Big Education in the Era of Big Data

Irwin King

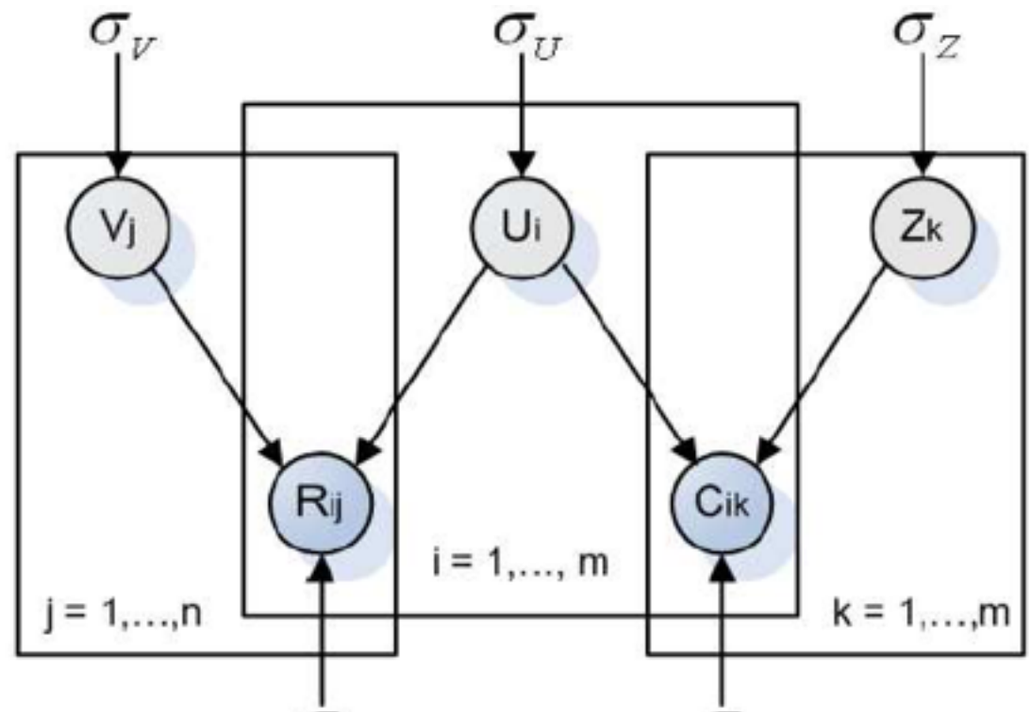
Associate Dean (Education), Faculty of Engineering, CUHK
Director, Shenzhen Rich Media and Big Data Analytics and Application Key Lab, SZRI, CUHK
PI, The Knowledge and Education Exchange Platform (KEEP), CUHK

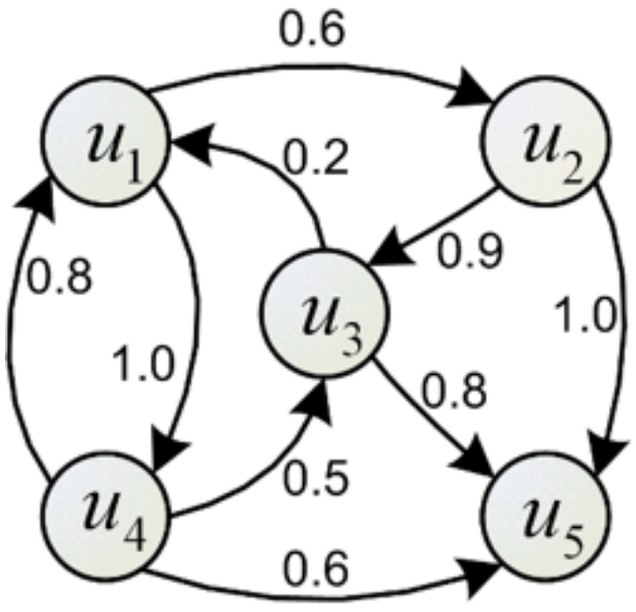
Department of Computer Science and Engineering
The Chinese University of Hong Kong

king@cse.cuhk.edu.hk
http://www.cse.cuhk.edu.hk/~king

Social Recommendations

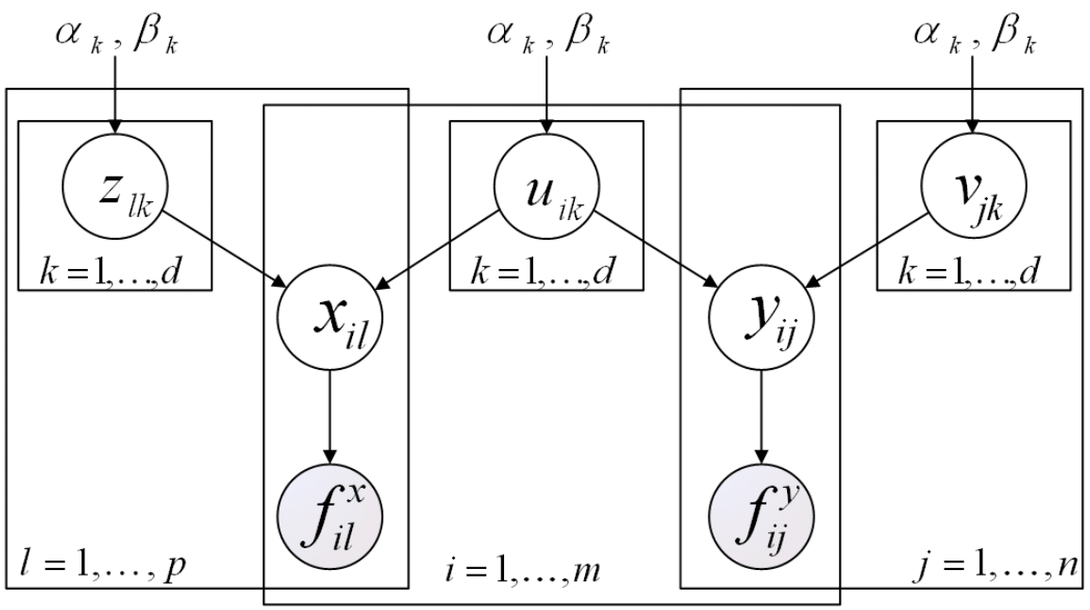
	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3





$$\mathcal{L}(R, C, U, V, Z) =$$
$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}(r_{ij} - g(U_i^T V_j))^2 + \frac{\lambda_C}{2}\sum_{i=1}^{m}\sum_{k=1}^{m} I_{ik}^{C}(c_{ik}^{*} - g(U_i^T Z_k))^2$$
$$+\frac{\lambda_U}{2}\|U\|_F^2 + \frac{\lambda_V}{2}\|V\|_F^2 + \frac{\lambda_Z}{2}\|Z\|_F^2,$$

Collective Probabilistic Factor Model



$$\mathcal{L}(U,V,Z;F^x,F^y)$$
$$= \sum_{i=1}^{m}\sum_{l=1}^{p}(f^x_{il}\ln x_{il} - x_{il}) + \sum_{i=1}^{m}\sum_{j=1}^{n}(f^y_{ij}\ln y_{ij} - y_{ij})$$
$$+ \sum_{i=1}^{m}\sum_{k=1}^{d}((\alpha_k - 1)\ln(u_{ik}/\beta_k) - u_{ik}/\beta_k)$$
$$+ \sum_{j=1}^{n}\sum_{k=1}^{d}((\alpha_k - 1)\ln(v_{jk}/\beta_k) - v_{jk}/\beta_k)$$
$$+ \sum_{l=1}^{p}\sum_{k=1}^{d}((\alpha_k - 1)\ln(z_{lk}/\beta_k) - z_{lk}/\beta_k) + \text{const.}$$

$$u_{ik} \leftarrow u_{ik}\frac{\sum_{j=1}^{n}(f^y_{ij}v_{jk}/y_{ij}) + \sum_{l=1}^{p}(f^x_{il}z_{lk}/x_{il}) + (\alpha_k - 1)/u_{ik}}{\sum_{j=1}^{n}v_{jk} + \sum_{l=1}^{p}z_{lk} + 1/\beta_k}$$

$$v_{jk} \leftarrow v_{jk}\frac{\sum_{i=1}^{m}(f^y_{ij}u_{ik}/y_{ij}) + (\alpha_k - 1)/v_{jk}}{\sum_{i=1}^{m}u_{ik} + 1/\beta_k},$$

$$z_{lk} \leftarrow z_{lk}\frac{\sum_{i=1}^{m}(f^x_{il}u_{ik}/x_{il}) + (\alpha_k - 1)/z_{lk}}{\sum_{i=1}^{m}u_{ik} + 1/\beta_k}.$$

$$u_{ik} \leftarrow u_{ik}\frac{\theta\sum_{j=1}^{n}(f^y_{ij}v_{jk}/y_{ij}) + (1-\theta)\sum_{l=1}^{p}(f^x_{il}z_{lk}/x_{il}) + (\alpha_k - 1)/u_{ik}}{\theta\sum_{j=1}^{n}v_{jk} + (1-\theta)\sum_{l=1}^{p}z_{lk} + 1/\beta_k}$$

The grass is greener on the other side...

Be inspired!

Stories and more stories...

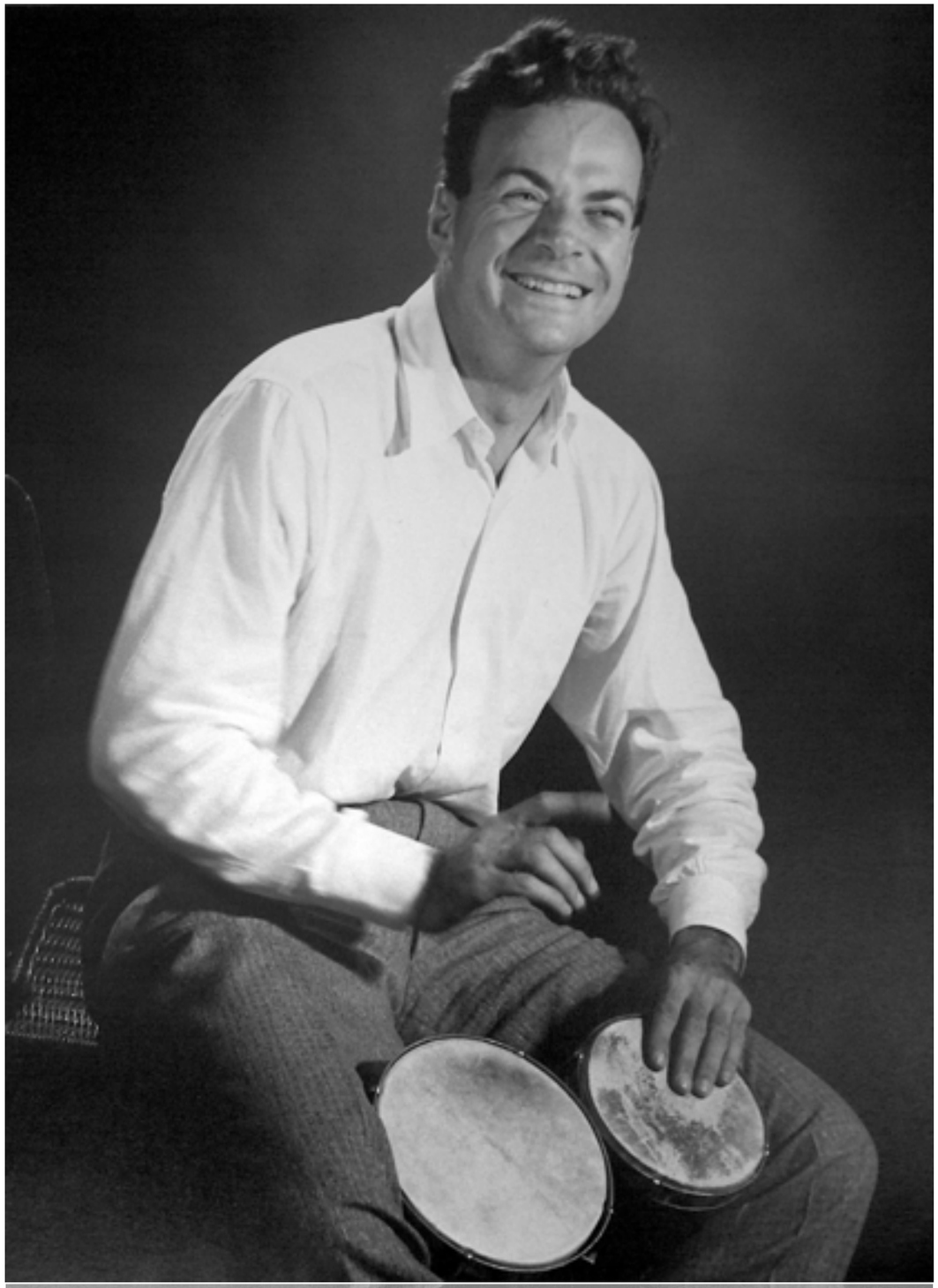
Be informed!

The devil is in the details...

Be challenged!

Our Education System

Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.



Richard Feynman

Words of Wisdom

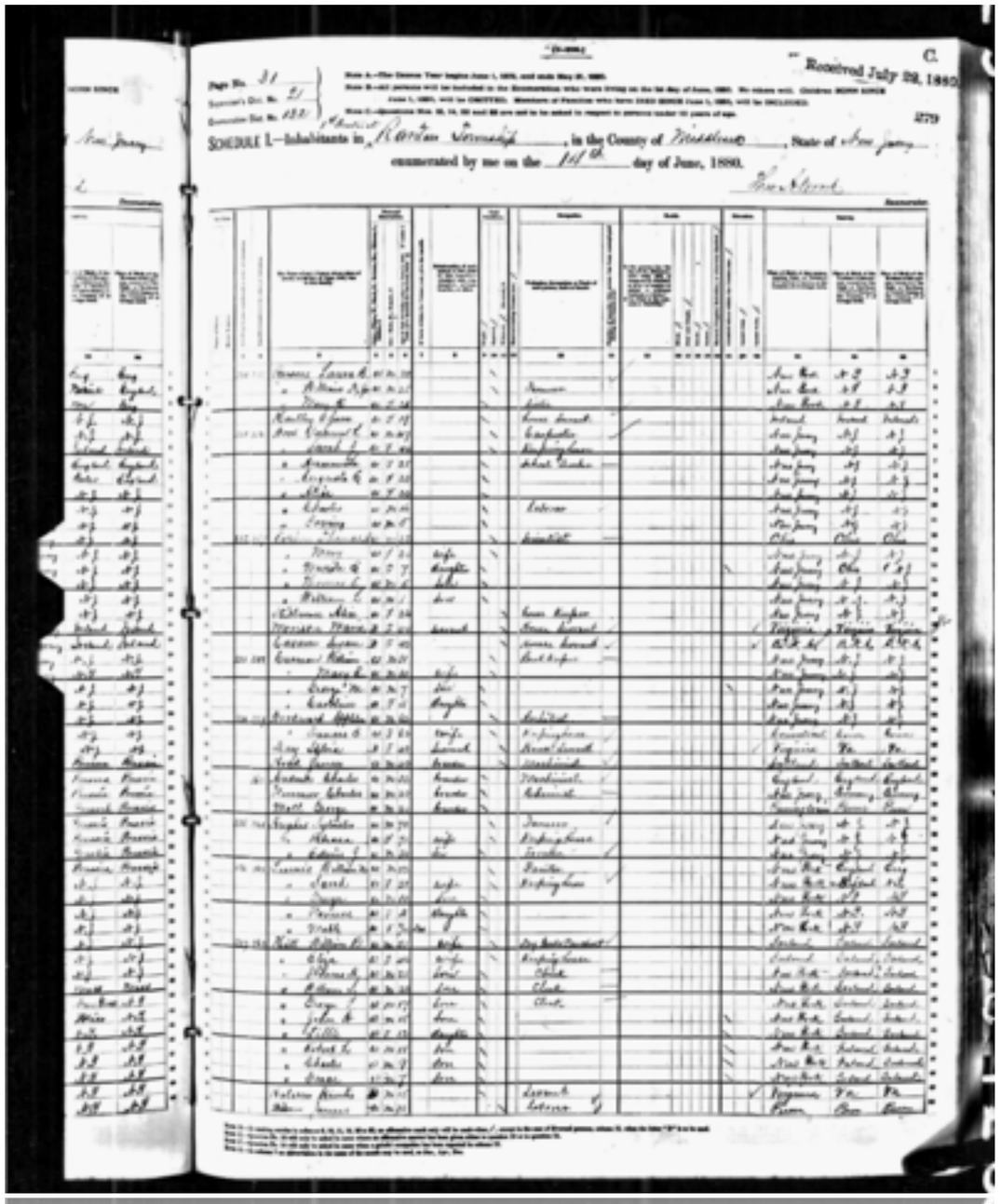
The **BEST** universities focus on **EDUCATION**!

The **BETTER** universities focus on citation numbers and impact factors…

The **GOOD** universities focus on counting the number of publications…

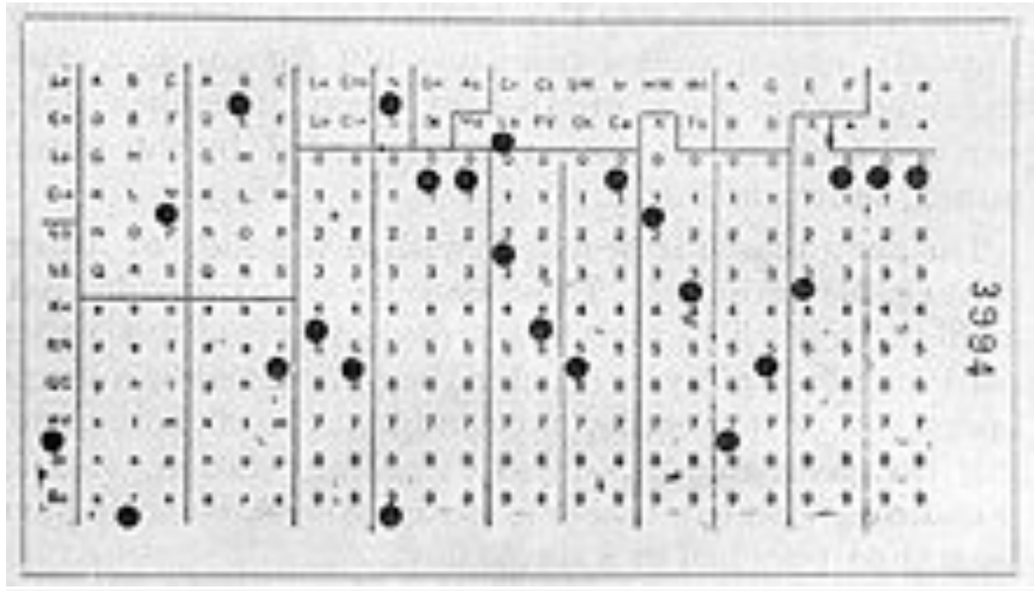
The First Big Data Challenge

- 1880 census
- 50 million people
- Age, gender (sex), occupation, education level, no. of insane people in household



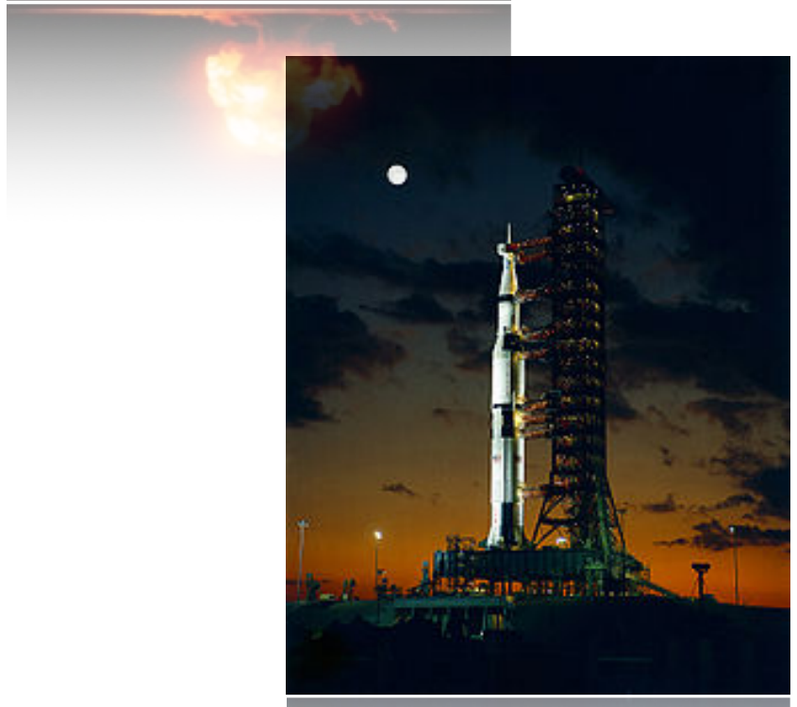
The First Big Data Solution

- Hollerith Tabulating System
- Punched cards – 80 variables
- Used for 1890 census
- 6 weeks instead of 7+ years



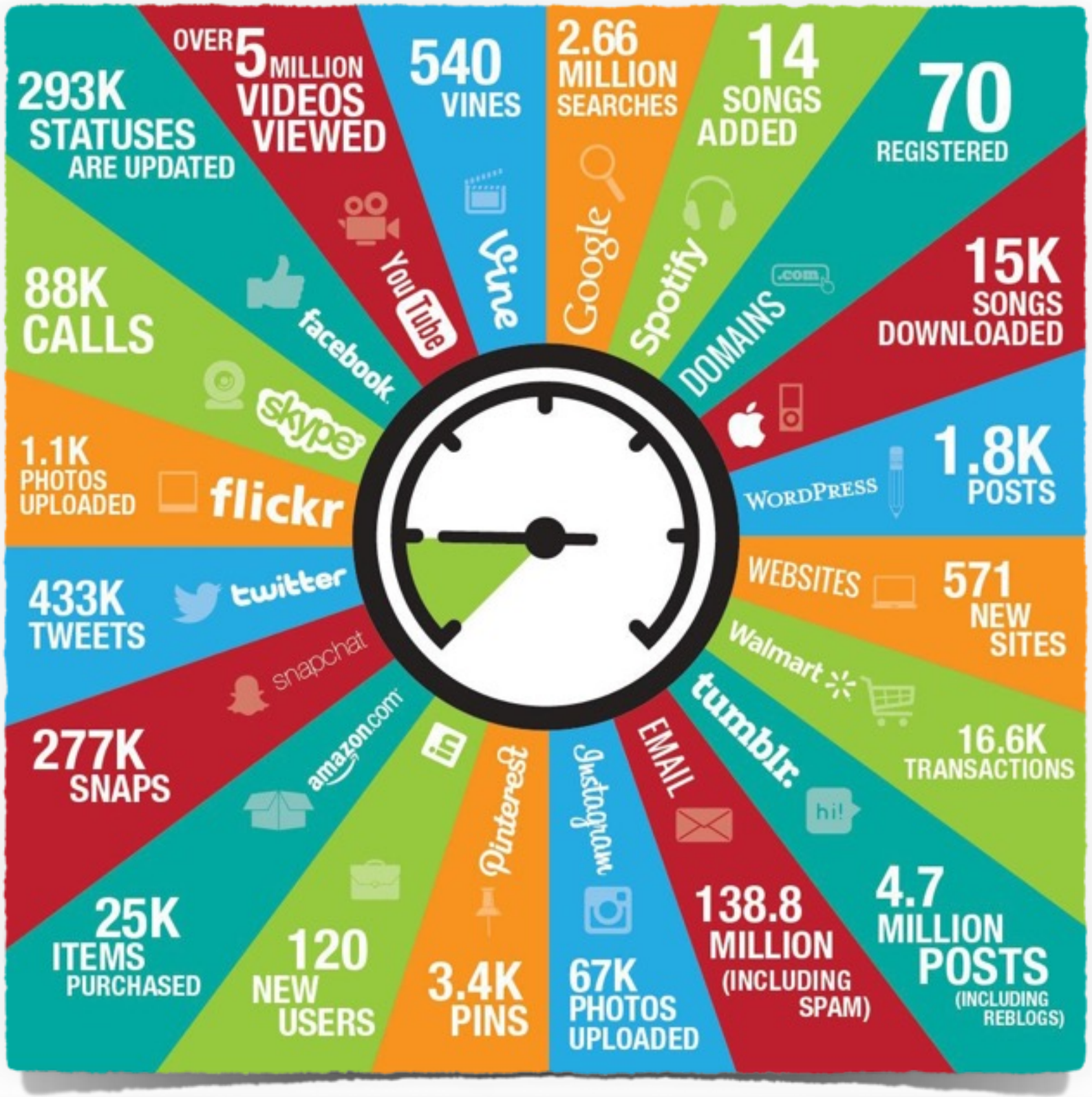
Big Projects

- Manhattan Project (1946 - 1949)
  - $2 billion (approx. 26 billion in 2013)
  - Catalyst for "Big Science"
- Space Program (1960s)
  - Began in late 1950s
  - An active area of big data nowadays



Big Science

- The International Geophysical Year
  - An international scientific project
  - Last from Jul. 1, 1957 to Dec. 31, 1958
- A synoptic collection of observational data on a global scale
- Implications
  - Big budgets, Big staffs, Big machines, Big laboratories


293K
STATUSES
ARE UPDATED
OVER 5 MILLION
VIDEOS
VIEWED
540
VINES
2.66
MILLION
SEARCHES
14
SONGS
ADDED
70
REGISTERED
88K
CALLS
facebook
YouTube
vine
Google
Spotify
DOMAINS
.com
15K
SONGS
DOWNLOADED
skype
1.1K
PHOTOS
UPLOADED
flickr
WordPress
1.8K
POSTS
433K
TWEETS
twitter
WEBSITES
571
NEW
SITES
snapchat
Walmart
277K
SNAPS
amazon.com
16.6K
TRANSACTIONS
tumblr.
hi!
25K
ITEMS
PURCHASED
120
NEW
USERS
Pinterest
3.4K
PINS
Instagram
67K
PHOTOS
UPLOADED
EMAIL
138.8
MILLION
(INCLUDING
SPAM)
4.7
MILLION
POSTS
(INCLUDING
REBLOGS)


Velocity
VALUE!
Veracity

Seven Typical Tasks

- **Querying**: spherical range-search $O(N)$, orthogonal range-search $O(N)$, nearest-neighbor $O(N)$, all-nearest-neighbors $O(N^2)$
- **Density estimation**: mixture of Gaussians, kernel density estimation $O(N^2)$, kernel conditional density estimation $O(N^3)$
- **Classification**: decision tree, nearest-neighbor classifier $O(N^2)$, kernel discriminant analysis $O(N^2)$, support vector machine $O(N^3)$, Lp SVM
- **Regression**: linear regression, kernel regression $O(N^2)$, Gaussian process regression $O(N^3)$, LASSO

Seven Typical Tasks

- **Dimension reduction**: PCA, non-negative matrix factorization, kernel PCA $O(N^3)$, maximum variance unfolding $O(N^3)$, Gaussian graphical models, discrete graphical models
- **Clustering**: k-means, mean-shift $O(N^2)$, hierarchical (FoF) clustering $O(N^3)$
- **Testing and matching**: MST $O(N^3)$, bipartite cross-matching $O(N^3)$, n-point correlation 2-sample testing $O(N^n)$, kernel embedding

Seven "Giants" of Data

- **Basic statistics**: means, covariances, etc.
- **Generalized N-body problems**: distances, geometry, etc.
- **Graph-theoretic problems**: discrete graphs
- **Linear-algebraic problems**: matrix operations
- **Optimizations**: unconstrained, convex, etc.
- **Integrations**: general dimension
- **Alignment problems**: dynamic programming, matching, etc.

Alexander Gray, GIT

Seven General Strategies

- **Divide and conquer/indexing**: trees
- **Function transforms**: series
- **Sampling**: Monte Carlo, active learning, etc.
- **Locality**: caching, hashing, etc.
- **Streaming**: online
- **Parallelism**: clusters, GPUs, etc.
- **Problem transformation**: reformulations

What Is Big Education?

Mr. Big Data + Ms. Education

Big Education Components

- **Pedagogy**
  - Theory and practice on education
- **Platforms** and **system technology**
  - Content creation, storage, delivery, etc.
  - Learning management systems
- **Algorithms** for Big Education
  - Learning analytics at scale

Big Education Applications

- **Statistics** for citation index, impact factors, h-index, bibliometrics, university ranking, etc.
- **Social computing** in students's learning networks, co-author networks, career networks, etc.
- Behavior and **learning analytics**
- Career and learning path **recommendations**
- Finding **domain experts**, **trendy research topics**, **communities**, etc.
- Identifying **plagiarism**, **at-risk students**, **frauds**, etc.
- Intelligent and automated **assessment and tutoring**


TURNIT IN
CLASS BLOGS
e-booKs
VIRTUAL PROFESSOR
SOCIAL NETWORK
eMAILS
BLACKBOARD DISCUSSIONS
TEAM'S INSTANT MESSAGES
CHAT
PODCASTS
WIKI
WI-FI
GPS
iPhone
CD
HD BACK
UPS
E-LEARNING OVERLOAD
© WWW.JOSETOONS.COM
NOV 19, 08

Education without values, as useful as it is, seems rather to make man a more clever devil.

C.S. Lewis

Trends in Big Education

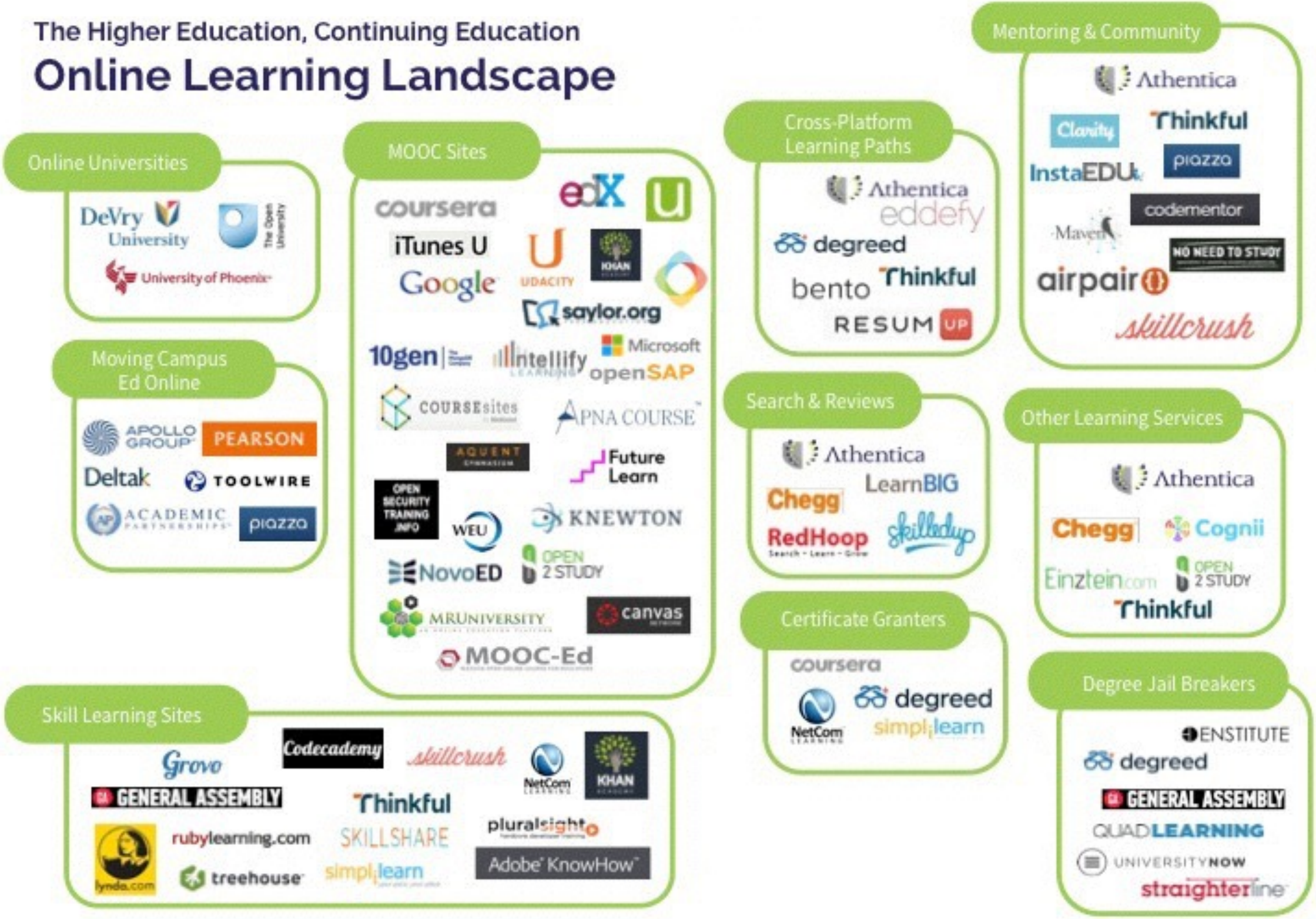
Cost
Effectiveness

Collaboration

Customization

The Higher Education, Continuing Education

Online Learning Landscape



http://athentica.com/wp-content/uploads/2013/10/Online-Learning-Landscape-Oct-2013.jpg

BIG MONEY

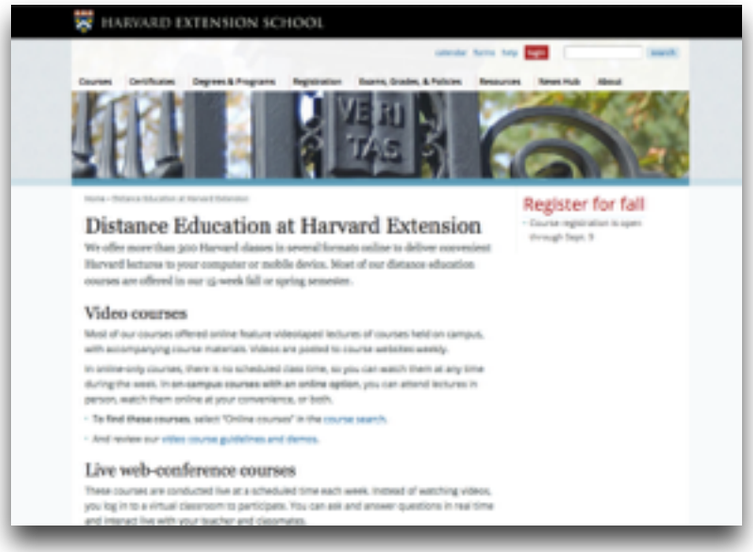
BIG USERS

BIG IMPROVEMENT

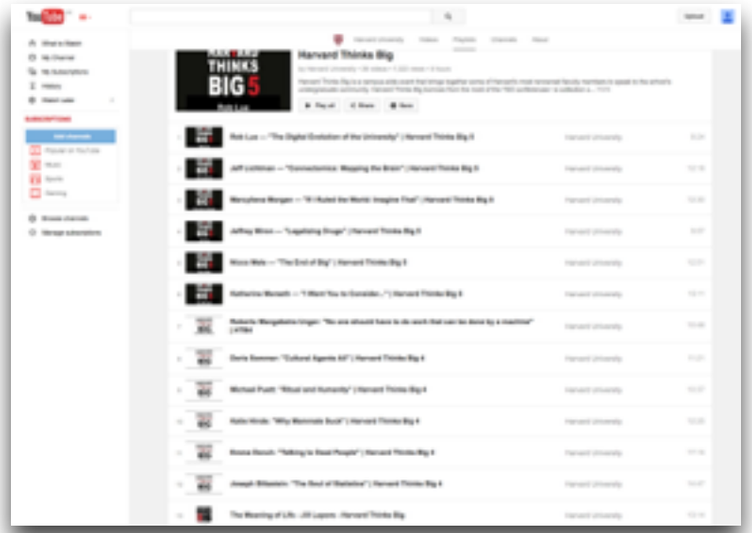
Big Education on Lifelong Learning

Multimodal Learning

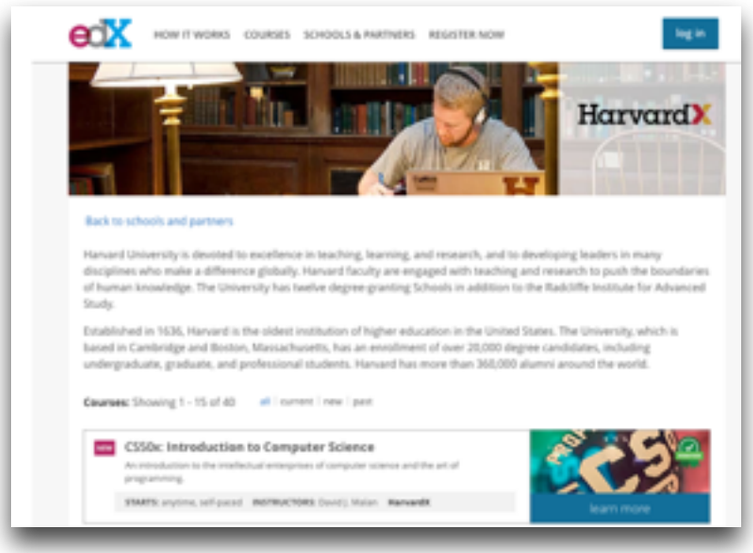
Harvard @ iTunes U



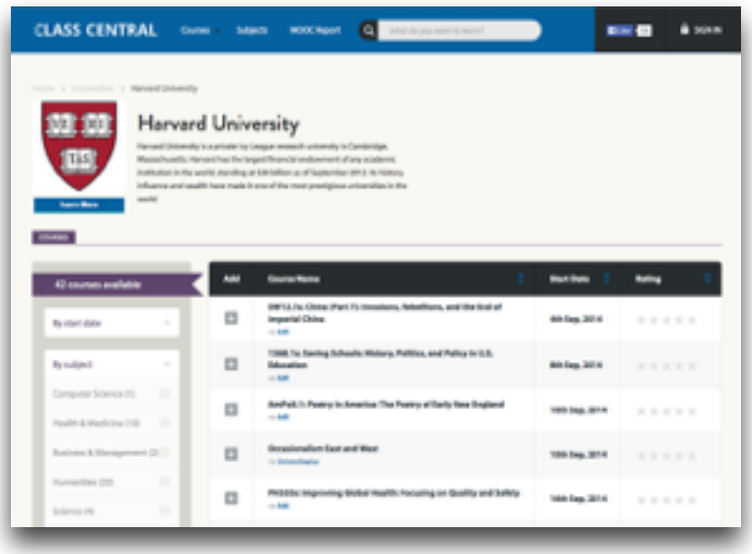
harvard.edu



Harvard @ YouTube



Harvard @ edX



Harvard @ Class Central

MOOC

Massive Open Online Course

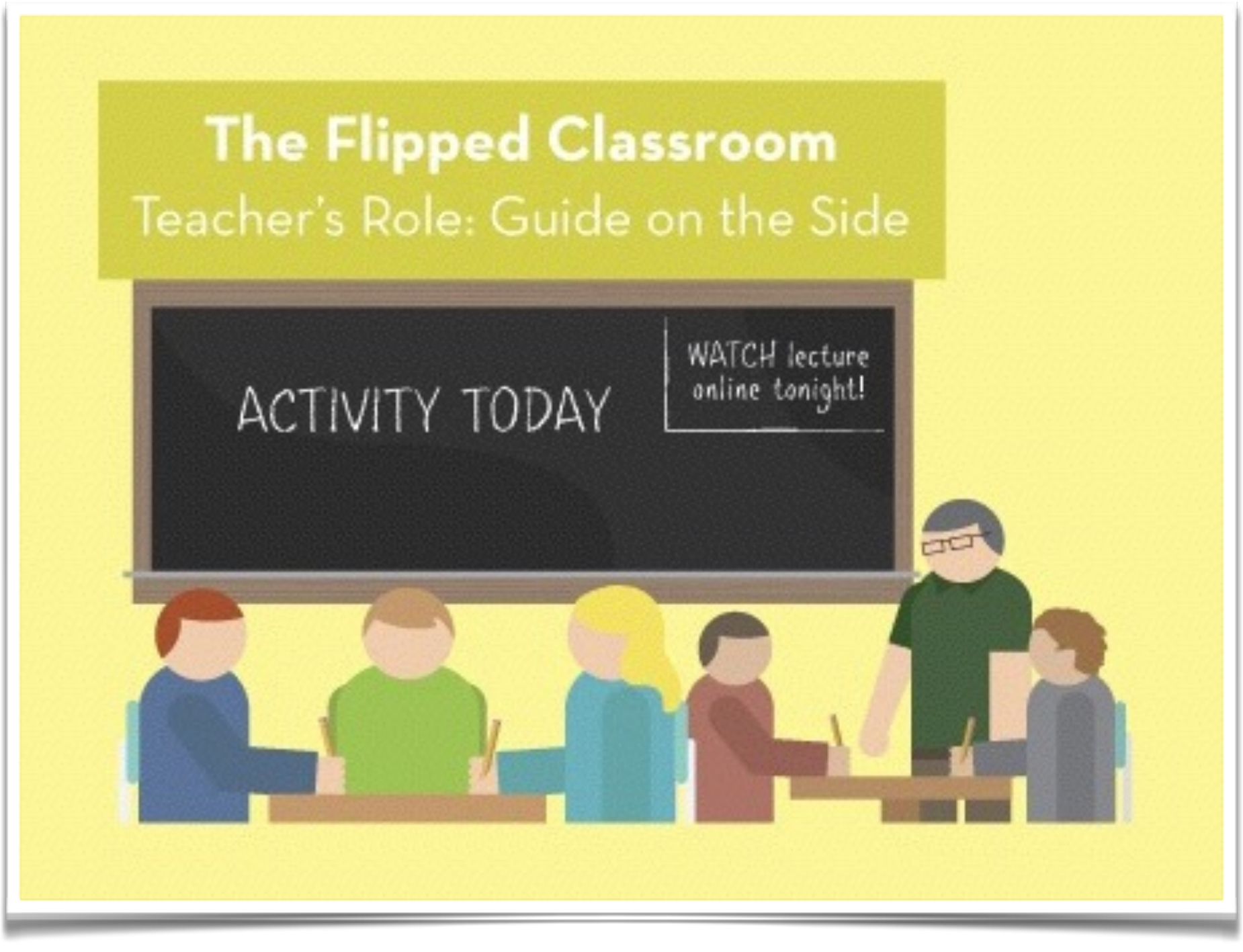
Small Private Online Course (SPOC) with Degree

UDACITY

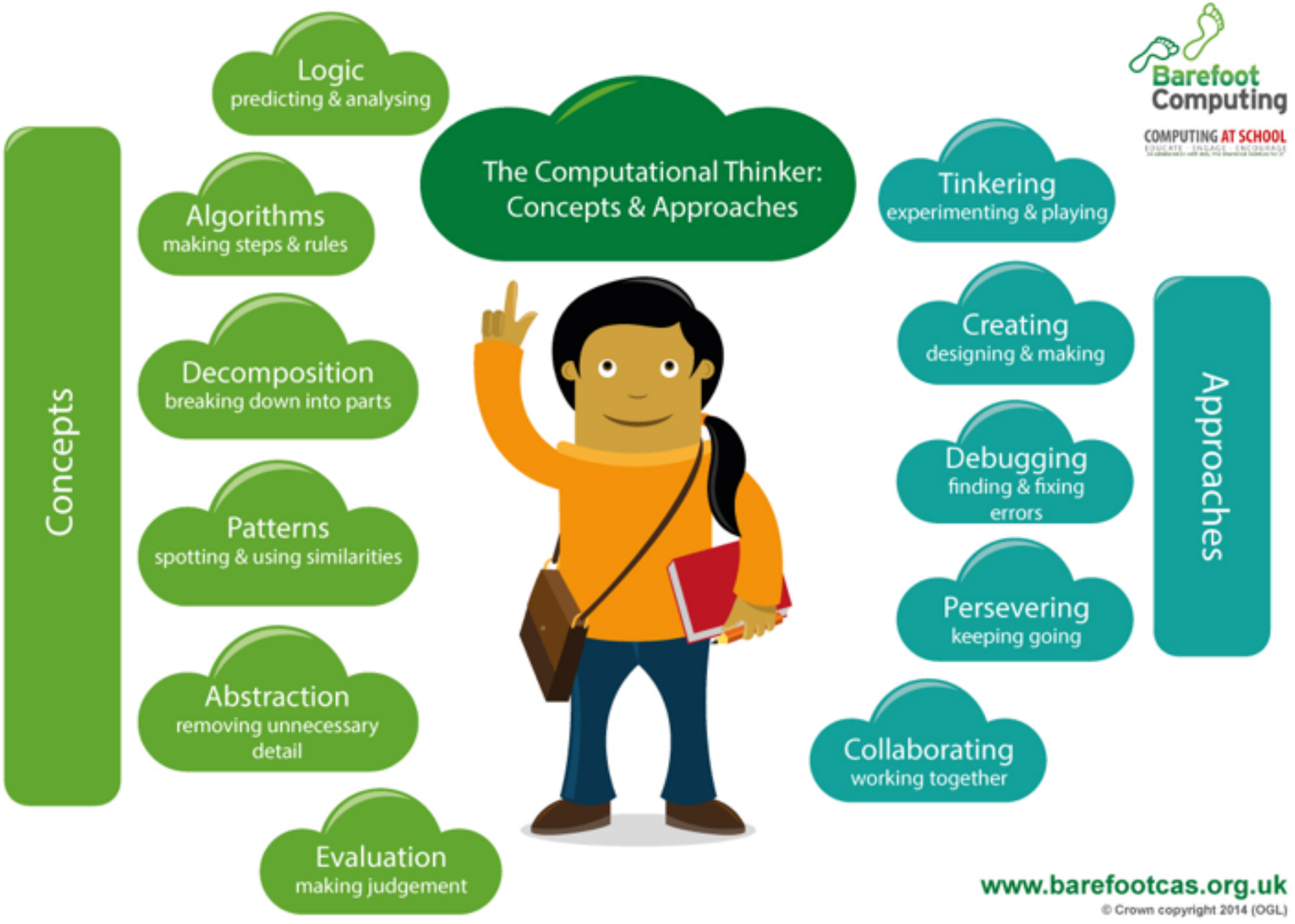
Georgia Tech College of Computing

at&t

Flipped Classroom

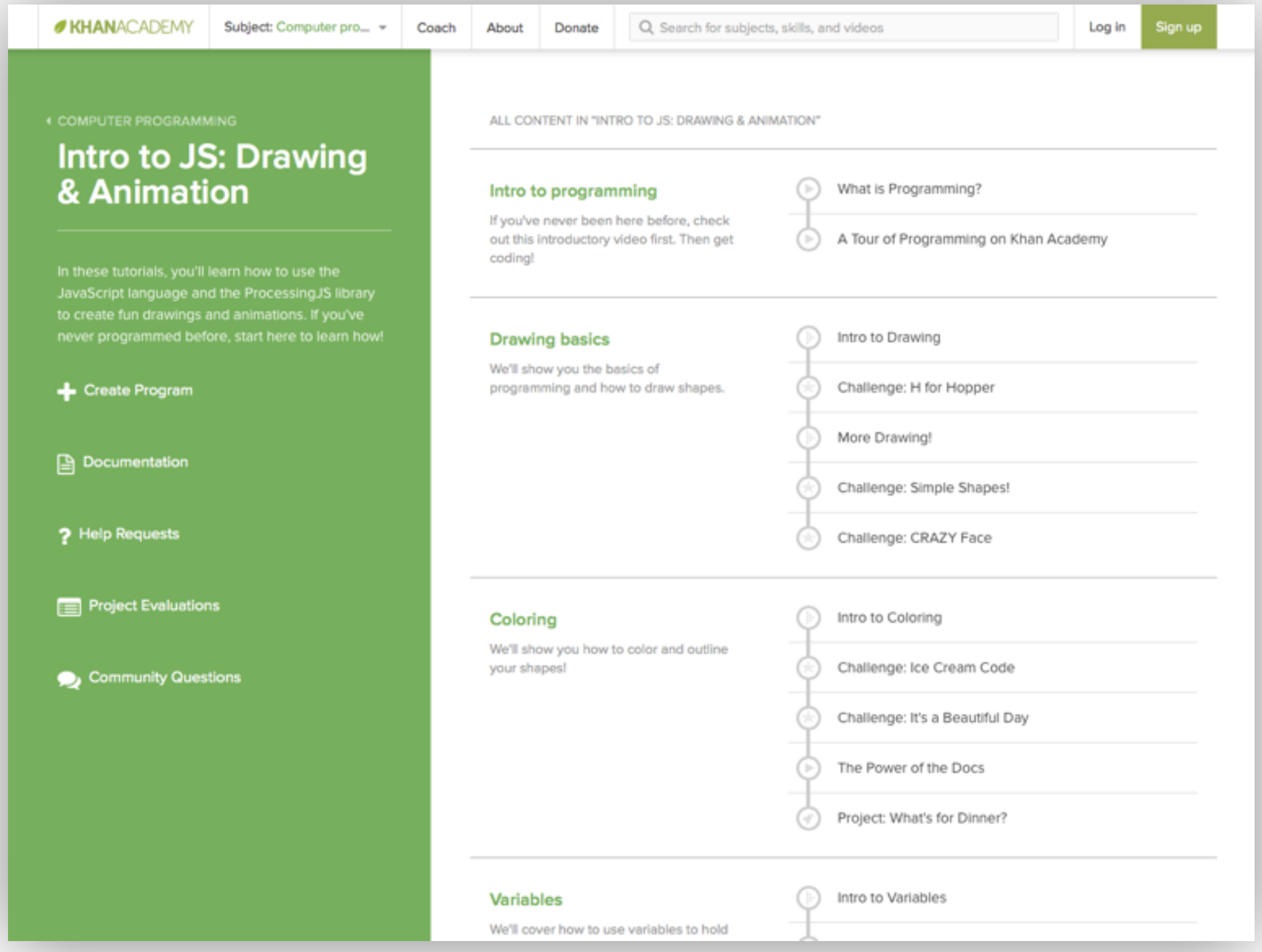


Computational Thinking

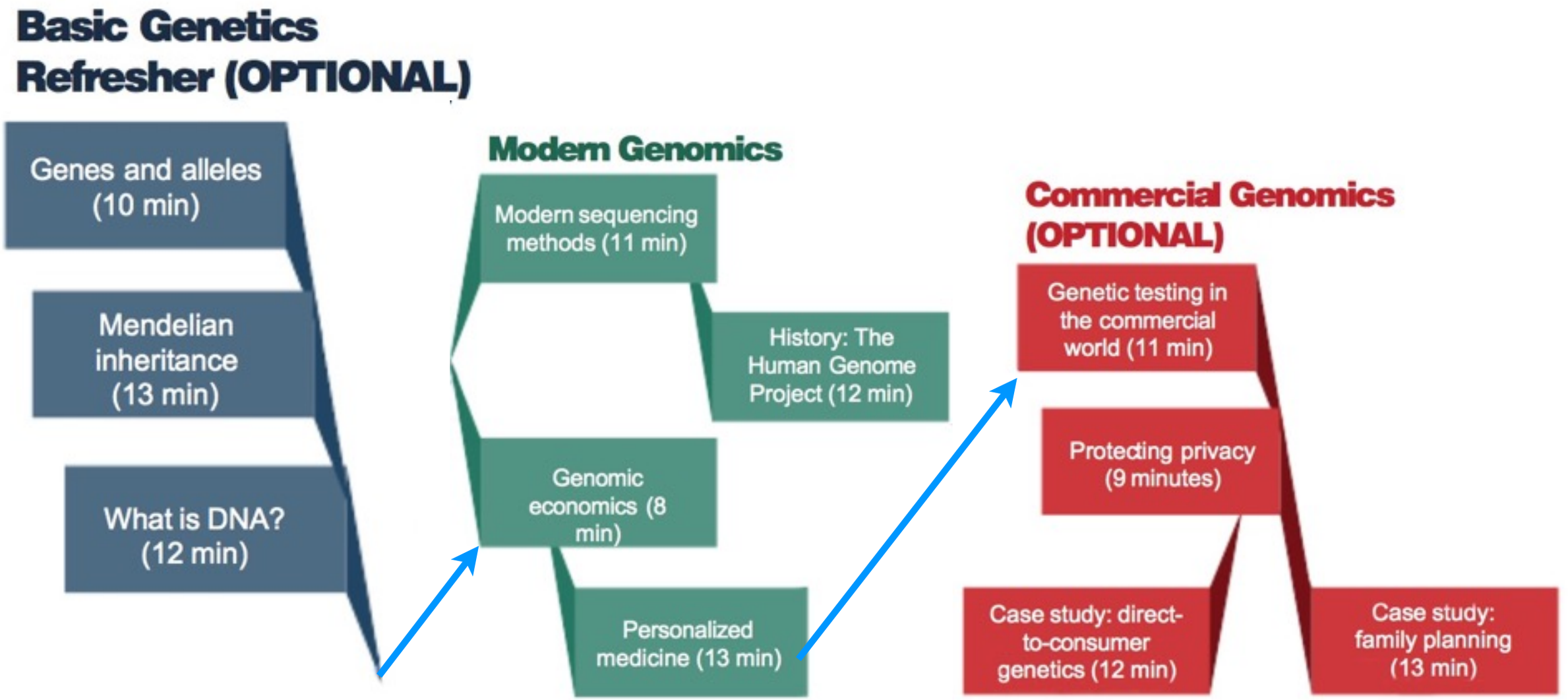


Barefoot would like to acknowledge the work of Julia Briggs and the eLIM team at Somerset County Council for their contribution to this poster.

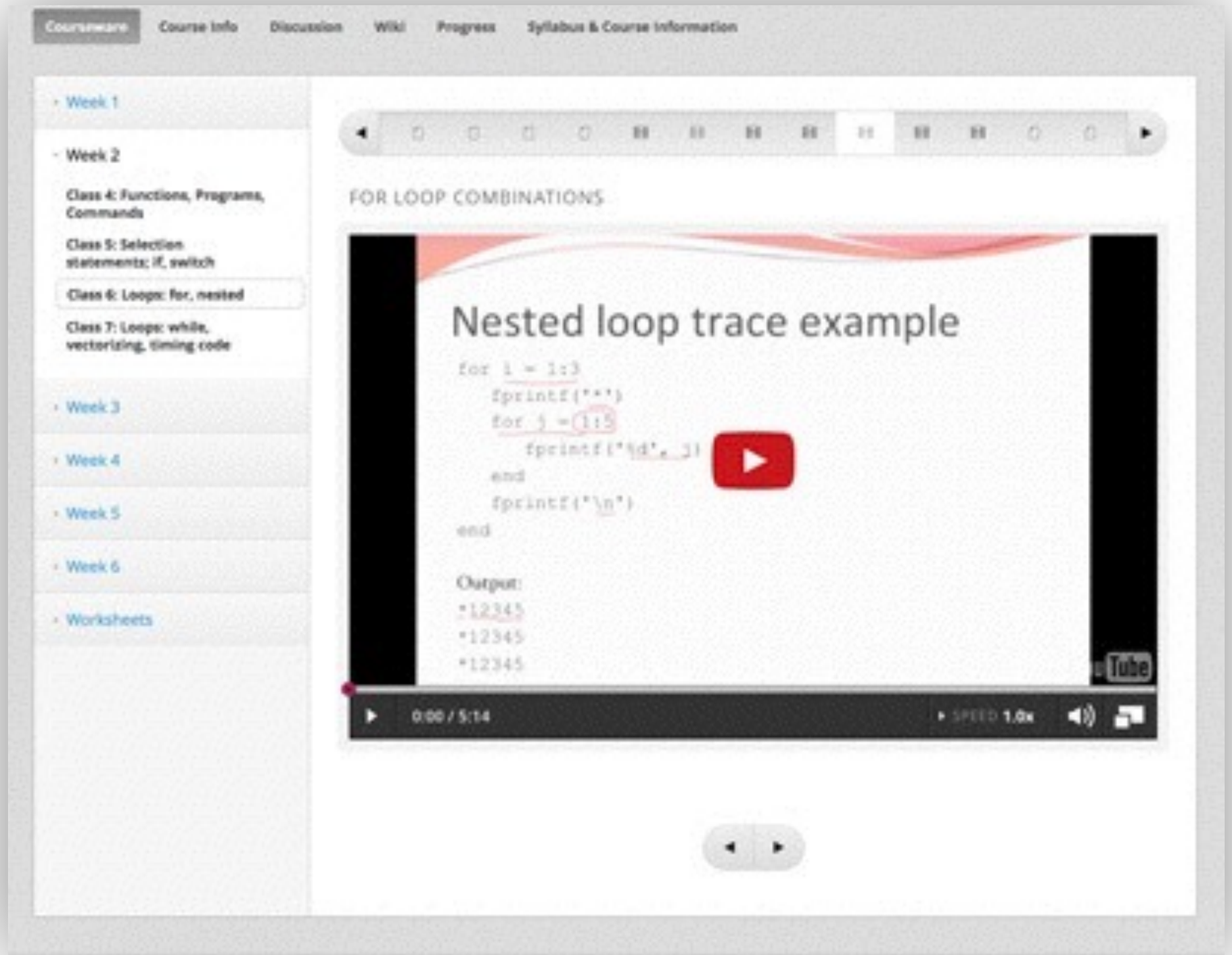
Microlearning



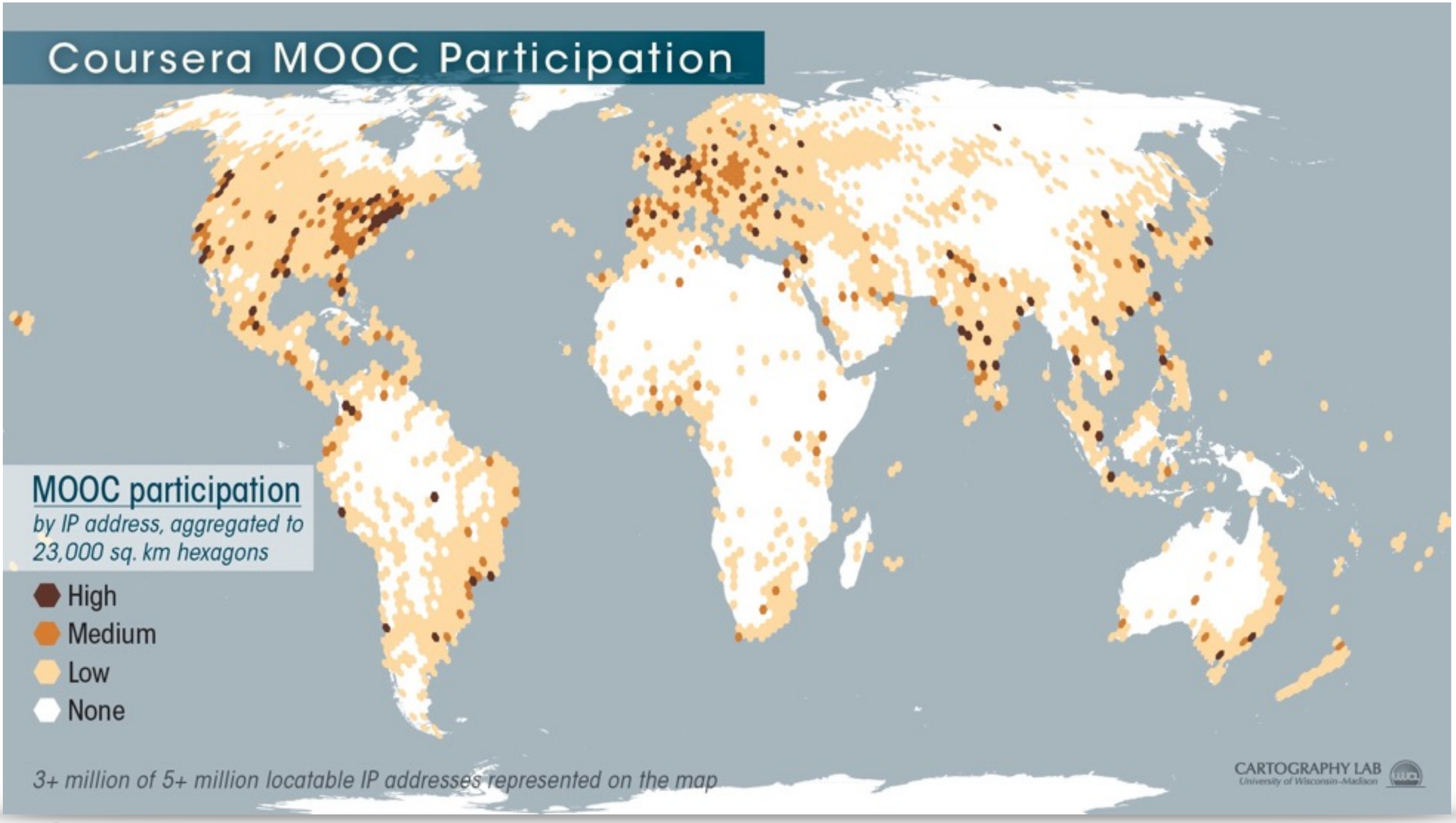
Personalized Learning



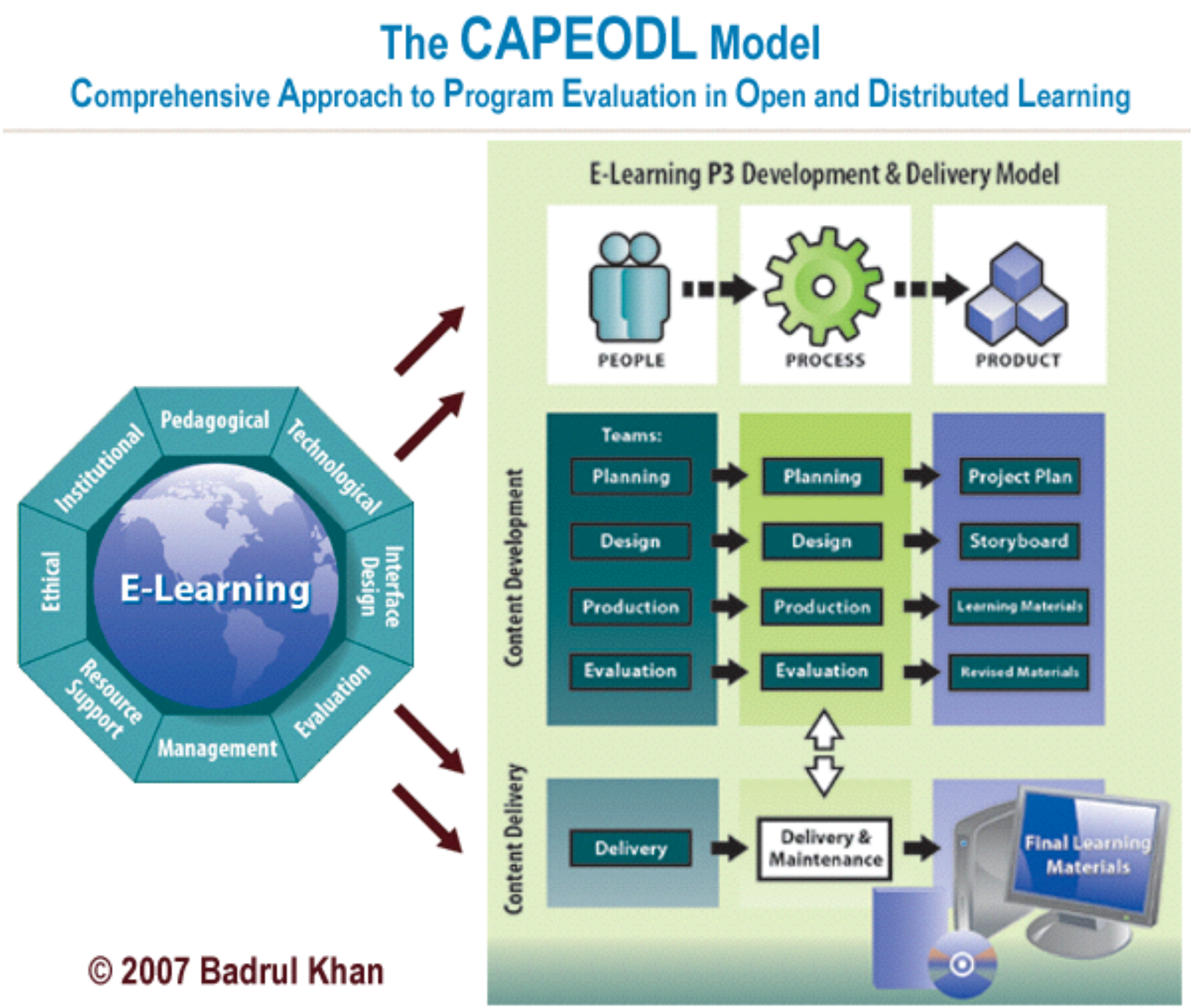
Active Learning



Peer Learning



Assessment Methods

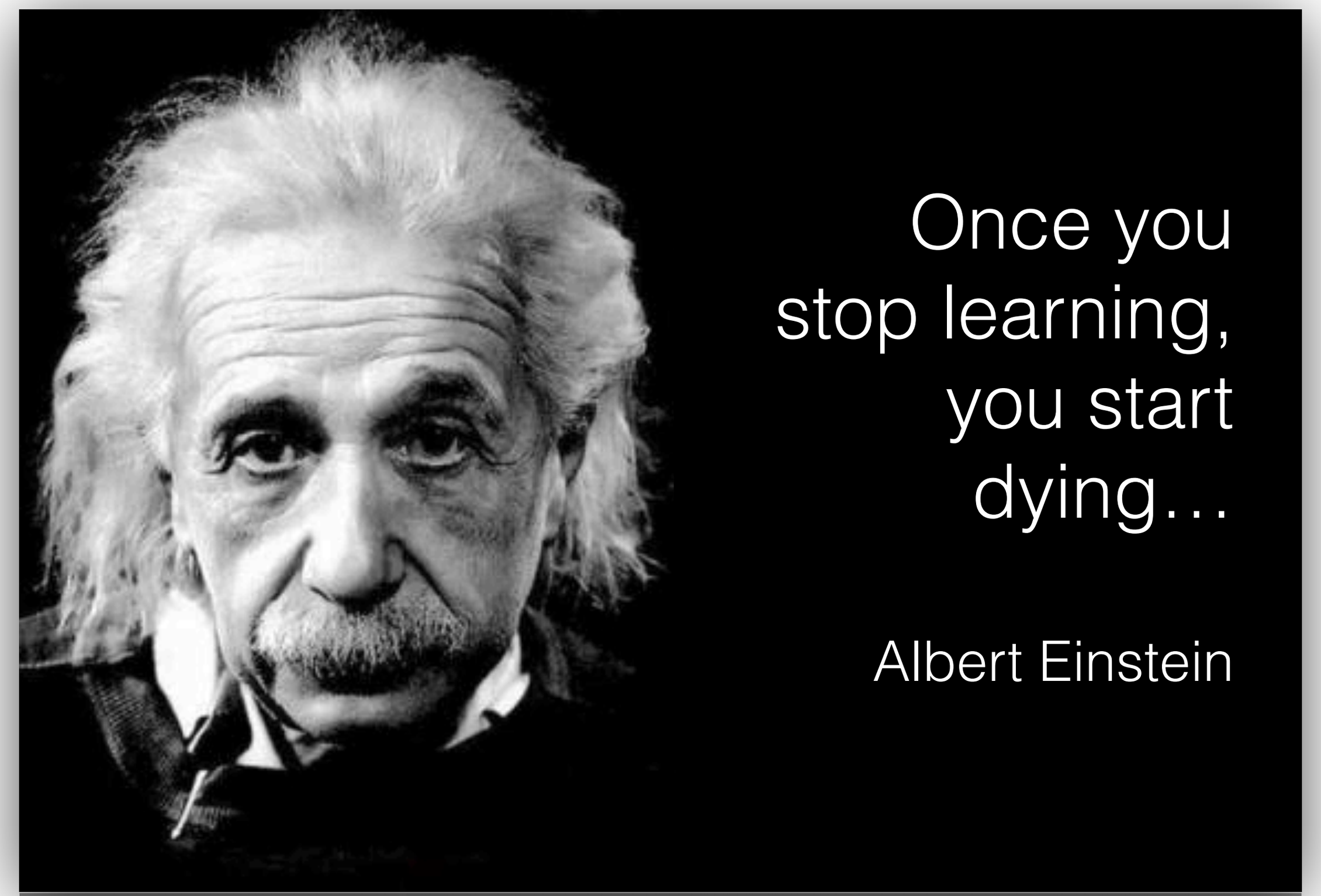


Gamification

Learning Analytics

Concluding Remarks

- *Be Inspired*
  - Big Education is the focus!
- *Be Informed*
  - Big Data in Education is the *VALUE* proposition!
- *Be Challenged*
  - Use technologies to transform education in the Big Data Era!


Once you
stop learning,
you start
dying…
Albert Einstein

KEEP

Knowledge and Education Exchange Platform

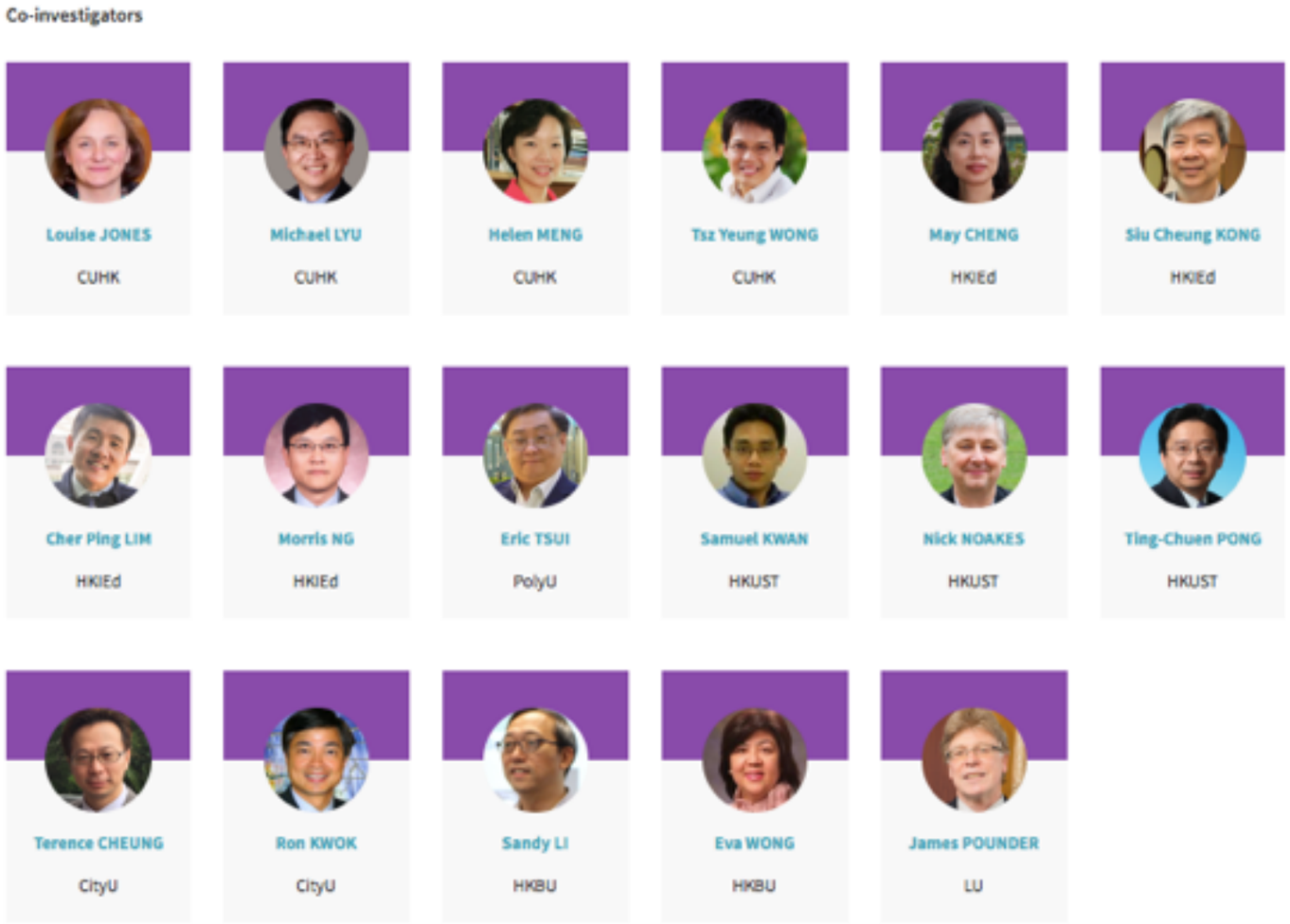
Acknowledgments

- Ken Chan (Ph.D.)
- Xixian Chen (Ph.D.)
- Chen Cheng (Ph.D.)
- Yuxin Su (Ph.D.)
- Haiqin Yang (Postdoc)
- Xiaotian Yu (Ph.D.)
- Connie Yuen (Ph.D.)
- Jichuan Zeng (Ph.D.)
- Hongyi Zhang (Ph.D.)
- Shenglin Zhao (Ph.D.)
- Jiani Zhang (Ph.D.)
- Tong Zhao (Ph.D.)
- Looking for more PhD students working on machine learning, Big Data, social computing, …

Acknowledgments

- Brian Lau
- Patrick Lau
- Jesse Lee
- William Leung
- Cathy Li
- Lin Tsang
- Albert Yang
- Raymond Yuen
- Owen Zhang

Big Education in the Era of Big Data by Irwin Ki



CUHK Excellence

- The only university in Hong Kong having Nobel Laureates as faculty with five Distinguished Professors-at-Large

Prof. Yang Chen-Ning, Nobel Laureate in Physics

Prof. Charles Kao Nobel Laureate in Physics

Prof. Sir James A. Mirrlees, Nobel Laureate in Economic Sciences

Prof. Yau Shing-Tung, Fields Medalist

Prof. Andrew Yao, Turing Award

- **Nine** academicians of Chinese Academy of Sciences and Chinese Academy of Engineering

The Chinese University of Hong Kong

Q&A