- The E-BioSci partners
- European Commission
  (contracts QLRI-2001-30266 and IST-2001-32688)