The background features a collage of protein structures. In the top left, there are white alpha-helices. In the top right, there are grey DNA double helices. In the middle right, there is a yellow and red space-filling model of a protein. In the bottom right, there is a large, multi-colored (red, yellow, pink) space-filling model of a protein.

New Online Curriculum: The PDB Pipeline & Data Archiving

Cathy Lawson, Rutgers University

July 22, 2018 / ACA-Toronto

BD2K Call for Proposals

- Develop open educational resources for sharing, annotating and curating "Biomedical Big Data"
- Target audience: librarians/instructors, for training biomedicine students and researchers

Enabling Data Science in Structural Biology (eDSB) Project

- Unique opportunity to document RCSB's development and management practices

Project Goals are Consistent with RCSB's Educational Mission

- Promote understanding of biomolecules and PDB
- Provide a structural view of biology and medicine

The screenshot displays the PDB-101 website interface. At the top, a navigation bar includes links for 'PDB-101', 'Molecule of the Month', 'Browse', 'Learn', 'Global Health', 'Teach', 'Geis Archive', 'Events', and 'More'. Below this is a search bar for 'Molecule of the Month articles and more' with a 'Go' button. The main header features the 'RCSB PDB-101' logo and the tagline 'Molecular explorations through biology and medicine', along with the URL 'http://pdb101.rcsb.org' and social media icons. A left sidebar titled 'Learn' lists various resources: Paper Models, Flyers, Posters & Other Resources, Videos, Interactive Animations, Coloring Books, Education Corner, Guide to Understanding PDB Data, and Structural Biology Highlights. The main content area is titled 'Learn: Education Corner' and includes a paragraph about the quarterly newsletter and a submission email. Below this, three featured issues from 2018 are shown with their respective images and titles: Issue 76 (January 2018) about a journey as an intern, Issue 77 (April 2018) about creating sculptural models, and Issue 78 (July 2018) about science policy.

PDB-101 Molecule of the Month Browse Learn Global Health Teach Geis Archive Events More

RCSB PDB-101 Molecular explorations through biology and medicine

Search Molecule of the Month articles and more Go

http://pdb101.rcsb.org

Learn

- Paper Models >
- Flyers, Posters & Other Resources >
- Videos >
- Interactive Animations >
- Coloring Books >
- Education Corner >
- Guide to Understanding PDB Data >
- Structural Biology Highlights >

Learn: Education Corner

Published quarterly in our [Newsletter](#), each *Education Corner* offers an account of how members of the community use the PDB to educate students. If you would like to submit an *Education Corner* column, please send an email to info@rcsb.org.

2018



ABRCMS 2017

Issue 76 - January 2018
From the PDB to Phoenix: My Journey as an RCSB PDB Intern by Jenna Abyad (Drew University) and Priscilla Marie Salcedo (California State University Northridge)



Issue 77 - April 2018
Creating sculptural models of proteins in a high-school engineering class by Keagan O'Mara (Bio-Med Science Academy STEM School)



Issue 78 - July 2018
From Structural Biology to Science Policy by Leah Cairns (Johns Hopkins University School of Medicine)

Project can Support Expanding Experimental Data Archives in Structural Biology

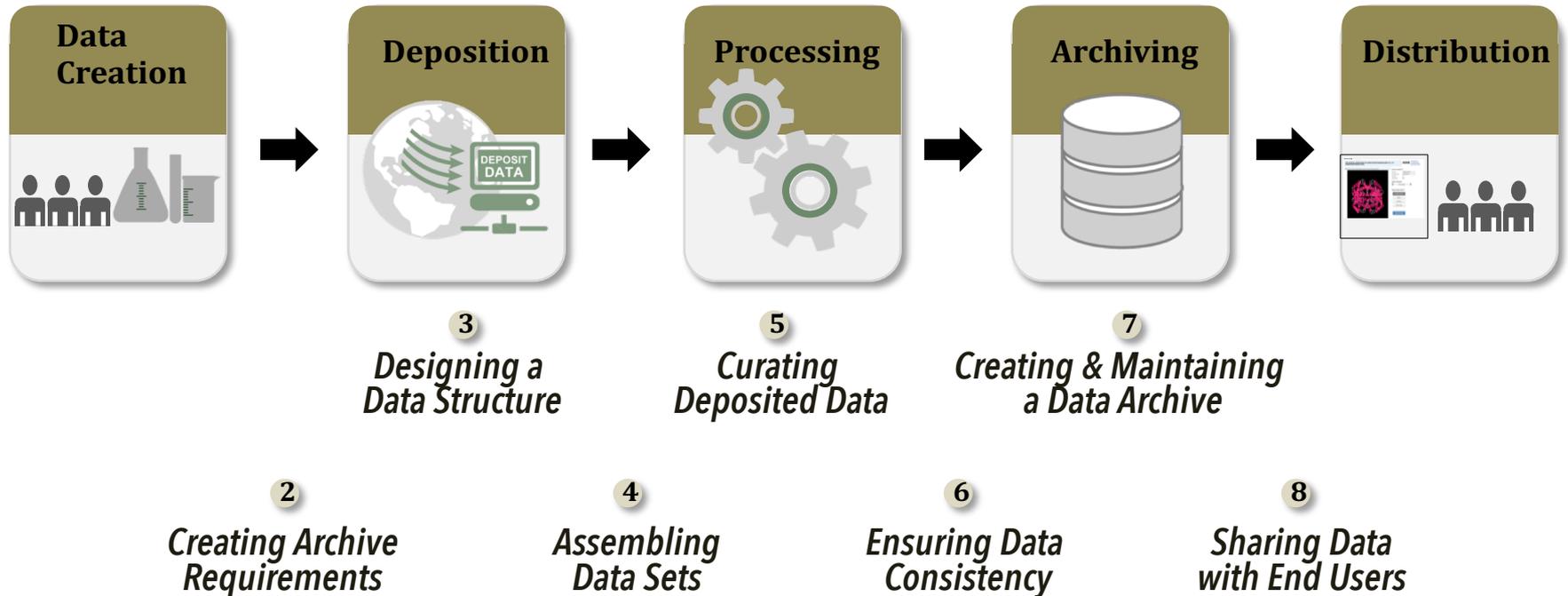
- Trained scientists are needed to develop federated data archives supporting new methods/model types (e.g. FRET, Mass Spectrometry ...)



Hybrid Methods Task Force EMBL-EBI, Hinxton 2014

Eight Curriculum Modules Follow the Data Pipeline

1 *Enabling Data Science in Biology: Overview*



Modules

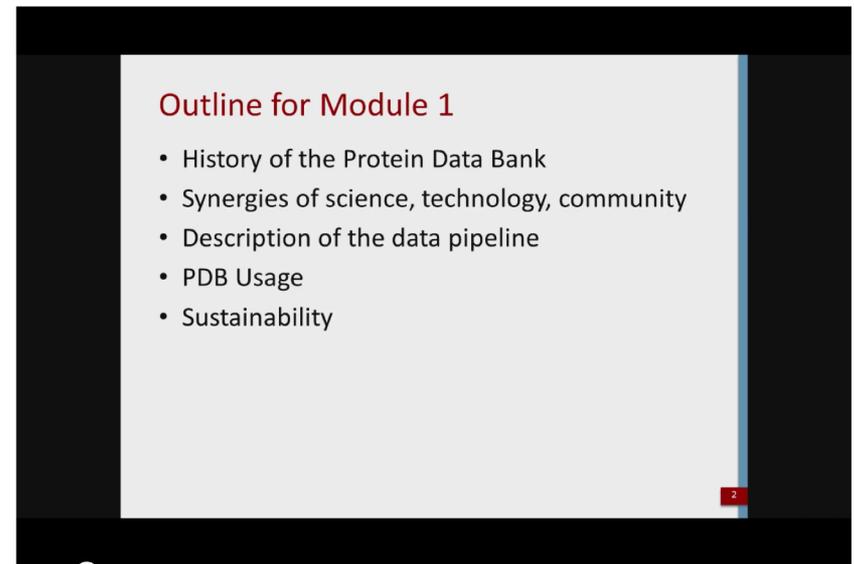
1. Overview	5. Curating the Data
2. Creating Archive Requirements	6. Ensuring Data Consistency
3. Designing the Infrastructure	7. Creating and Maintaining an Archive
4. Data Deposition	8. Data Distribution

Learning Objectives/Skills

- Recognize what is involved in designing and maintaining an archive for shared data
- Identify key stakeholders
- Develop requirements for what data to include
- Understand how to develop a data dictionary with the appropriate level of granularity
- Construct a deposition and annotation workflow based on a data dictionary

Lectures

- Lectures were developed and delivered by RCSB PDB group members according to their expertise
- 3-5 video segments per Module
- Transcripts were carefully curated to support closed-captioning



Exercises/Homework

- Students are guided step-by-step to design, create, and query a database on a topic of their own interest
- Exercises introduce tools needed to complete assignments
- Worked example included in all assignments

Homework Flow

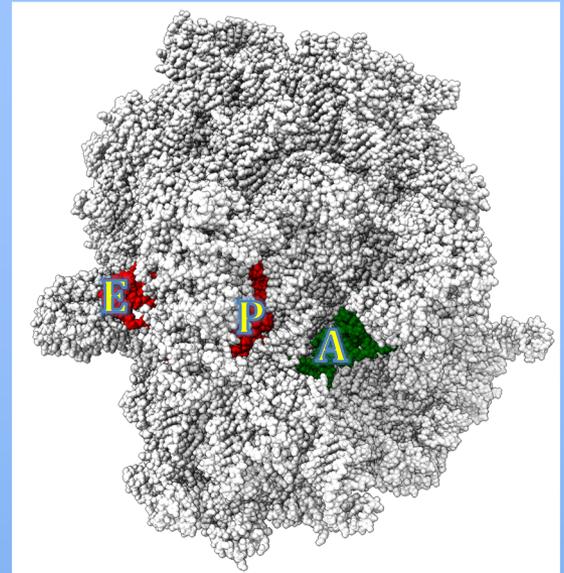
Module	Goal
1	Select set of PDB entries on topic of interest (50-100)
2	Create PDB data reports, get primary citations
3	Define questions about your topic, create new data terms
4	Create a deposition form for your new terms and fill it in
5	Review validation reports for your PDB entries
6	Check filled data for errors
7	Create a database combining PDB data and your new data
8	Perform queries to answer the questions about your topic

Tools used:

RCSB PDB website search/browse/reports, simple text editor, Excel or equivalent, Google Forms, MySQL Server and MySQL Workbench

Worked Example

- Recent *E. coli* ribosome cryoEM structures (61)
- Example Questions:
 - How many structures have both ribosomal subunits?
 - Which structures include messenger RNA?
 - What type of tRNA is bound in the P (peptidyl) site?
 - Do ribosome structures with bound antibiotics have good model quality?



Distribution of tRNA types in the peptidyl site of recent *E. coli* ribosome structures:

COUNT(pdb_id)	p_site_trna_aa_type
1	Glycine
2	Aspartate
3	Proline
4	Unknown
17	none
34	Initiator Methionine

Initial Implementation

- The curriculum was pilot-tested at Rutgers in Spring 2016, and then again in Spring 2018
- Students included:
 - Rutgers Graduate Students (Chemistry, Mol Bio)
 - Information Specialists from Rutgers Libraries
 - International Scientists interested in developing data archives

Use in a Flipped Classroom



Dissemination

- All materials will be accessible via PDB-101 and <http://edsb.rcsb.org>
 - Lectures: Slides, Transcripts, Videos
 - Exercise, Homework Slides
 - Links to Additional Resources
 - Licensing: Attribution-NonCommercial-ShareAlike 4.0 International
- Coursera MOOC: to be developed



Project Personnel



Catherine Lawson
Project PI



Helen M. Berman
Pilot 1 Lead



Maggie Gabanyi
Video Production Lead

Our
Advisory Committee
Members:

[Michael Lesk](#)
[Jill Trehella](#)
[Ann Watkins](#)



John Westbrook



Jasmine Young



Shuchismita Dutta



Brian P. Hudson



Ezra Peisach



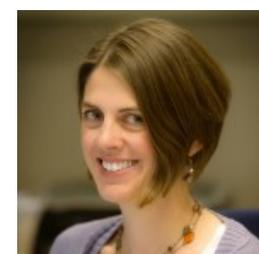
Peter Rose



Jose Duarte



Stephen K. Burley



Amy Sarjeant
CCDC

Interested in using
this curriculum?

Let us know:

edsb@rcsb.org