responsible operations
data science, machine learning, and ai in libraries

oc.lc/responsibleoperations

thomas padilla, oclc research
sarah shreeves, university of arizona
1. what & how
2. a path
3. next steps
... a library community research agenda that charts a path for (responsible) operationalization of data science, machine learning, and AI.
advisory group

Kenning Arlitsch, Montana State University
Jon Cawthorne, Wayne State University
Karen Estlund, Colorado State University
Josh Hadro, IIIF Consortium
Bohyun Kim, University of Rhode Island
Trevor Owens, Library of Congress
Ben Schmidt, Northeastern University
😎🌵 Sarah Shreeves, University of Arizona🌵😎
MacKenzie Smith, University of California Davis
Claire Stewart, University of Nebraska Lincoln
Melissa Terras, University of Edinburgh
Diane Vizine-Goetz, OCLC Research
John Wilkin, University of Illinois at Urbana Champaign
Kate Zwaard, Library of Congress
landscape group

Ruth Ahnert, Queen Mary University of London
Taylor Arnold, University of Richmond
Helen Bailey, Massachusetts Institute of Technology
Ted Baldwin, University of Cincinnati
Daina Bouquin, Harvard University and the Smithsonian Institution
Karen Cariani, WGBH
Michelle Cawley, University of North Carolina Chapel Hill
Rumman Chowdhury, Accenture
Jason Clark, Montana State University
Nicole Coleman, Stanford University
Rebecca Dikow, Smithsonian Institution
Quinn Dombrowski, Stanford University
Virginia Dressler, Kent State University
Jon Dunn, Indiana University
Ixchel Faniel, OCLC
Maggie Farrell, University of Nevada Las Vegas
Lisa Federer, National Institutes of Health

Barbara Fister, Gustavus Adolphus College
Kathleen Fitzpatrick, Michigan State University
Themba Flowers, Yale University
Alex Gil, Columbia University
Jean Godby, OCLC
Tiffany Grant, University of Cincinnati
Jane Greenberg, Drexel University
Harriett Green, Washington University St Louis
Umi Hsu, ONE Archives Foundation
Richard Johansen, University of Cincinnati
Bohyun Kim, University of Rhode Island
Lauren Klein, Georgia Tech
Emily Lapworth, University of Nevada Las Vegas
Shari Laster, Arizona State University
Matthew Lincoln, Carnegie Mellon University
Meris Longmeier, The Ohio State University
Dominique Luster, Carnegie Museum of Art
Nandita Mani, University of North Carolina Chapel Hill
And more!
1 year of activity
60 hours of interviews
1 face to face event
3 conferences
144 challenges
150 comments
7 drafts
an ode to big data, GPUs this is not
nor will you hear unfettered enthusiasm for “scale”
Microsoft denied police facial recognition tech over human rights concerns

The company has sold the technology to at least one US prison though

By James Vincent | Apr 17, 2019, 5:16am EDT

Amazon shareholders will vote to ban facial recognition tech

Amazon's board opposes it and wants government regulation instead.
rather
responsible operations

a commitment to fostering individual, organizational, and community capacities for responsible operationalization of data science, machine learning, and AI.

h/t rumman chowdhury for the inspo
getting from concept to operations
Responsible Operations: Data Science, Machine Learning, and AI in Libraries

Thomas Padilla

oc.lc/responsibleoperations
doi.org/10.25333/xk7z-9g97
7 areas of investigation

18 challenges

51 recommendations
1. Committing to Responsible Operations
   A. Managing Bias
   B. Transparency, Explainability, and Accountability
   C. Distributed Data Science Fluency
   D. Generous Tools

2. Description and Discovery
   A. Enhancing Description at Scale
   B. Incorporating Uncertain Description
   C. Ensuring Discovery and Assessing Impact
3. Shared Methods and Data
   A. Shared Development and Distribution of Methods
   B. Shared Development and Distribution of Training Data

4. Machine-Actionable Collections
   A. Making Machine-Actionable Collections a Core Activity
   B. Broadening Machine-Actionable Collections
   C. Rights Assessment at Scale

5. Workforce Development
   A. Investigating Core Competencies
   B. Committing to Internal Talent
   C. Expanding Evidence-Based Training
6. Data Science Services
   A. Modeling Data Science Services
   B. Research and Pedagogy Integration
7. Sustaining Interprofessional and Interdisciplinary Collaboration
1. what & how
2. a path
3. next steps
1a - managing bias

3a - shared development and distribution of methods

4b - broadening machine actionable collections

5b - committing to internal talent
1a - managing bias

3a - shared development and distribution of methods

4b - broadening machine actionable collections

5b - committing to internal talent
responsible operations call for sustained engagement with human biases manifest in training data, machine learning models, and outputs.
Bias management activities have precedent and are manifest in collection development, collection description, instruction, research support, and more.
Discriminating Systems: Gender, Race, and Power in AI

April 2019

monoculture cannot effectively manage bias.

diversity is not a nice to have, it is an imperative.
a recommendation

Explore creation of a “practices exchange” that highlights successes as well as notable missteps in cultural heritage use of data science, machine learning, and AI. Commit to transparency as a means to work against repeated community mistakes — a pattern of negative behavior in Silicon Valley that Jacob Metcalf, Emanuel Moss, and danah boyd have referred to as “blinkinged isomorphism.”
1a - managing bias

3a - shared development and distribution of methods

4b - broadening machine actionable collections

5b - committing to internal talent
LIBRARIES USING COMPUTER VISION IS GOOD?

SO MANY THINGS!
venues and mechanisms for refinement are few
impacts assessment & broader uptake
a recommendation

Develop venues, publication outlets, and funding sources that facilitate the sharing of methods and benchmarks for machine learning and artificial intelligence in the cultural heritage community.
1a - managing bias
3a - shared development and distribution of methods
4b - broadening machine actionable collections
5b - committing to internal talent
“we are in an imagination battle.”

adrienne maree brown
"Natural Language" Processing Digital Divide

As of August 2019:

961 resources for English
121 for American English
216 for German
180 for French
130 for Spanish
103 for Mandarin Chinese
103 for Japanese.

Remaining ~7000 languages have far fewer resources.

Emily M. Bender, The #BenderRule: On Naming the Languages We Study and Why It Matters
a recommendation

Prioritize the creation of machine-actionable collections that speak to the experience of underrepresented communities. Inform this work through collaborations with community groups that have ties to collections, subject experts, and reference to resources … like Design for Diversity. Per community input, decisions to not develop a machine-actionable collection are as positive as decisions to develop a machine-actionable collection.
1a - managing bias
3a - shared development and distribution of methods
4b - broadening machine actionable collections
5b - committing to internal talent
a recommendation

Form a working group to investigate the development of organizational models that avoid silos and support hybridity between core and emerging services - models of this kind may encourage natural diversification and/or deepening of skills over time.
1. what & how
2. a path
3. next steps
Responsible Operations: Data Science, Machine Learning, and AI in Libraries

Thomas Padilla

oc.lc/responsibleoperations
doi.org/10.25333/xk7z-9g97
7 areas of investigation

18 challenges

51 recommendations
reflections

oc.lc/responsibleoperations