

Unboxing the Repository Model

Princeton Data Commons (PDC) Research Data Repository,
a use case for thinking differently

Hector Correa – Senior Library Software Engineer

Hannah Hadley – Manager, Open Publishing & Repository Services

Example of a “Boxed” Model

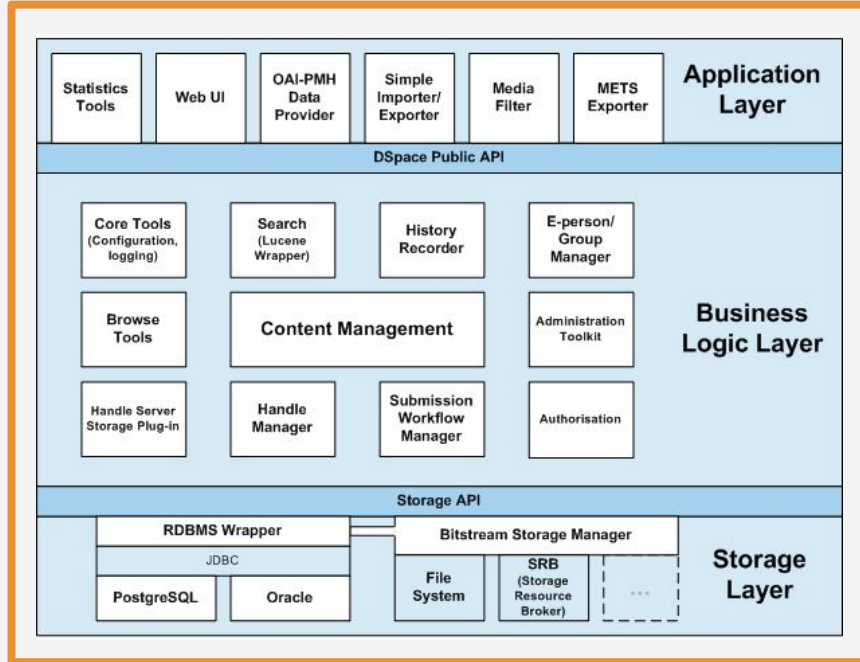


Image shows: [DSpace version 5.x](#)

- Centralized design
- Components are indivisible from the whole system
- Inflexible for our needs



Infrastructure Challenges

- Customized DSpace 5x
- Inflexible metadata options
- Curatorial workflow was rigid and had too many steps
- Increasing need to support larger submissions, including to move them for curation and publication.
- No dedicated development team

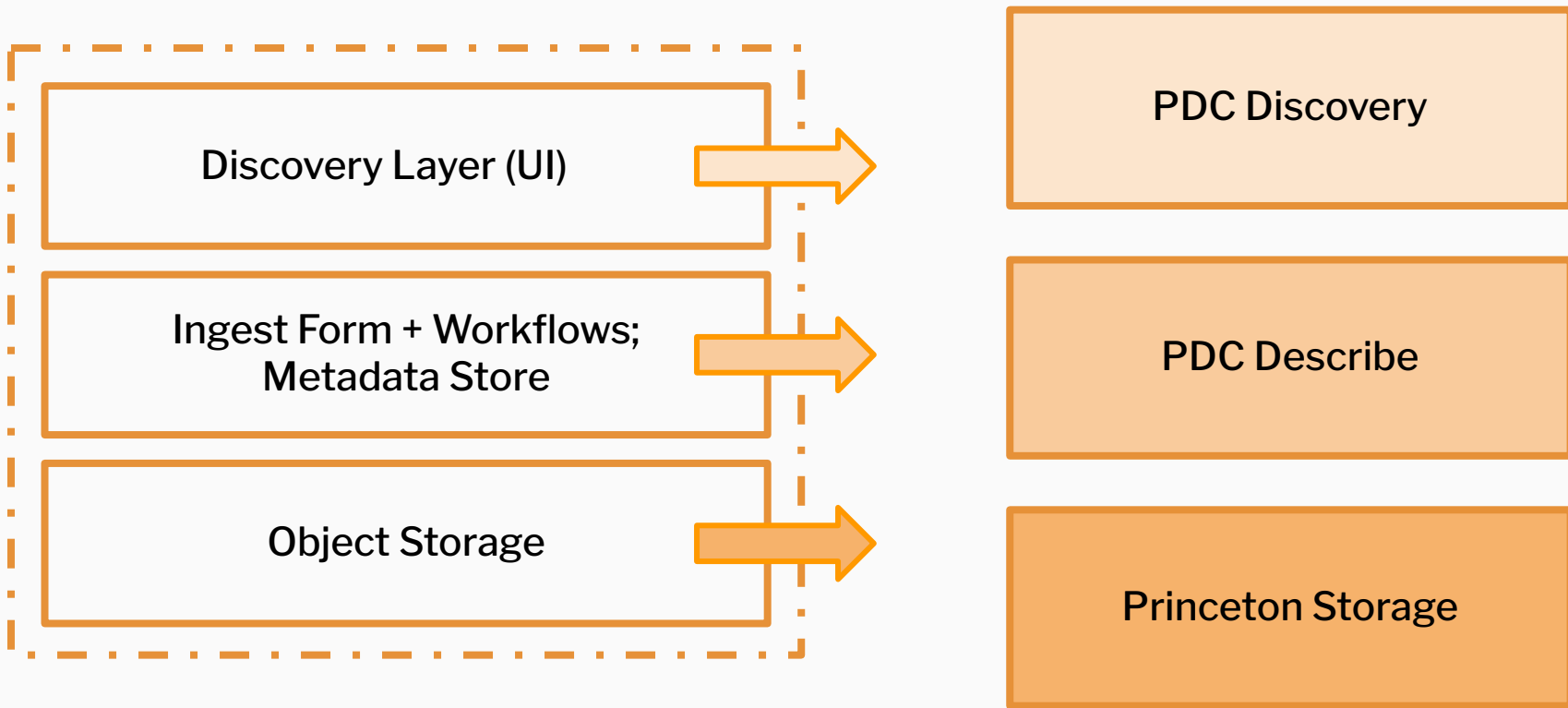
Identify the Essential Functions

Discovery Layer (UI)

Ingest Form + Workflows;
Metadata Store

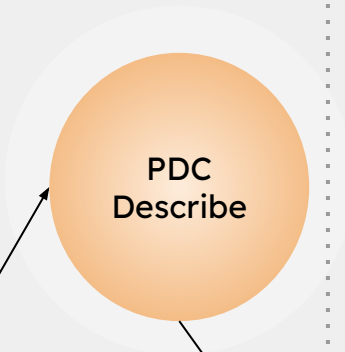
Object Storage

Unboxing the Model





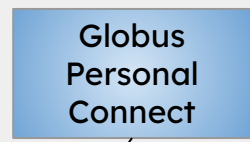
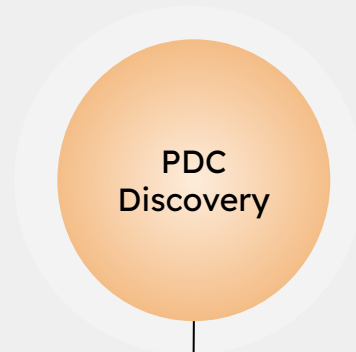
Princeton
Researcher interacts
with Data Curators



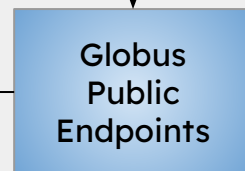
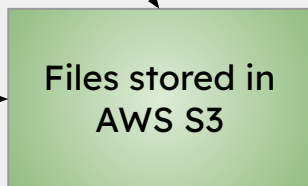
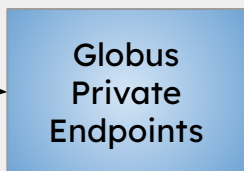
small files



Available Data
for End Users



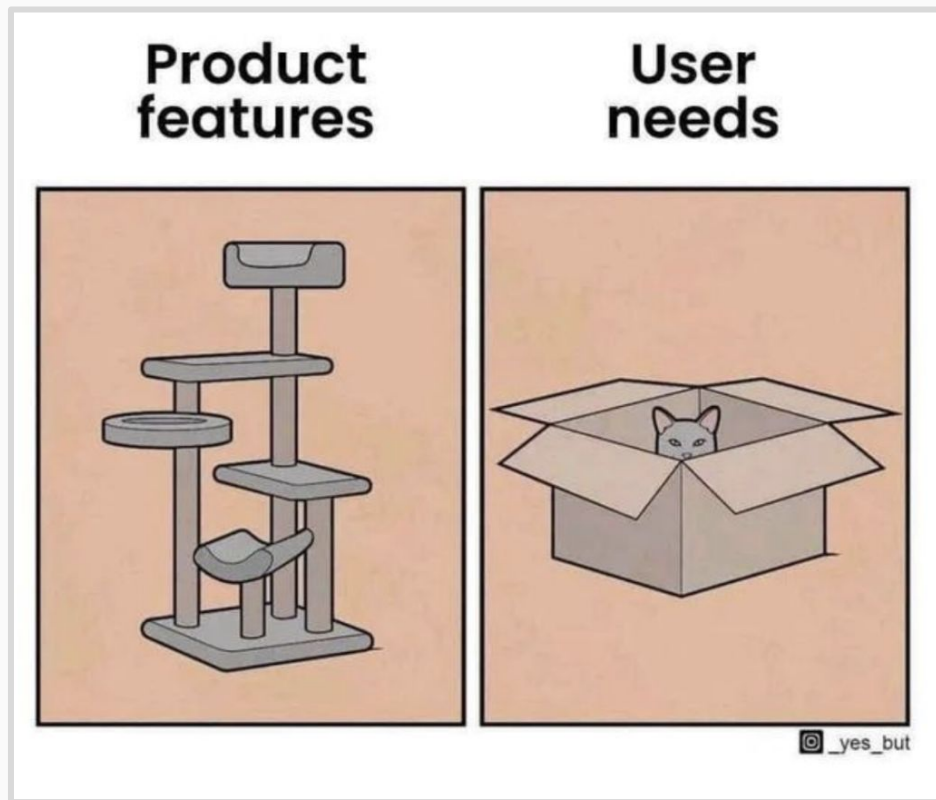
large files



Collaborative Challenges

Different expectations for an MVP

1. Stakeholders wanted all “the best” features
2. **Curators minimally needed features to operate their service.**
3. Developers wanted a more minimal design to start with, so iterations would test what users really wanted.



“Any software system you use to enable you to preserve and provide access to digital content is by necessity temporary. You need to be able to get your stuff out because it will likely not last forever.”

– Trevor Owens

Owens T. (2018). *The theory and craft of digital preservation*. Johns Hopkins University Press.



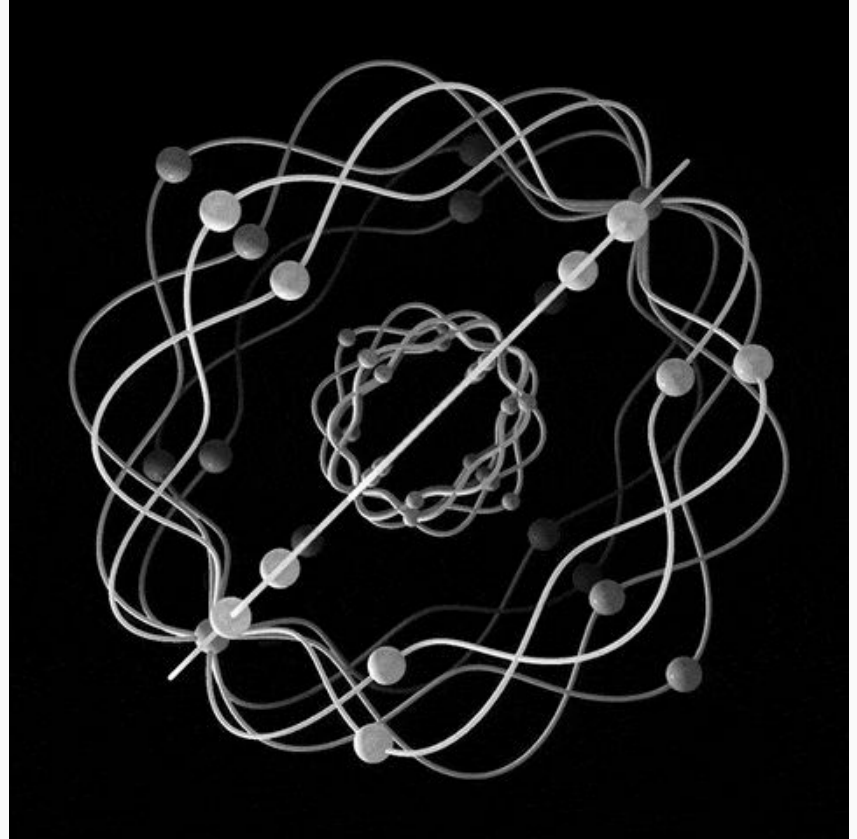
Image shows an automobile graveyard

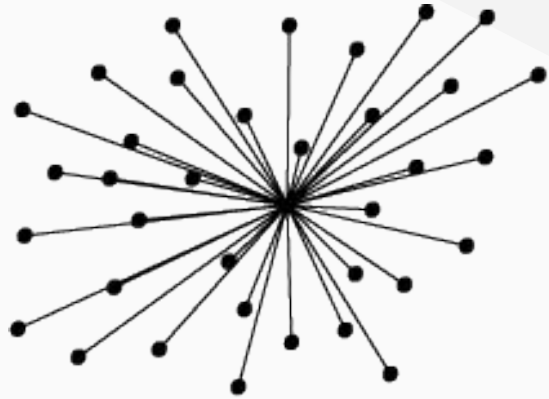
“Infrastructures are not things that can be built all at once, nor are they static. They are complex ecologies with many moving parts.”

– Christine L. Borgman

Borgman, Christine (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*

Xponentialdesign (2022). *Black and White Art GIF*. Giphy. Retrieved from: <https://giphy.com/gifs/loop-atom-sinusoidal-IUsAr6rYwWEhBEDGYv>

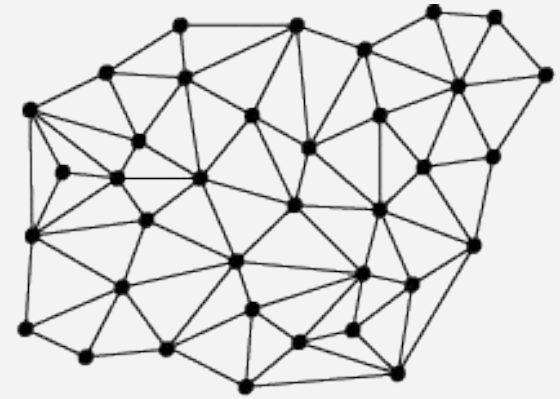




centralised

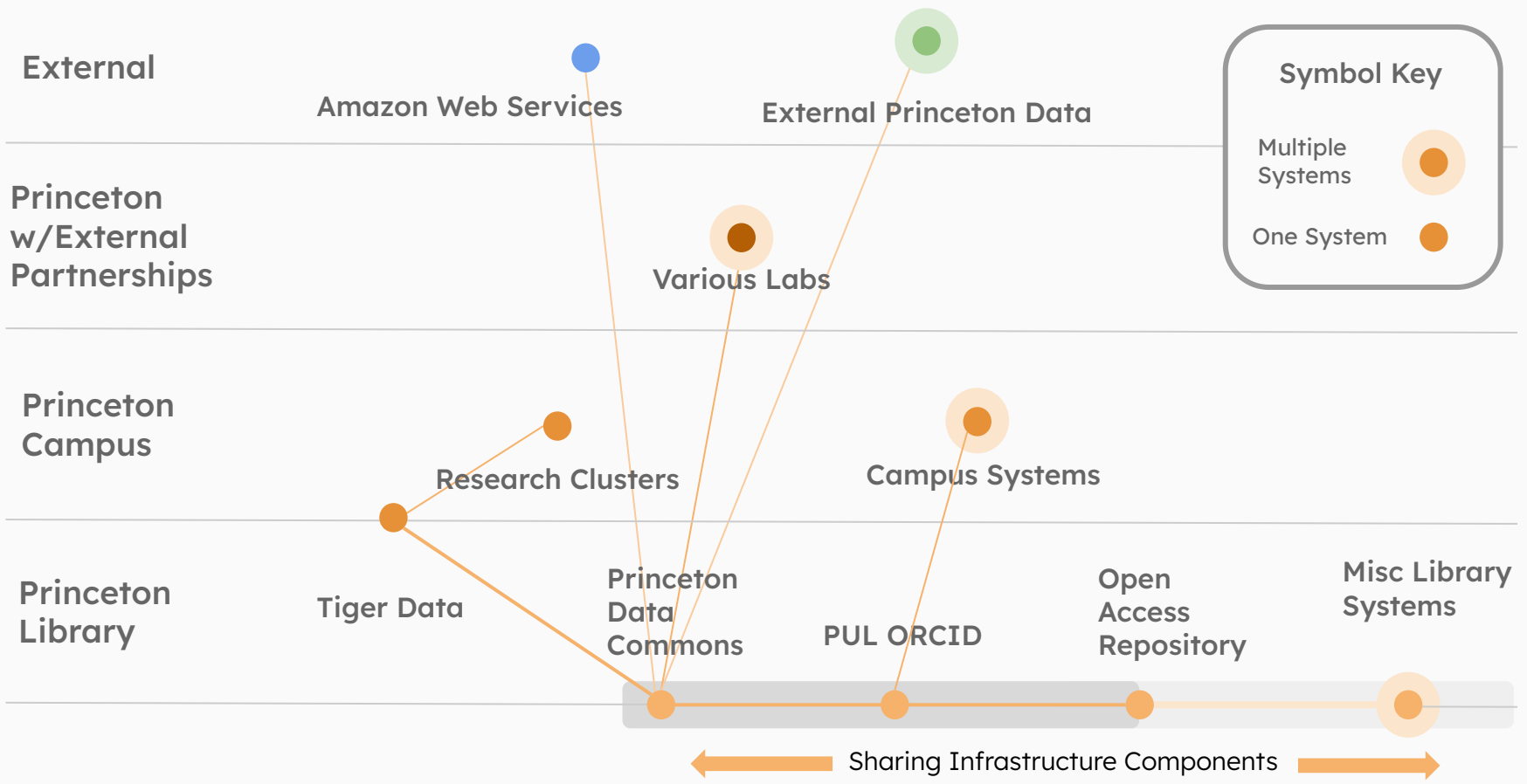


decentralised



distributed

Image by 1983-enwiki Retrieved from:
<https://commons.wikimedia.org/wiki/File:Centralised-decentralised-distributed.png>





Some Takeaways

- Prioritize listening and understanding across groups
- Consider flexibility in your design to address unknown future needs
- Address capacity issues in both human and technical infrastructure
- Think sustainably. Leverage what you have.

Thanks!

Hector: hector_correa@princeton.edu

Hannah: hhadley@princeton.edu