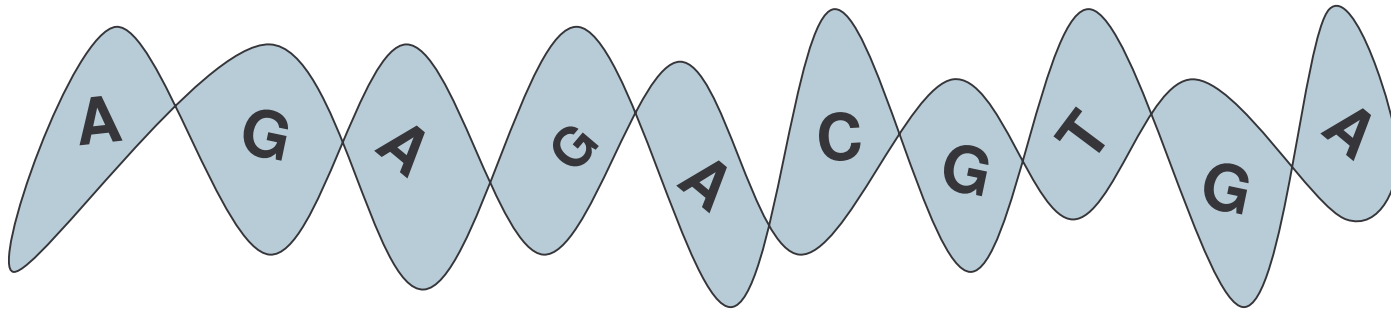


B
i
o
i
n
f
o
r
m
a
t
i
c
s

P
e
r
-

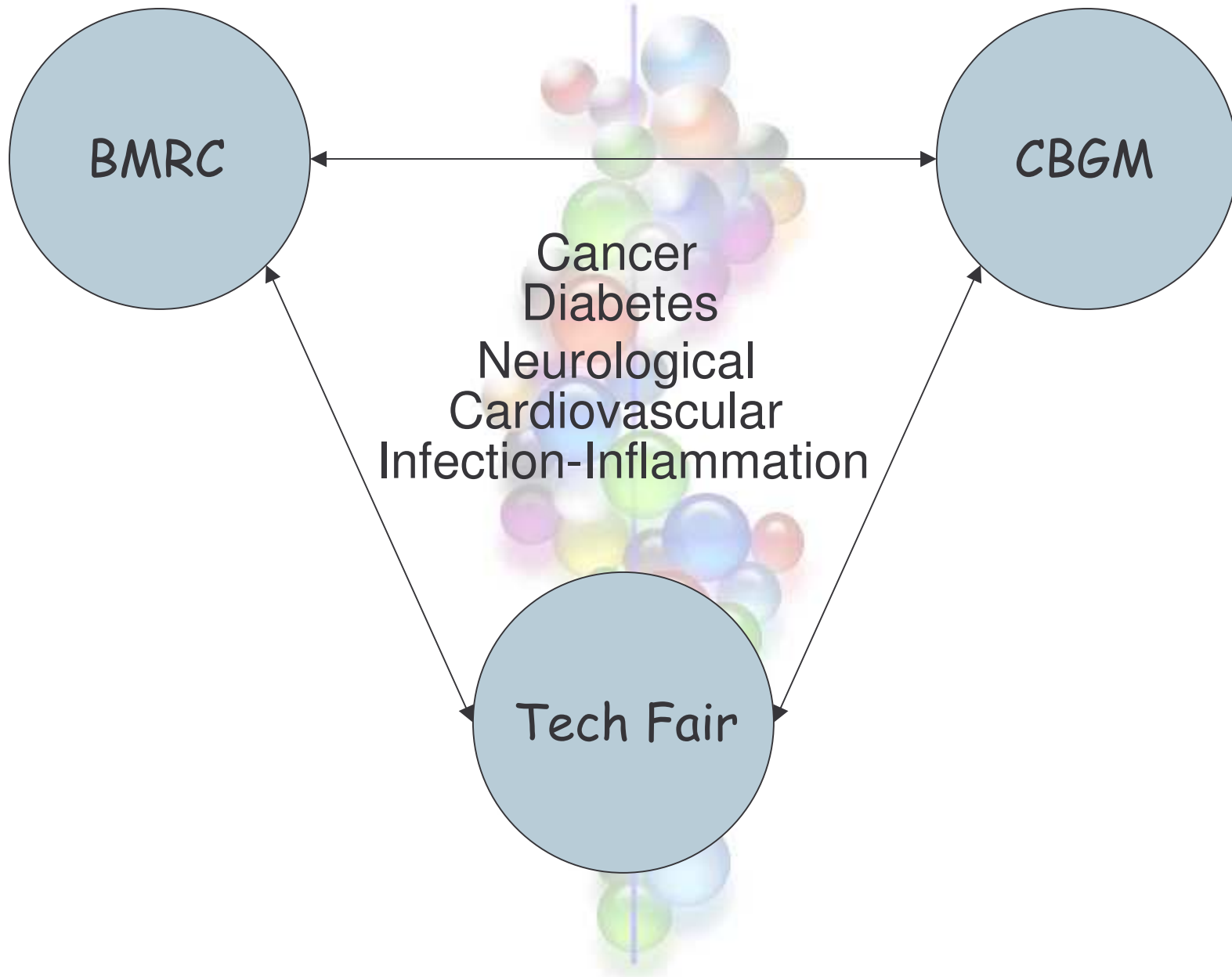


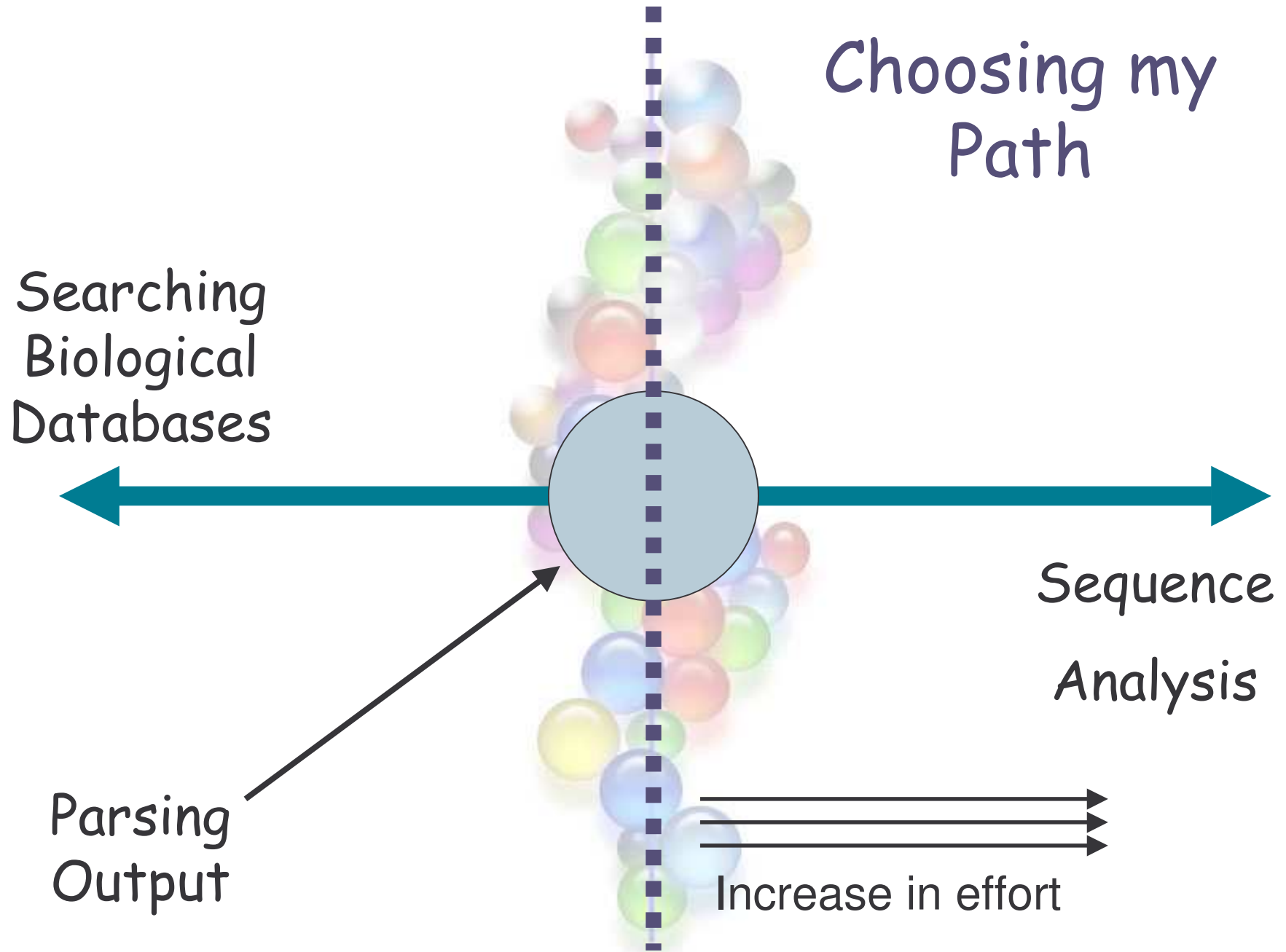
Douglas James Joubert

Digital Information
Librarian

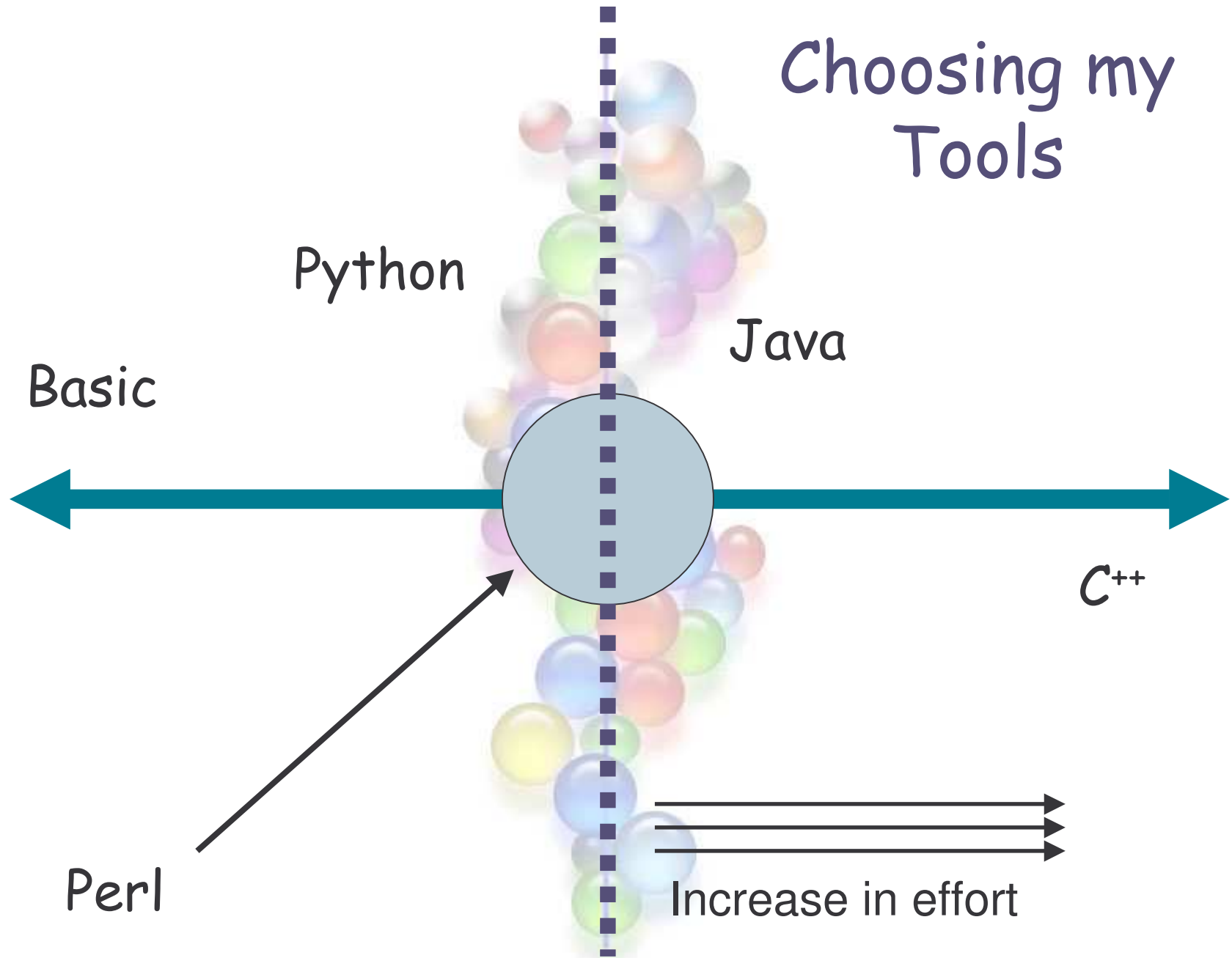
Medical College of Georgia







Choosing my Tools



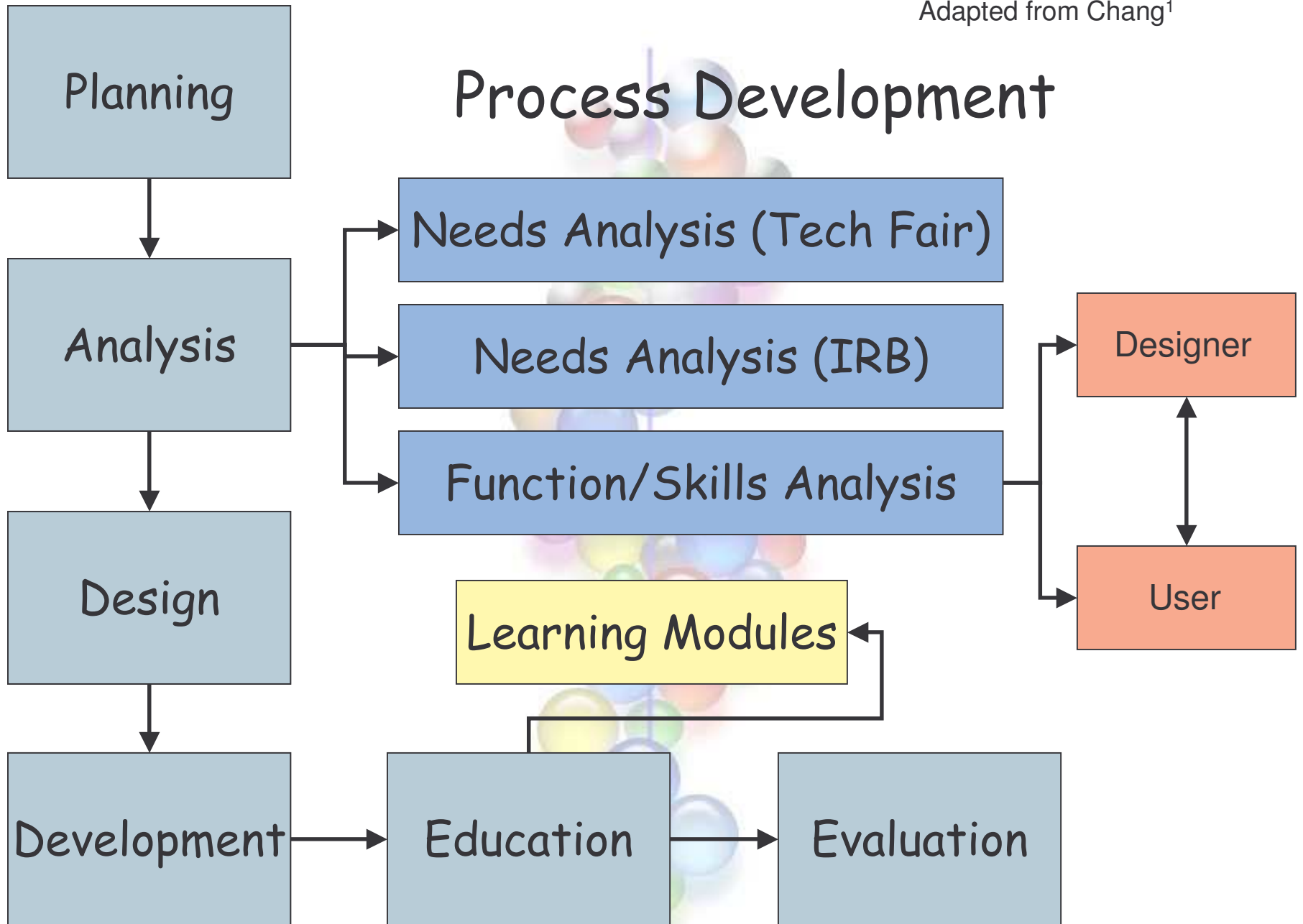
Goals



- To develop and foster a transferable skill base
- To increase visibility of the library in terms of knowledge and personnel resources
- Depending on the success of this project, filter this information to other librarians
- Is this possible with current staffing and resources?

Adapted from Chang¹

Process Development



Process Development



- Leysens defines six project patterns that develop in Bioinformatics
- My project shared many characteristics with **Visionary** classification²
 - Single individual
 - Creativity limited by time and resources
 - Must be strong advocate of vision and the management of resources

Perls of Wisdom



- Open source and free
- “Relatively” easy to learn
- Extensive use of regular expressions
- Rapid prototyping
- Portable (relative→BioPerl)
- Fairly easy to maintain
- Flexible (modules)

Perl-Tools



- Perl
 - CPAN
 - ActiveState
- Solid Text Editor
 - Notepad
 - Arachnophilia (free but now Java-based)
 - Perl Builder (not free and not cheap)

Domains of Knowledge

Run Perl	Scalars	Arrays
Hashes	Control Structures	Subroutines
Perl Operators	File Manipulation	Reg Expressions



BioPerl

Design Phase

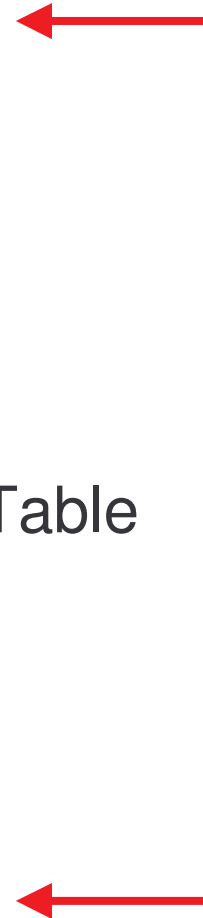


- Programming shares many characteristics of the research process
 - Identify the problem
 - Separate the problem into simpler components
 - Evaluate your needs
 - Selecting the appropriate data and input characteristics
 - Critically evaluate your results

New Data—New Formats⁴

- FASTA
 - NCBI Seq-Id Syntax (database specific)
- GenBank/EMBL/DDBJ
 - Flat-files
- SWISS-PROT
 - Flat-files, Field Definitions, and Feature Table
- Pfam
 - Flat-file, and field definitions
- PROSITE
 - Flat-file, and field definitions

U
n
i
f
o
r
m
a
l
i
t
y



To BioPerl, or not?



- Based on modules, which are one way Perl handles external programs
- BioPerl automates a number of tasks with the 500+ modules
 - Alignments
 - Databases
 - Features and genes on sequences
 - Sequences

To BioPerl, or not?



- Things to consider
 - Object-oriented
 - Installation is not as intuitive, must use ppm
 - Open source so in constant development
 - Script transferability between systems (heavily UNIX-based)
 - Documentation is spotty
 - Size is overwhelming

User Input



- Developed survey for second phase of project (waiting IRB approval)
 - Demographic data
 - Databases and NT Resources
 - Databases and PROT Resources
 - Software
 - Training (library initiated)
- If project goes to second phase, evaluation will occur in the form of qualitative and quantitative survey.

Education



- Developed learning modules using Qarbon Viewlett Builder
- Done to reinforce basic concepts and expedite the learning of repetitive tasks
- Able to deliver content on demand at anytime
- Files are accessible via the web in *swf* format→no Flash learning curve



Perl Builder Viewlet

Print Resources



- Pierce, C. (2002). Sams teach yourself Perl in 24 hours. Sams Publishing.
- Tisdall, J. (2001). Beginning Perl for bioinformatics. O'Reilly & Associates.
- Claverie, J.M., & Notredame, C. Bioinformatics for dummies. Wiley Publishing Inc.
- Schwartz, R. (2003). Perl objects, references, & modules. O'Reilly & Associates.

Print Resources



- Nucleic Acids Research January and July Issues
- Brooks, F. (1995). The mythical man-month. Addison-Wesley.
- Gibas, C., & Jambeck, P. (2001). Developing bioinformatics computer skills. O'Reilly & Associates.
- Markel, S., & Leon, D. (2003). Sequence analysis in a nutshell. O'Reilly & Associates.

Bibliography



1. Chang, C.C. (2003). *An electronic performance support system for learning and doing instructional design tasks*. Educational Technology, 43(4), 46-51.
2. Leysens, M. (2003). *Patterns in bioinformatics project management*. Paper presented at the Bioinformatics Technology Conference.
3. Gery, G. (1991). *Electronic performance support systems*. Cambridge, MA: Ziff Communications.
4. Markel, S., & Leon, D. (2003). *Sequence analysis in a nutshell*. Sebastopol, CA: O'Reilly & Associates.