

Research information sharing... and information synthesis skills

Contents

Introduction	2
What is the sharing of research information?	2
Agenda	2
Relevance	3
Part 1. "Genome Race"	3
Background	3
Benefits of knowing the human genome sequence	4
The race.....	4
About the public effort	5
Celera's effort	5
The pace of sequencing	6
Sharing the human genome sequence data	6
So who won the genome sequencing race?	7
What happened in the end?	8
What does this genome sequencing race show us?.....	8
Part 2. Open access publishing	9
What is open access publishing?	9
Sample open access publications	9
Value of open access publishing	10
Part 3. Synthesis skills	11
Definition	11
Motivation.....	11
Kinds of synthesis.....	12
Cultivating synthesis skills.....	13
Summary	13
References	13

Introduction

SLIDE 2

This is a guest lecture for Wake Forest University's LIB 100 class, Information Literacy. (November 18, 2010)

Good morning, my name is Jeffery Loo, and I am a librarian at the University of California, Berkeley.

It is an honor to speak with you today. I have visited the Wake Forest campus many times, and am always struck by the unique environment you are in. Wake Forest is like the best of both worlds. It has the sense of community and accessibility of a small liberal arts college, but at the same time there are research and learning opportunities in areas like law, business, and medicine that are comparable to larger universities. If I were to describe Wake Forest to someone, I would fondly describe it as cozy and cosmopolitan.

What is the sharing of research information?

SLIDE 3

Today we'll be discussing the sharing of research information. I have a very loose and general definition of sharing. When you share, you are letting others see your research work, use it, and build upon it.

Agenda

SLIDE 4

There are three goals in our discussion today.

1. Demonstrate the research practice of sharing research information
2. Describe a model of research information sharing known as Open Access publishing and then identify its impact
3. As a response to the increasing volumes of shared information, identify ways to develop "synthesis" skills (i.e., joining pieces of information together to create new ideas or new meaning)

Relevance

The relevance of this lecture is twofold.

1. By recognizing a trend of sharing research information, you can look to free research works that may contribute to your academic, personal, and future professional life.
2. Additionally, we're going to discuss how to develop information synthesis skills that may help you manage the increasing volume of information.

Part 1. "Genome Race"

SLIDE 5

I want to tell you one of my favorite stories that demonstrates how important sharing information is to scientific research.

(This story is a re-telling of the H. Allen Orr's piece titled "The Genetic Adventurer," a review of the book *A Life Decoded: My Genome: My Life* by J. Craig Venter. Orr's review was published in *The New York Review of Books*, March 20, 2008.)

Often times, science is a quiet affair. A few laboratories work on a scientific problem, and when they get results, they publish in an academic journal that few in the public will ever see.

This was not the case in the research of sequencing the human genome.

Sequencing the human genome was a massive research project to determine the order of the chemical "letters" that makes up our DNA.

Background

SLIDE 6

Before I go any further with this story, let's examine the science behind human genome sequencing.

As you may remember from biology, our bodies are made up of cells. Within cells, we have chromosomes, and our chromosomes are made up of DNA.

DNA is the blueprint of life. It contains information that can control your physical characteristics - like your hair and eye color - and it can also predispose you to certain diseases - like diabetes or cancer.

DNA is made up of a chain of chemicals that are represented by the letters A, T, G, and C.

Genes are regions in your DNA and their specific sequence of letters tells your body how to make a certain protein.

Proteins are very important in your body. They build your tissues, they facilitate communication within your body, and they can lead to chemical reactions that make your body work.

Genes may sometimes lead to proteins that may result in illness as well.

All of your DNA is known as the human genome and it is very large. It's approximately three billion letters long with about 30,000 genes.

How do you determine the sequence of DNA?

There is automated DNA sequencing technology. You take DNA from an organism and inject it into a machine. Through chemistry and lasers, the machine can read the sequence.

Benefits of knowing the human genome sequence

SLIDE 7

Knowing the sequence of your DNA may contribute to medicine and human biology in a number of ways:

- detect genes that leads to diseases
- may lead to "personalized medicine" - for example, because of your genetic makeup, drug X may work best for you
- may lead to gene therapy - for example, altering your genes to treat disease
- facilitate comparative genomics - for example, to compare the human genome with the chimpanzee genome to understand how similar or different they are and to study evolution (by the way, the difference between the human genome and the chimpanzee genome is only a little more than 1 percent of the DNA sequence)
- identify "model organisms" - find organisms like the mouse or Drosophila flies that may have similar genes to humans, so we may experiment with these model genes and then try to translate the findings to humans

The race

SLIDE 8

Back to the story of the Genome Race.

It was a race between two programs to be the first to identify the sequence of the human genome.

The competitors represented the public and the private sectors:

- International Human Genome Sequencing Consortium (IHGSC) (which we'll interchangeably call the public effort, public program, or Human Genome Project). This was an international effort funded by international governments centered at the National Institutes of Health with support from Department of Energy in the United States. There was extensive funding from the independent NGO charity, the Wellcome Trust. The research was performed in universities and research centers from the United States, the United Kingdom, Japan, France, Germany, and more.
- Celera Genomics, a private company

About the public effort

SLIDE 9

Who started the race first?

It was the public program (International Human Genome Sequencing Consortium (IHGSC)).

The US federal government decided to coordinate the massive project to sequence the human genome among international partners.

This project began in 1990 by the US National Institutes of Health and the Department of Energy.

In the history of biomedical research, this was the largest cooperative undertaking ever. The cost would ultimately be \$3 billion.

The approach of the public effort was divide and conquer.

1. Researchers divided the human genome into large fragments. This was a painstaking process of carefully mapping large fragments so that every fragment was accounted for.
2. Afterwards, the DNA fragments were allotted to an international consortium of laboratories, and each would sequence the fragment.

Celera's effort

SLIDE 10

In 1999, a private company joined the race. The company was known as Celera Genomics and their goal of sequencing the human genome was led by the scientist, Dr. Craig Venter.

Celera took a different approach. It was a faster and more automated approach that required very powerful computers.

Their technique was known as whole genome shotgun sequencing.

1. Many copies of the human genome were broken into small pieces. This was done randomly.
2. These pieces were injected into an automated sequencer machine.
3. After reading the sequence of these little random pieces, they used computers to find overlap between the pieces. If they found an overlap, they would link and map the smaller pieces together.

The pace of sequencing

SLIDE 11

When Celera began their project in September 1999, the public effort had already completed sequencing 25% of the human genome.

But Celera moved at a fast pace, decoding fifty to one hundred million letters of DNA per day.

Here is a graph showing the pace of human genome sequencing overall. As you can see, after Celera joined the race in 1999, the pace picked up.

- <http://www.strategicgenomics.com/Genome/index.htm>

Sharing the human genome sequence data

SLIDE 12

What did they do with the sequence information?

During the race, Celera and the public effort took different approaches to storing their genome sequence information.

- Every day, the public program was loading their results into a database known as GenBank - developed by the U.S. National Center for Biotechnology Information with sister organizations in Europe and Japan . This database was publicly accessible so anyone could search and see the sequence data. Researchers around the world could use this information in their research.
- Celera took a private approach with its research data and did not publicly share their results.

- Additionally, Celera was seeking patent protection on some of the genes they were sequencing (see news story <http://news.bbc.co.uk/2/hi/science/nature/487773.stm>). A patent means that Celera would have exclusive rights to the genetic information and if scientists wanted to work with or use that information, they may need to request permission from Celera and possibly pay a fee to do so.

SLIDE 13

Here's an interesting note about the information sharing behind the human genome sequencing project. Craig Venter, the lead scientist in the Celera effort, wrote in his book "A Life Decoded" that Celera was using the public data to help with their sequencing project.

Because of this, some argue that the race to sequence the human genome is unclear since Celera used the public data. In other words, "how can you 'beat' the runner who hands you the baton?" (Orr, 2008)

So who won the genome sequencing race?

SLIDE 14

It was a tie.

The public effort and Celera were concerned that the other would be the first to finish sequencing the human genome.

Then, President Clinton intervened and after discussion, the parties collaborated.

On June 26, 2000, at the White House, President Clinton along with Dr. Craig Venter of Celera and Dr. Francis Collins of the public effort jointly announced the sequencing of the human genome was near completion with a working draft

SLIDE 15

In February 2001, Celera and the public effort scientists published details of their drafts - one day after the other, in separate publications.

In any case, both the private and public efforts were crucial and important to the success of the Human Genome Project.

On April 2003, it was essentially announced that complete genome has been identified (<http://news.bbc.co.uk/2/hi/science/nature/2940601.stm>).

What happened in the end?

SLIDE 16

In 2005, Celera Genomics announced that it will stop selling genomic information. All of its genomic data was made available in the public databases.

(<http://www.nature.com/nature/journal/v435/n7038/full/435006a.html>)

What does this genome sequencing race show us?

SLIDE 17

This case shows us that:

1. There are different approaches to scientific research - specifically, collaborate with different researchers to complete a massive project, or not.
2. There are different ways of dealing with research results. One of which is to publicly share your data and results.
3. Information technology has a powerful impact on research approach and the sharing of research results

SLIDE 18

Specifically, compare Celera's approach with the public effort.

- Celera conducted their work as an individual organization using complex and powerful computer systems to facilitate this massive project.
- The public effort divided the work of the sequencing project among a consortium of laboratories internationally. They cooperated and used computers to coordinate their efforts and merge their results.

SLIDE 19

And compare the management of research results by the public effort with Celera's.

- The public effort deposited their data everyday into a database that was publicly accessible. There was an online repository of genetic sequence data. Researchers around the world could use the genetic sequence information for their research on human health and biology.
- Celera was seeking to patent some of their genomic information, which may have limited public use.

Part 2. Open access publishing

SLIDE 20

Let's examine the sharing of research information further.

There is a model of research information sharing known as Open Access publishing. We're going to define it and look at its impact.

What is open access publishing?

Open access means the publishing of scholarly works that is:

- Digital and online
- Free of charge to the reader
- Free of most traditional copyright and licensing restrictions – so that readers and creators can re-distribute, remix, and reuse the content (so long as there is acknowledgement to the creator) (<http://www.earlham.edu/~peters/fos/overview.htm>)

SLIDE 21

However, It is important to note that open access publications are free for the reader, but the author, the author's organization, or the funding agency needs to pay for the article to be published.

In traditional publishing, researchers submit their publication, and when it is published, readers or libraries on behalf of readers pay a subscription fee to see the published work. Furthermore, in many cases, the publisher owns the copyright to the publication so that the author has limits on making copies or allowing others to adapt or re-use the work.

Sample open access publications

SLIDE 22

There are prominent open access scientific publications. One is the Public Library of Science, which includes:

- PLoS One
- PLoS Biology
- PLoS Medicine

You can also find open access journal articles in the life and health sciences via:

- PubMed Central
- BioMed Central

SLIDE 23

To find an open access publication in your field, search the Directory of Open Access Journals

- DOAJ - <http://www.doaj.org/>

There are also open access databases of data.

- GenBank for genetic information, <http://www.ncbi.nlm.nih.gov/sites/gquery>
- PubChem for small molecule information, <http://pubchem.ncbi.nlm.nih.gov/>

Value of open access publishing

SLIDE 24

Because OA works are free to the reader and disseminated online, there is a greater audience that can access and make use of research work. Since a lot of research is funded by public tax dollars from the government, some argue that there is an obligation to publish the results so that the general public may view it - a return on public investment.

Since traditional copyright is not in place, authors retain ownership and control of their work and can authorize others to use the work freely. (Traditionally, when you publish a work, the publication owns the copyright on the work, and the author may lose the ability to duplicate and share their work on their own terms.)

Because open access publications are free to read wherever there is Internet access, it may advance science and technology because knowledge and ideas are shared (which may lead to new ideas).

There are also obligations to make researchers share their work publicly available. Here are two prominent examples:

- NIH Public Access Policy - <http://publicaccess.nih.gov/> If you receive funding from the NIH, you need to deposit a copy of your article within 12 months of publication to PubMed Central repository.
- Open access policy of the Wellcome Trust - <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Policy/index.htm>

Part 3. Synthesis skills

SLIDE 25

With open access leading to the sharing of research publications at no cost to the reader, the volume of information available will grow and grow. How do we cope with this?

Definition

Howard Gardner is a developmental psychologist at Harvard. In his book *Five Minds for the Future*, he describes the synthesizing mind. In my opinion, the synthesizing mind is one response to the increasing volume of information.

- Gardner, Howard. *Five minds for the future*. Boston: Harvard Business School Press, 2006.

To paraphrase: Synthesizing is taking information from disparate sources, understanding and evaluating that information objectively, and putting it together in ways that make sense to you and to others.

Motivation

Gardner argues that people crave coherence and integration, they want to make sense of information and put it to use (p. 4).

Kinds of synthesis

SLIDES 26-31 - Paraphrased below is Gardner's identification of the kinds of synthesis.

Kind of synthesis	Definition	Example
Narratives	Weaving information into a coherent narrative. Telling a story.	Tolstoy's War and Peace - A social commentary, a work of fiction, and a historical chronicle about the Napoleonic era in Tsarist Russia all rolled into one.
Taxonomies	Categorizing items in terms of salient characteristics.	Linnaean classification of plants and animals and the periodic table of elements.
Complex concepts	Developing a new concept to blend or unify a range of phenomena.	Sigmund Freud developed the concept of the unconscious.
Rules and aphorisms	Capturing and conveying many ideas/wisdom through short phrases that are memorable and widely applicable.	Think first and act second
Powerful metaphors, images, and themes	Invoking metaphors to bring concepts to life.	For example: <ul style="list-style-type: none"> • Corporations create brands with words, graphics, and jingles. • Mashups - combining digital works to create a new work. For instance, PadMapper, which brings together Google Maps with apartment rental listings. http://www.padmapper.com/ • Information graphics are visual representations of information, data, or knowledge. http://www.infographicsshowcase.com/
Embodiments without words	Works of art that captures a lot of ideas	Consider Picasso's Guernica which tries to express the horror of the Spanish Civil War in a cubist-style mural
Theories	Joining concepts into a theory (i.e., a plausible explanation of phenomena)	Adam Smith's theory of a market economy weaves together ideas of supply and demand, labor, production, profit, and loss
Metatheory	Creating an overall framework for knowledge	George Wilhelm Friedrich Hegel developed the Hegelian dialectic "There are three stages in development: a thesis, giving rise to its reaction, an antithesis, which contradicts or negates the thesis, and the tension between the two being resolved by means of a synthesis" (i.e., "reconciliation and forming a new proposition"). (Wikipedia, 2010)

Cultivating synthesis skills

SLIDE 32

To develop these synthesis skills, Gardner (p. 155) recommends

- digesting new information (i.e., recognizing, learning, and understanding)
- organizing it
- applying the new knowledge in the ways outlined above.

SLIDE 33

Here is a college student example:

- In assignments, collect different types of information, and then organize them in a helpful manner. Experiment with visuals, layout, and organization.

And a professional example:

- At work, recognize new knowledge or skills that may be important to your field. Learn it and apply it.

Summary

SLIDE 34

To summarize:

- Sharing information and collaboration are important research practices to consider
- Open access publishing is a model for sharing research information
- With greater sharing, there will be a larger volume of information and knowledge available
- One response to this increasing volume of information is to develop synthesis skills for digesting, organizing, and applying new knowledge

References

SLIDE 35

Gardner, Howard. *Five Minds for the Future*. Boston: Harvard Business School Press, 2006.

Orr, H. Allen. "The Genetic Adventurer." A review of the book *A Life Decoded: My Genome: My Life* by J. Craig Venter. *The New York Review of Books*, March 20, 2008.