LIBRARY OF CONGRESS

September 23 2021

Partnering Across the Library of Congress on CIP e-book Processing During Expanded Telework

Cataloging in Publication (CIP) Program
Camilla Williams
Digital Content Management (DCM) Section:
Mark Cooper

LIBRARY LIBRARY OF CONGRESS

Agenda

- Introduction to CIP and CIP e-books
- 2020: LC pivoting to remote work
- Collaboration across Library divisions
- Results from project and increased automation
- Conclusions and future work

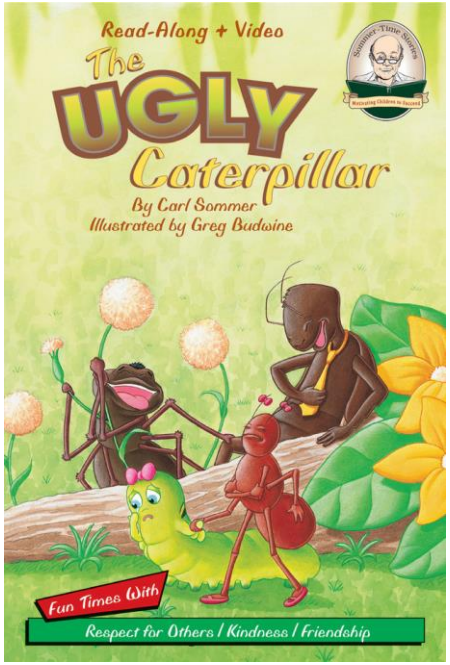
LIBRARY LIBRARY OF CONGRESS

Cataloging in Publication (CIP) Program

- Began in July 1, 1971
- **Mission**: to help libraries expedite and reduce the cost of cataloging the books they acquired for their collections
- Publishers submit an online application via PrePub Book Link
- Library of Congress staff cataloging the book and creates a bibliographic record and CIP data block
- Publisher prints the CIP data block on the Copyright page
- Publisher sends a copy of the book to the Library

CIP E-books Program

- Began in 2012 as a pilot
- Started creating e-book bibliographic records
- Began ingesting e-books in July 2014
  - The Ugly Caterpillar by Carl Summers
- FY14 – 100 e-books, 10 publishers
- FY20 – 49,413 e-books, 120 publishers and aggregators
  - 37,475 bulk delivery e-books through PCN



CIP E-Book File Delivery Pathway

- Publisher delivers e-book files
- LC content management system inventories files
- Attempts to use ISBN from filename to identify an existing e-book ILS record
- Task is created for staff to review the e-book file, make cataloging updates, and confirm acceptance of delivery
- Accepted deliveries move to preservation storage

LIBRARY LIBRARY OF CONGRESS

CIP E-Book Delivery Mechanisms

- Process originally conceived specifically for CIP e-book deliveries expanded to other types of e-book content
- Content delivered to CIP e-book file stream now includes:
  - E-books with CIP e-book ILS records
  - E-books with CIP print ILS records
  - E-books without ILS records

LIBRARY LIBRARY OF CONGRESS

Staff Starting Telework, March 2020

- Rapid switch to telework projects
- MARCEdit to create e-book records
- Trained staff in ABA, Music and G & M divisions to process e-books and to create e-book bibliographic records
- Manually assigning review tasks to a much larger number of staff

LIBRARY LIBRARY OF CONGRESS

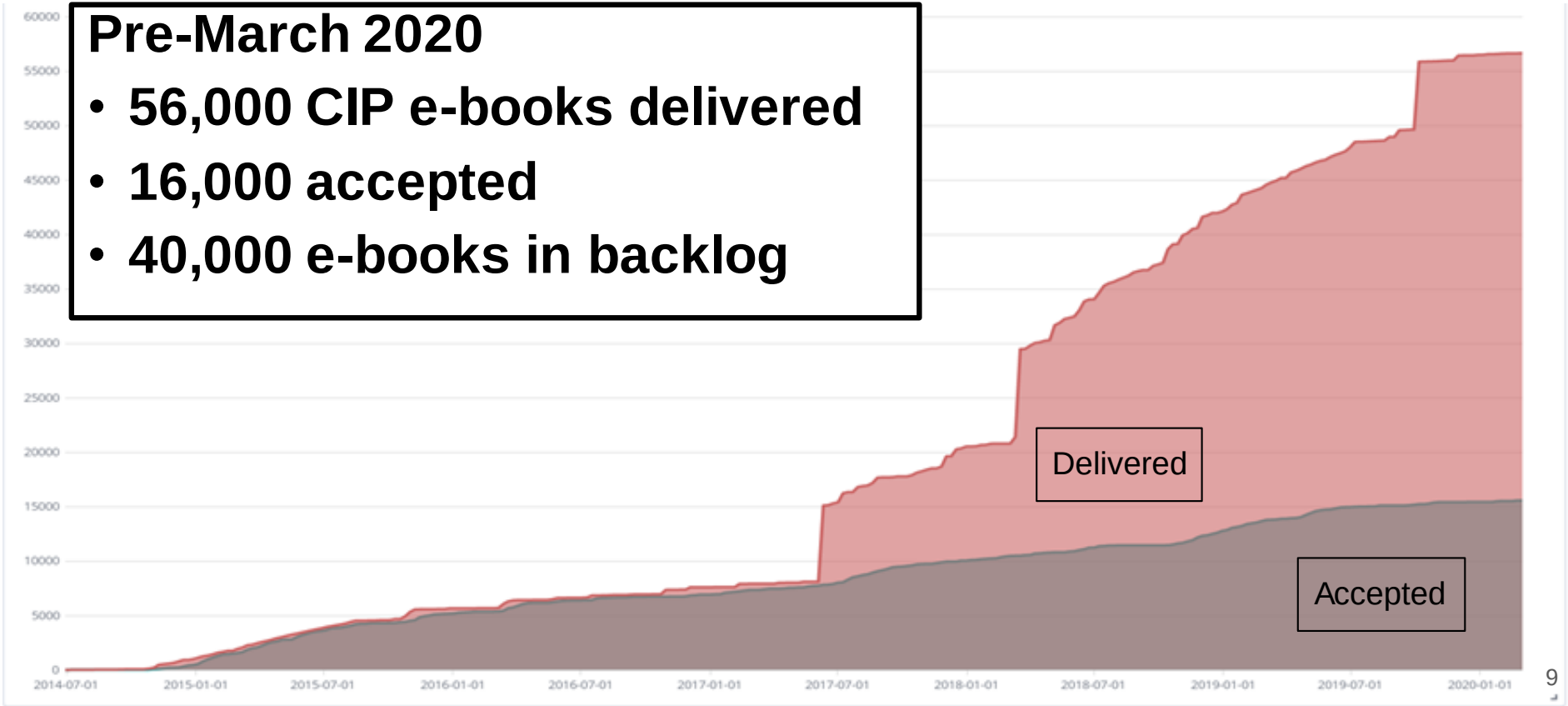
Collaboration

- CIP Program – Caroline Saccucci, Camilla Williams
- Digital Content Management (DCM) – Mark Cooper, Lauren Seroka, Mark Lopez
- Integrated Library System Program Office (ILSPO) – Dave Reser
- DCMS/ILSPO/CIP collaboration and support:
  - Leveraged batch CTS capabilities
  - Reviewed unaccepted e-book backlog
  - Helped provide a steady stream of work to newly remote staff
  - Conducted ILS record cloning & analysis

LIBRARY LIBRARY OF CONGRESS

Pre-March 2020

- 56,000 CIP e-books delivered
- 16,000 accepted
- 40,000 e-books in backlog



LIBRARY LIBRARY OF CONGRESS

Reviewing Existing Processes

- Initial focus on tracking current processes for ingest, processing, description, and preservation of CIP e-books
- Examined current processes, types of files being delivered, and associated metadata
- Criteria for acceptance and preservation using the Library's permanent preservation inventory system

LIBRARY LIBRARY OF CONGRESS

Challenges and Opportunities

- Processes conceived for largely manual work; limited developer resources in short-term
- Immediate availability of large numbers of regularly analog staff for digital work
- Newly developed presentation system for onsite access of restricted content provides access pathway (Stacks; released November 2019)
- DCM entering its third year with a focus on improving batch capabilities of existing systems

LIBRARY LIBRARY OF CONGRESS

Leveraging Batch Processes

- First priority: providing meaningful work to newly remote staff
- DCM had been building a library of Python scripts to interact with LC's inventory system
- Quickly implemented processes to review existing e-book acceptance tasks and assign to staff for completion, expanding upon existing Python script functionality

LIBRARY LIBRARY OF CONGRESS

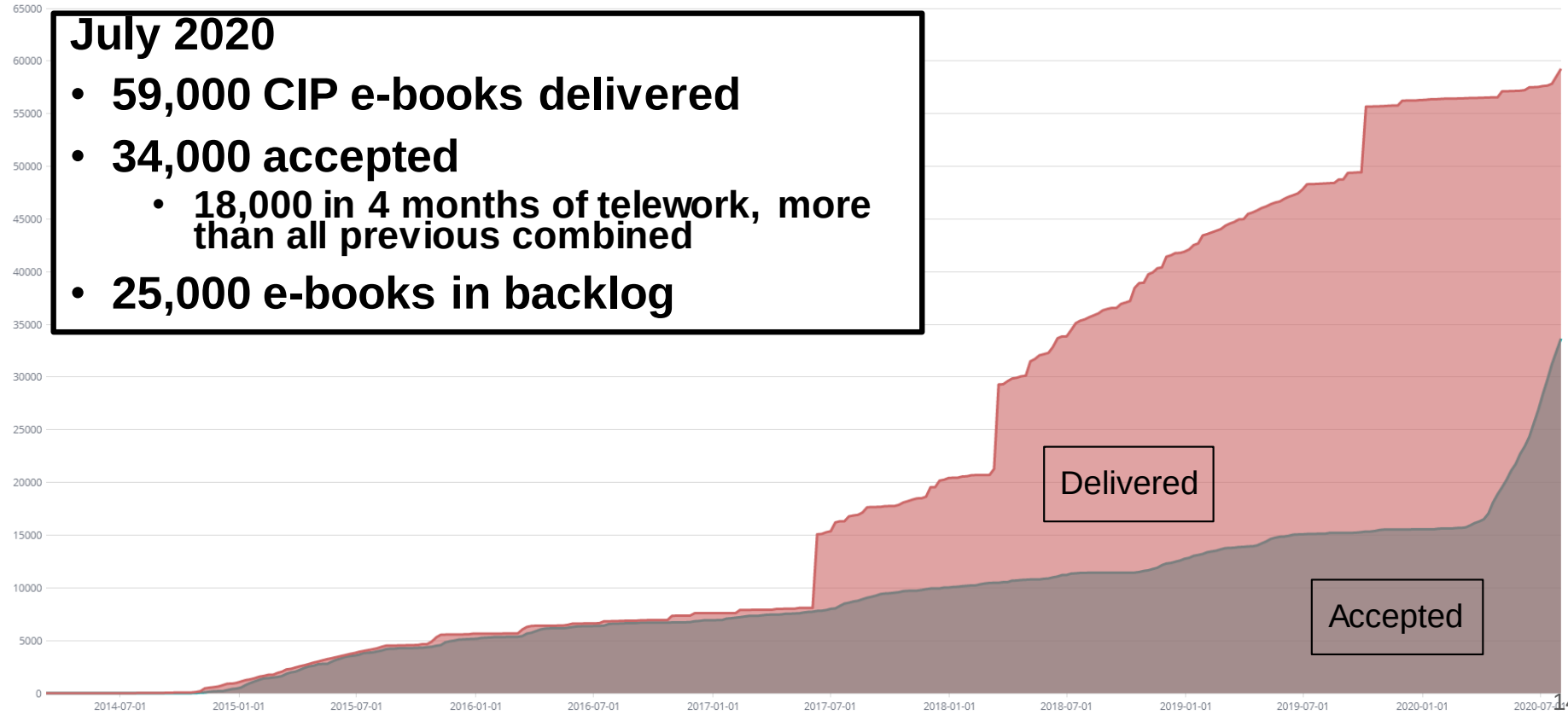
Innovating on Existing Processes

- Analysis of incoming books; extracting embedded file metadata to match books to records
- Determining if e-book or print LC records exist from limited data points
  - Using OCLC and other record sources to identify ISBNs and other information to track down existing records
- Identifying duplicate or alternate edition deliveries

LIBRARY OF CONGRESS

July 2020

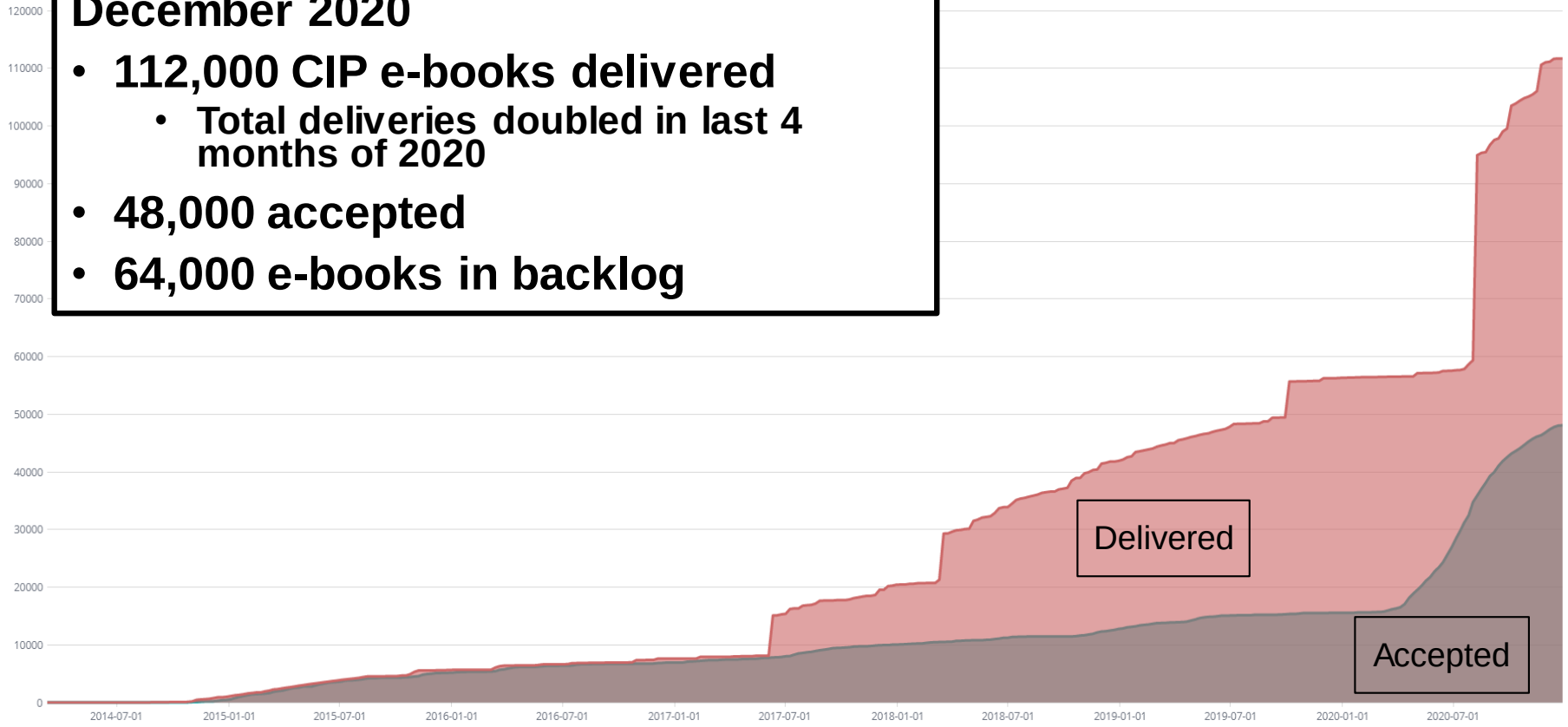
- 59,000 CIP e-books delivered
- 34,000 accepted
  - 18,000 in 4 months of telework, more than all previous combined
- 25,000 e-books in backlog



LIBRARY LIBRARY OF CONGRESS

December 2020

- 112,000 CIP e-books delivered
  - Total deliveries doubled in last 4 months of 2020
- 48,000 accepted
- 64,000 e-books in backlog



LIBRARY LIBRARY OF CONGRESS

Batch ILS Record Challenges

- Many with completed CIP e-book ILS records; others with only CIP print records, or no identifiable record
- Requirements of LC record distribution set initial limits on batch processes
- Policy approval to clone fully cataloged, distributed print records into distributable e-book records
- To accept without manual review, need high confidence that the file was what it purported to be

LIBRARY LIBRARY OF CONGRESS

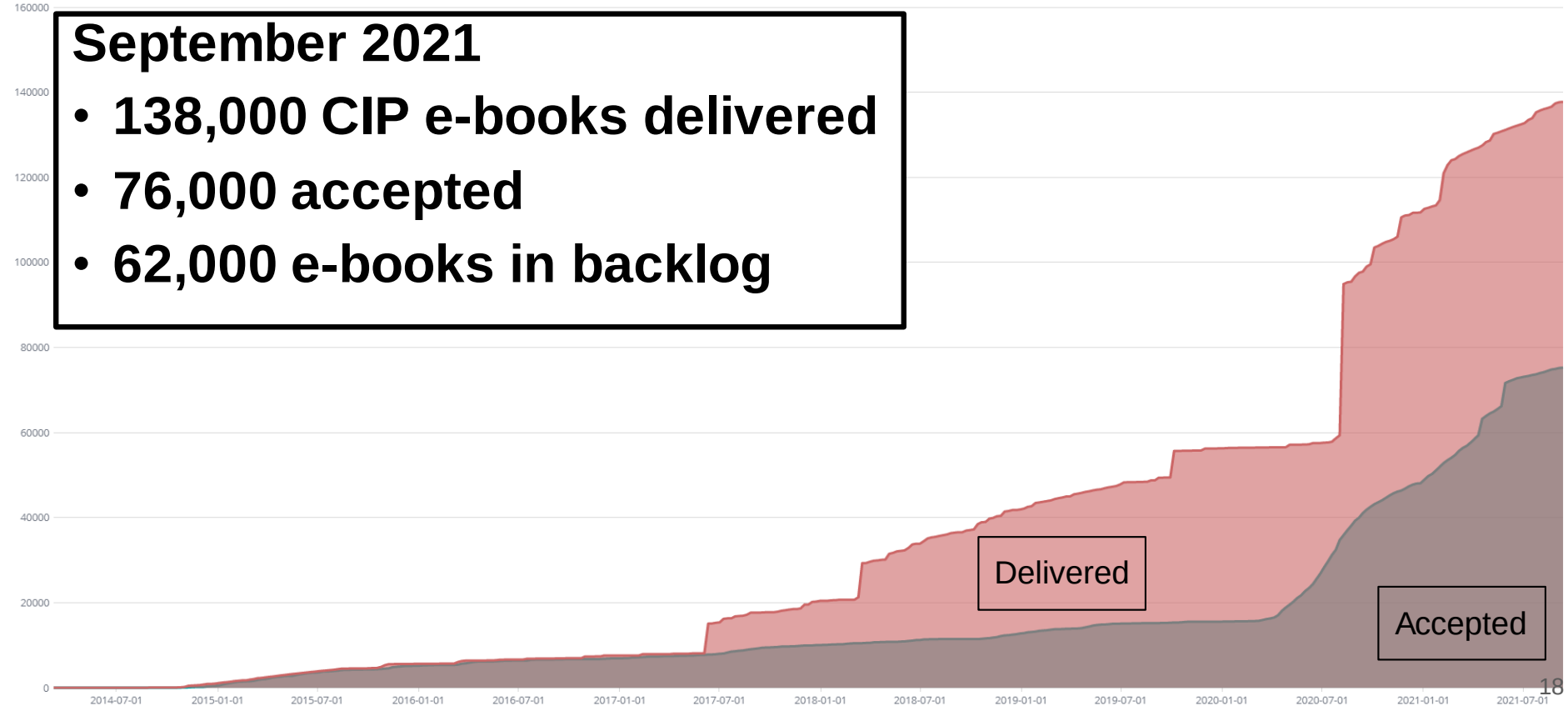
Bulk ILS Record Creation

- Expanding policy and practice for bulk e-book record cloning and creation
- PCN delivery: single publisher; 37,000 e-books
  - Records created utilizing vetted publisher-provided metadata
  - Content and records processed in bulk; accessible in Stacks
- Expanding and standardizing across other LC e-book delivery streams, including Copyright, Purchase, and Open Access

LIBRARY LIBRARY OF CONGRESS

September 2021

- 138,000 CIP e-books delivered
- 76,000 accepted
- 62,000 e-books in backlog



LIBRARY LIBRARY OF CONGRESS

CIP E-Books Accessible through LC Systems

- Over 110,000 e-books accessible onsite in Stacks
  - 72,000 CIP e-books
  - 36,000 through related PCN bulk delivery process
- CIP Open Access books accessible on loc.gov Open Access Books digital collection

LIBRARY LIBRARY OF CONGRESS

Thank You

- Camilla Williams: cewi@loc.gov
- Mark Cooper: maco@loc.gov