Digital archives, big data and image-based culturomics for social impact assessment: opportunities and challenges

Kate Sherren [a], John R. Parkins [b], Michael Smit [c], Mona Holmlund [d], Yan Chen [a]

a School for Resource and Environmental Studies, Dalhousie University, Halifax, NS, B3H 4R2, Canada, kate.sherren@dal.ca

b Department of Resource Economics and Environmental Sociology, University of Alberta, Edmonton, AB, T6G 2H1

c School of Information Management, Dalhousie University, Halifax, NS, B3H 4R2

d Department of History, Dalhousie University, Halifax, NS, B3H 4R2

Last submitted version of Viewpoint for Environmental Impact Assessment Review, 67 (November), 23-30. Please see DOI: 10.1016/j.eiar.2017.08.002 for final published version.

Abstract

Social impact assessment (SIA) is well-established but uses conventional approaches that have become less effective in recent decades, particularly in relation to declining survey response rates and a lack of youth engagement. Images from digital archives and social media sources are poised to advance the research and practice of SIA by transcending text-based methods with insights into changing landscapes, and human engagement with them. This viewpoint describes progress, challenges and cautions towards the development of such tools (defined as culturomics), using hydroelectricity cases to illustrate potential approaches. These tools build on foundational work in a range of disciplines, including the humanities and computer science. We describe necessary advances in machine learning, image digitization, and data aggregation and visualization techniques, as well as ways to ensure that such tools are carefully tested, applied and interpreted. Challenges include the automation, acquisition and management of datasets, and using these tools appropriately and equitably. Critically, culturomics of any

kind must not be used as a replacement for engagement with people, but as complementary to inclusive stakeholder engagement.

Keywords

Archives; computational social sciences; cultural ecosystem services; digital humanities; social media; machine learning; volunteered geographic information

Highlights

- Image-based digital archives and social media present opportunities for new SIA tools.
- Application-ready big data approaches are emerging across many fields.
- SIA scholars, practitioners and stakeholders should engage with culturomics.
- Challenges include automation, digitization, interpretation and justice.
- Culturomics of any kind must not replace engagement with people.

Acknowledgements

The authors would like to acknowledge funding for Chen from the Social Sciences and Humanities Research Council of Canada (Grant #435-2012-0636, 2012-2016, Parkins PI) and a Nova Scotia Graduate Scholarship, and thank Elena Bennett for insightful comments on an earlier draft of this paper.

The need for new tools for SIA

Environmental assessment is a long-established practice with wide-ranging procedures for data collection and analysis, focused on the anticipation of impacts from policy or project proposals. Within social impact assessment (SIA) in particular, conventional tools for assessment include local surveys, analysis of secondary data (e.g., census data), interviews with key informants, and a range of methods for anticipating social impacts such as the calculation of multiplier effects, straight line analysis, and comparative case studies (Asselin and Parkins, 2009). Although there are some efforts to update these methods with more sophisticated approaches to public participation (Sinclair and Diduck, 2017) or simulation models (Karami et al., 2017), many of these advances remain within the realm of academic work, with limited application in the practice of SIA. Although the SIA literature has traditionally focused on *ex ante* assessments of impacts, this focus is changing with growing attention to *ex post* assessment where long-term policy or project monitoring comes into focus. This shift in the definition of SIA is reflected in the work of Vanclay et al. (2015), where SIA is described as:

> *"the process of identifying and managing the social issues of project development, and includes the effective engagement of affected communities in participatory processes of identification, assessment and management of social impacts" (Vanclay et al., 2015, p. iv).*

Although the identification and management of social issues is a point of focus in SIA literature, methods for doing this work have not advanced significantly in recent decades. SIA remains entrenched in conventional public engagement procedures and analysis using local household surveys and census data (Parkins and Mitchell, 2016). Such conventional methods are rendered increasingly ineffective because of steep decreases in survey response rates and associated growing biases in those who participate (Connelly et al., 2003, Stedman et al., 2016). Additionally, after a development is approved, follow-up social monitoring is rarely undertaken to confirm the predicted effects or to assess the effectiveness of impact mitigation.

A real weakness of SIA is the bias introduced by voluntary public participation, meaning younger demographics are often underrepresented (Checkoway et al., 2005). Though apathy toward politics and public affairs has generally increased over the past 30 years, people under the age of 25 show the biggest decrease in interest (Delli Carpini, 2000). Younger generations have very different conceptions of civic

engagement, but should not be ignored (Blandford et al., 2015). Much of the engagement around public issues that does occur within this demographic cohort – Generation Y or so-called Millennials who were born from the early 1980s to early 2000s – happens on social media where this generation documents their lifestyles and attitudes (Delli Carpini, 2000). Despite its increasing ubiquity, social media is generally only used in SIA to share information and occasionally monitor online sentiment (e.g. *Twitter*). Such uses are important for tracking opposition around specific projects but less nuanced when it comes to understanding the social and cultural implications of those proposals (Hanna et al., 2016).

In an increasingly digital world, practitioners of SIA have a growing opportunity to leverage a wider range of datasets, including digital images and associated text in archives and social media sources (Esteves et al., 2012). The term 'culturomics' encompasses this growing field of research, described as the use of text-based corpora, or collections, to understand culture. The inventors of this term leveraged the *Google Books* database to track the usage of words and phrases (and thus one lens on culture) over time (Michel et al., 2011). They liken culturomics to the microscope or telescope, as it makes visible for study a dimension that was previously largely hidden. Even using only the text associated with online activity, such as search engine queries and *Twitter* traffic, researchers have been able to track cultural trends and monitor environmental conditions (Di Minin et al., 2015, Ladle et al., 2016). Given the growth in image-dominated datasets, however, we are thinking very small indeed if we do not leverage the increasing volumes of online images, particularly those generated by citizens (Sherren et al., 2017). Unlike predominantly text-based sources such *Twitter*, which is more explicitly polemical, messages carried by landscape images on social media often *imply* landscape perceptions, preferences, and lifestyles. Research using social media is typically nonreactive, as subjects are unaware of being studied. This may not mean that such research is immune to biases such as social desirability. Social desirability oriented towards peers and other expected audiences will undoubtedly bring different perspectives to SIA, however, than bias introduced by government or proponent intervention. Image collections that provide a longer historical perspective are also held in other increasingly digital databases, such as newspapers/sites and archives.

Advances in publishing technology have come with the widespread means to easily produce and publish image, audio, and video and therefore a need to widen the lens through which we observe human/environment issues. We focus here on the potential of leveraging online images and technical

capabilities in image processing to improve SIA tools. We exclude here the understanding of environmental and social change that can be determined through analysis of aerial imagery such as from satellites, either using feature recognition or spectral analysis. Drawing on Vanclay's (2002) distinction between social impacts and social change variables, we see culturomics contributing to several social change domains. These include demographic processes (e.g., influx of tourists and seasonal workers, displacement and dispossession), and socio-cultural processes (e.g., perceived differences between community groups, changes in cultural values). Geographic processes are particularly relevant here, with attention to how social change accompanies land use conversion and diversification, urbanization, gentrification, changing transportation systems and other physical changes and uses of the landscape. Given the exponential growth in digital archives and the recent emergence of culturomics, in this commentary we describe the interdisciplinary state-of-the-art that is converging to enable new tools for SIA, using hydroelectricity as a case study, and describe some of the opportunities and challenges that accompany the development and use of such tools.

The interdisciplinary state-of-the-art

A 'pictorial' turn is afoot across many fields that is increasingly well-described (Bachmann-Medick, 2015) and already plumbed for insight in the arts, humanities and social sciences (e.g. Graham et al., 2011, Pink, 2003). It includes approaches to describing contemporary culture, as well as tracking cultural change through time. For the latter, we can build on approaches in art history that observe cultural shifts such as landscape perceptions and cultural norms (Halkes, 2006). More recently fields such as material and visual culture studies explore cultural changes in environmental behaviour through artefacts such as paintings, advertisements, television shows and everyday objects (Bennett and Joyce, 2010, Shove et al., 2007). These fields demonstrate that structured analysis of images can help us understand not only their contents and meaning, but also the emotional power that images acquire as they move between individuals and across cultures through various media (Belting, 2011). One can easily imagine what longitudinal insights could be established using increasingly digitized newspaper and archival image corpora.

The integration of social media into contemporary society provides emerging opportunities for describing today's culture. Images are being shared online with a frequency and density that is culturally unprecedented. These images reveal personal values, lifestyles, landscape uses and perceptions. Because

of shifting technologies – which social media software is being used, its uptake and demographic biases, and its internal rules for use – tracking change over time is less possible with social media. Images hold great potential, however, for understanding how the connections between features and values are made over space. Existing methods for doing this, for example photo-elicitation or values mapping with research participants (Brown and Fagerholm, 2015, da Silva Vieira and Antunes, 2014), are conceptually rich but time-consuming in both data collection and processing. Through passive (i.e. non-reactive) data collection, social media offers a clear opportunity to increase sample size and compensate for biases introduced by researchers using more active approaches (i.e. those *generating* rather than collecting data). Barry (2013), for example, demonstrated such image-based approaches can help develop nuanced understandings of public concerns about grazing on public lands, and help avert conflict over land management issues.

In recent years the techniques for automation in the processing of such data volumes, whether social media or archival, have matured in their sophistication and application. 'Big Data' refers to the phenomenon of having large volumes of data in a variety of unstructured formats accumulating at high velocity. The human brain can readily comprehend and analyze images and unstructured text, but only in relatively small amounts. Computer scientists are developing sophisticated algorithms to help in cases where we have too much data and not enough humans. Machine learning algorithms can learn how humans processed a set of images and text, and apply that understanding on a vastly larger scale. Image recognition algorithms, once trained, can generate descriptions of the events, objects, and interactions occurring in an image (Fang et al., 2015, Karpathy and Fei-Fei, 2015). Automated approaches can also be used to determine where photographers sharing work online actually live, even if their profiles are anonymous (Rugna et al., 2012). Combining images and caption text within analyses may indeed help balance the varying weaknesses of each source, including the use of colloquial language and intentional misspellings (Agarwal and Yiliyasi, 2010). Technology companies often deploy these algorithms for real-world applications through cloud-based web services (e.g. IBM Bluemix Watson APIs, Google Cloud Machine Learning, Microsoft Azure Machine Learning Studio). Application of digital media analysis also extend to emerging areas of computational impact assessment and social computing, for instance to assess the individual and broader impacts of social change documentaries and projects (Diesner et al., 2014, Diesner and Rezapour, 2015, Rezapour and Diesner, 2017).

Leveraging images to track social, environmental and cultural trends builds on advances in other fields, sitting at the nexus of several new and established methodologies. As mentioned above these ideas are related to the proliferation of Big Data and ubiquitous public connectivity, reflecting the turn toward digital humanities and computational social sciences (Cioffi-Revilla, 2010) as well as artistic and aesthetic data visualization (Kim and Park, 2013, Viégas and Wattenberg, 2007). In biophysical terms, such an undertaking is supportive of landscape sustainability science (Musacchio, 2013, Wu, 2013), and cultural ecosystem services (Chan et al., 2012). At the intersection of these is conservation social science (Bennett and Roth, 2015) and visual and landscape sociology (Beilin and Bohnet, 2015, Harper, 2012). It builds on volunteered geographic information (Goodchild, 2007), produced intentionally through citizen science or as through user content generated as a by-product of social media participation.

New modes of analysis and visualization

The high dimensionality of image-based culturomics may call for approaches to data analysis and visualization, such as spatial and conceptual mapping, aggregation and clustering that will be relatively new to SIA practitioners.

The largest and fastest moving body of research to date using social media-derived images to inform environmental decision-making is the mapping of cultural ecosystem services. Such cultural benefits of nature (e.g. aesthetics, recreation, inspiration, education, sense of place) have limited monetary proxies that can be included on balance sheets to inform decision-makers (Chan, Satterfield, 2012). Because of this limitations some researchers seek to map cultural ecosystem services so that they can be considered in planning decisions alongside other phenomena (Kopperoinen et al., 2017). These approaches include participatory mapping exercises as mentioned earlier, but also, more recently, the use of online secondary datasets such as photo-sharing sites and social media, including *Panoramio*, and *Flickr* (Martínez-Harms and Balvanera, 2012). These mapping tools demonstrate the utility of leveraging images, but also reveal the need for carefully understanding source data and methods. For instance, some online photo-sharing sites focus on intact 'natural' ecosystems even though that excludes human-modified landscapes where some cultural services are more prevalent (Braat and de Groot, 2012). Map-based approaches may also simply reveal observation bias created by existing roads and population densities. More recent work on cultural ecosystem service mapping takes a 'causal approach' (Plieninger et al., 2013), to understand what is driving the revealed values (Oteros-Rozas et al., van Zanten et al.,

2016). The application of such approaches in SIA tools would allow for the identification of hotspots (and their causes): of public values, lifestyles and preferences related to specific proposals.

It may be, however, that diagrammatic visualizations are more powerful for some SIA applications. Tourism researchers have developed approaches for the statistical association of themes and features coded from images and associated text (Stepchenkova and Zhan, 2013). The resulting landscape perception diagrams are conceptual, rather than spatial, and thus relatively robust to biases that result from differences in methods. Impacts to the landscape are one of the key challenges in matching development proposals to evolving societal values. Scholars, development proponents, opponents and decision-makers alike could use such diagrams to understand how landscape changes are likely to affect local residents. For instance, impacts may be revealed by identifying at the features that are expected to change and exploring the strength of associations to specific values or activities. Algorithms that can separate locals from visitors can provide additional insight, such as illustrating potential conflict around landscapes of production (livelihoods) and landscapes of consumption (leisure) (Gill and Reed, 1997, Walker and Fortmann, 2003).

Although most approaches to analysing social media involves translation into text (Bachmann-Medick, 2015), such as via the process of coding or classification (whether manual or automated), alternatives do exist. The most ground-breaking work of this kind is being carried out in the fields of artistic and aesthetic data visualization (Kim and Park, 2013), where non-text corpora are analyzed – averaged, clustered, visualized – without conversion to text at any point. They reveal the cultural patterns "too mundane, or too abstract" to have been written down (Ginosar et al., 2015, p. 1). For instance, over a decade ago artist Jason Salavon averaged Playboy centrefolds and yearbook photos, by decade, to illustrate shifting styles, ideals and behaviours. More recently Ginosar and Rakelly (2015) tackled yearbook images computationally, identifying high school 'types' within each decade using clustering algorithms, and tracking increasing smile curvature. Landscape presents similar opportunities. In 2008 Salavon simulated a number of photos of Chicago based on favourite tourist views, and overlaid them using transparency (Figure 1) (Viégas and Wattenberg, 2007). Today, such photos would not need to be simulated, thanks to enthusiastic photo sharing online. Automated filtering and clustering of such images could reveal archetypal views of the landscape, informing land management decisions and SIA.

Figure 1 Jason Salavon, *City (westward)*, 2008, Retrieved from *http://salavon.com/work/City/* and used with the permission of the author. This art piece was created by adding together simulations of popular tourist views of Chicago.

Hydroelectricity case studies

It is an interesting time for hydroelectricity, with dam construction and dam removal happening in parallel, neither without controversy (Sherren et al., in review). We do not seek here to endorse either, but to use two Canadian hydroelectricity proposals as an opportunity to explore the value of engaging with public images in the practice of SIA. Text-based analysis of *Instagram* and *Twitter* coverage of these two proposals demonstrated not only the different ways these social media platforms are used – *Twitter* to share opinions, *Instagram* to share lifestyles – but also the limitations of purely text-based assessments (Chen, 2015). By contrast, image-based analyses have driven new insight. The methods used to develop these visualizations are fully described in the references given below, and are only briefly outlined here.

A new large dam is under construction at what is called Site C, the controversial third dam on the Peace River near Fort St. John, British Columbia. Recently Chen et al. (in review) has manually coded, and statistically associated, features, activities and values evident in a year's worth of landscape *Instagram* images and captions from the planned Site C headpond area (Figure 2). This approach is a nonreactive mirror of cognitive approaches to understanding sense of place, where respondents are asked in surveys about place attachments (connection, utility, personal identity) and how those relate to what the place physically comprises (Brehm et al., 2013). Derived from predominantly youth perspectives, based on user profiles, the resulting conceptual diagram allows us to predict the way the dam will affect this neglected demographic cohort. Based on Figure 2, changes to the river and inundation of riparian land that will accompany the dam, might degrade landscape aesthetics and reduce the likelihood of youth taking road trips to experience the area. Moreover, the sense of home of young residents will also likely be impacted by the flooding of farmland. These insights may not have been considered in the Site C SIA, given the general lack of participation of this cohort in stakeholder processes.

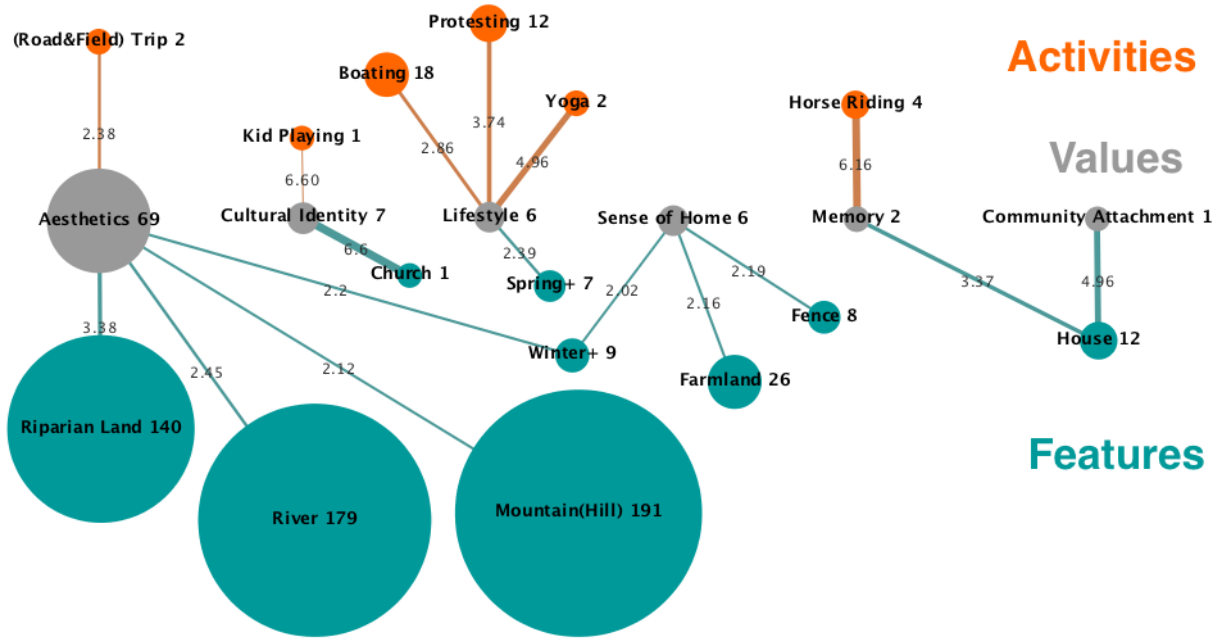

Figure 2. Sample of a conceptual landscape perception map based on statistical relationships between activities, values and features coded from a year of landscape images and captions on *Instagram* from the proposed headpond area of the now-approved Site C dam, Peace River, British Columbia, Canada (Chen, Parkins, in review).

Consider also the case of the Mactaquac Hydroelectric Dam near Fredericton, New Brunswick. In operation since 1968, this dam can supply 12% of the province's power and serves as a transportation bridge connecting two highways. Construction flaws required a decision to be made in 2016 about its future. Options included repairing or removing the facility, each at a cost in the billions of dollars (CAD). An extensive public consultation process recently sought input from across the province on the dam's future. When a year of geocoded landscape *Instagram* posts from its headpond area were coded based on captions and mapped, new insights emerged. For instance, it became evident that the town that was built to hold those displaced from the dam (called Nackawic (Figure 3)) had not accrued the same range of overlapping values in the decades since construction as had the older extant communities, even Meductic with a quarter the population (Chen et al., in press). This finding is consistent to others suggesting that length of association is critical to sense of place (Smaldone et al., 2008, Vorkinn and Riese, 2001), and is useful for post-hoc reflection on the implications of forced relocation (Si, 1993).

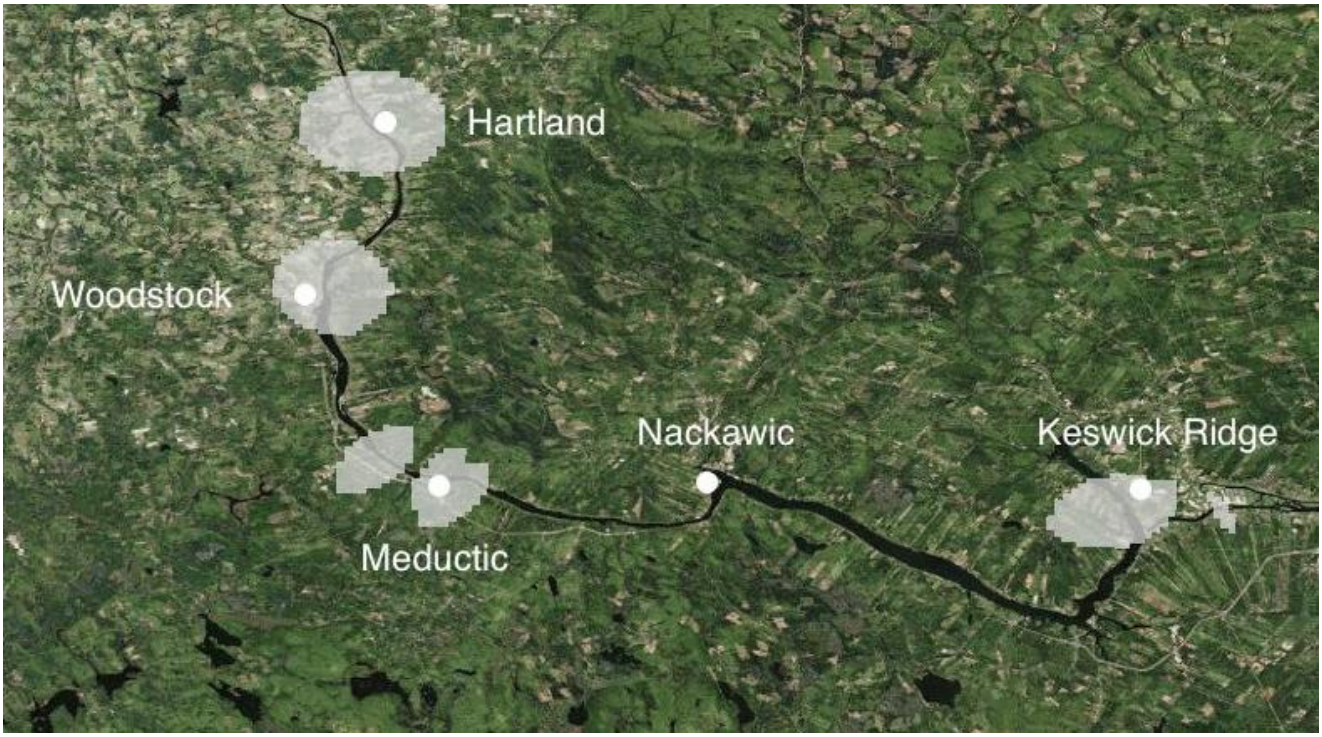

Figure 3. A map showing where all coded values coded to landscape images overlapped for the Mactaquac headpond area, New Brunswick, based on a year of geo-coded Instagram posts. Coded values included aesthetics, sense of home, community attachment, cultural identity, lifestyle, and memory (adapted from Chen, Parkins, in press).

As image producers and publishers, the public creates and curates their own visual expressions of their values, aesthetics and ideology. It is important to note that these are not only *in addition* to what they may communicate through other means, but often, crucially, even in contradiction to what is communicated verbally or textually, and, in particular, what is said in traditional fora like meetings and surveys. Those engaged with the Mactaquac public engagement process consulted social media platforms for text-based insights (National Public Relations and Corporate Research Associates, 2016). They may have found the two posts shown in Figure 4, but neither includes useful text. Yet inferences may be able to be made based on the combination of image and caption. The careful composition of the Instagram image on the left might suggest that the author accepts the dam as part of the landscape and considers it aesthetically pleasing. The Twitter image on the right may suggest the author values the transportation aspect of the dam, as well as seeing it as a worthy backdrop for a "photo-op" roadside stop. Both posts are geo-tagged and included dozens of pieces of metadata that could further inform analysis. This level of inference is achievable using an automated image classifier with sufficient training. These interpretations of images may be unnerving to some readers. We agree that image-by-image analysis and interpretation is highly subjective, but a higher data volume including similar features could aggregate to generate new insight. These insights may reveal consensus as well as diversity in values and experience with the landscape and proposed landscape change.

Digital image repositories in archives and newspapers also offer powerful ways to monitor and understand the impact of past hydroelectric change, for use during recommissioning or recertification processes. For instance, the Oldman Dam was constructed in 1992 near Pincher Creek, AB, to remedy inconsistent rainfall and enable crop diversification, as well as for hydroelectricity and recreation. It was controversial for its forced relocations, violation of First Nations rights and bypassing of environmental impact legislation (Glenn, 1999). Text-based analysis of newspaper coverage of the issue suggests that journalistic interest waned with the controversy (de Loë, 1999). Analysis of local newspaper text and images beyond that which explicitly covers the dam could tell us much more about the place, its people and their changing relationships to local landscapes. A baseline of pre-dam lifestyles and concerns could be compared with those emerging from coverage over the subsequent years. Location and other keywords could reveal spatial discontinuities or cultural inequities in such patterns (Yzaguirre et al., 2015). Such opportunities support the need to build robust and continuous archival data sets to fully realize the monitoring aspect of new SIA tools.

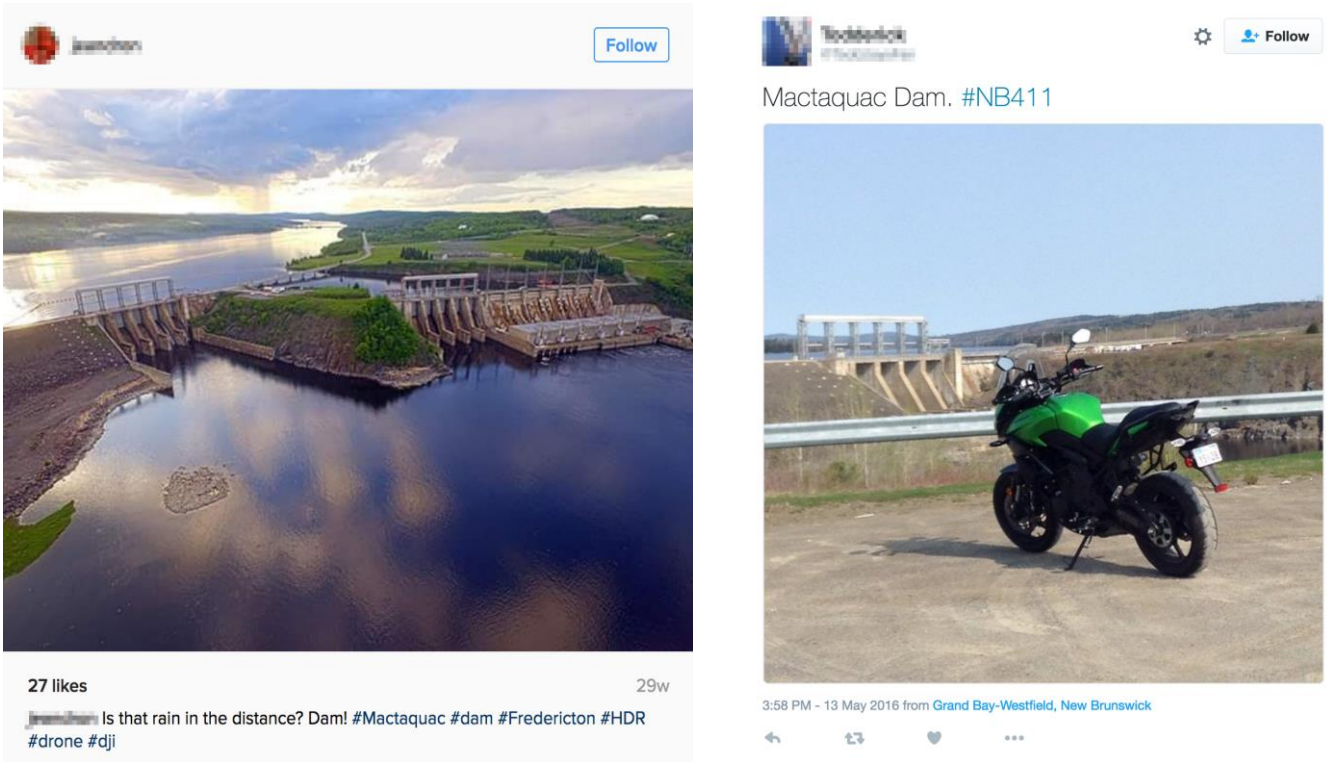

Figure 4. Sample images and captions from *Instagram* (left) and *Twitter* (right) posts about the Mactaquac Dam, New Brunswick, that might confound text-based analysis.

Toward better SIA tools leveraging online images

We call for the development and careful testing of tools that leverage image data types to enrich SIA, as well as advance environmental and social justice objectives via scholarship and professional practice. For instance, beyond the practice of SIA, such tools could be used to identify and visualize environmental racism (Moore et al., 2017, Pearsall and Pierce, 2017, Reed and George, 2011). While Figures 2 and 3 represent a powerful synthesis of hundreds of images in a rural setting, automation will be required to: leverage larger databases; extend to longitudinal analysis; undertake post hoc monitoring; track changing use of the landscape, emerging expressions of cultural practice and their material and symbolic manifestations (cultural memes); and, understand landscape change more generally. Rich data with the capacity to provide insight on values, lifestyles, history and culture reside in the databases of newspapers or private companies that provide social media services, as well as public institutions such as archives. One could imagine an automated tool that generated maps, diagrams or line graphs to reveal how features, activities, and values varied in time or space (building on Google's *Ngram* Viewer (Aiden

and Michel, 2013)). Queries could be driven by location, time period or keyword. Of course, such powerful tools do not come without challenges and cautions (Boyd and Crawford, 2012, Cresswell, 2014, Mah, 2017). Given the nature and number of these challenges and cautions, it is important that those with interests and expertise around society, culture and justice become as central to the development and testing of such tools as technical experts and SIA practitioners.

Some challenges relate to the source datasets themselves, their completeness, diversity and how we use them. The ubiquity of image production and sharing opportunities forces questions about intention, authenticity, and bias (Pink, 2011). Online sources contain systematic biases, as do archives, which house collections donated by or bought from prominent people, and newspaper photography that filters only the 'newsworthy' (and perhaps even decreasing amounts of that, given shrinking budgets at conventional media outlets). Knowing the provenance of images is relevant to identifying and mitigating bias (Bachmann-Medick, 2015), as automated analysis is vulnerable to codifying the biases of data creators (Boyd and Crawford, 2012). For example, power imbalances in decisions of what to digitize and what not to digitize can transfer those imbalances to analytical results (Avery and Holmlund, 2010).

As is the case with most research methods, it will be important that users realize there will be important limits to what any such tools can ever achieve, particularly in relation to disadvantaged populations that hold significant stake in development decisions. For instance, certain points of view (e.g., indigenous populations, religious minorities) and experiences (e.g., rural poor, the elderly) are likely to remain relatively invisible online, either due to imbalances in access, or an understandable desire to protect personal, cultural or local practices (Mah, 2017). Even if these imbalances are remedied, historical records cannot be corrected and longitudinal analysis using such data will never represent these underrepresented voices and experiences.

Concerns about the relative incompleteness of digital media are particularly pressing because the United Nations Declaration of the Rights of Indigenous Peoples (United Nations, 2008) enshrines the legal right of Indigenous groups to withhold consent to development. Justice in such tools will require more systemic justice: Indigenous communities brought on par with others in terms of digital and transportation connectivity, services such as education and health, and press coverage of all of the

above. Meaningful and inclusive SIA is particularly critical in the Canadian context, in light of the recent recommendations of The Truth and Reconciliation Commission of Canada (2015) that seeks to forge new relationships between Indigenous and non-Indigenous Canadians after centuries of injustice and harm.

Other challenges relate to how we might use these tools. Triangulation of insight using a range of data collection and analysis approaches will be critical to achieve balance in SIA or other social justice applications. Any such approaches require careful testing and comparison with more conventional methods of insight generation (as demonstrated by, e.g. Hausmann et al., 2017). Attention must be paid to diversity as well as aggregation. For instance, while an image-enriched SIA tool may theoretically suit such rich Indigenous visual cultures, care must be taken that such a tool is sensitive to different ideas of the sacred and of cultural knowledge and values. Culturomic tools must also not replace direct engagement with people, but rather be used to complement it. Continuing to advance the rich history of SIA that combines both technical as well as participatory methods, as summarized recently by Karami et al. (2017), will be an ongoing challenge given likely pressures to use culturomics as a short-cut. Culturomics cannot replace the richness of interpretation that often comes from close contact through in-depth interviews with research participants. We would not endorse any implementation of culturomics in SIA that seeks to by-pass genuinely participatory engagement.

Other challenges remain technical. Training machine learning algorithms is non-trivial and typically requires labelled data, which has contributed to slow adoption outside of lab environments. Some citizen science projects have inspired smartphone users to generate training databases for large search tasks, by turning image capture and classification tasks into games. For longitudinal analysis to be possible, progress in machine learning must also be matched by the digitization, cataloguing and public access of images held in public archives and newspapers. Such techniques will depend on the widespread availability of, and access to, digital image databases with strong spatial, temporal as well as thematic metadata to aid selection. Ethics and copyright issues also arise when using volunteered images in research that can in part be solved through aggregation (as in Google *Ngram Viewer* (Aiden and Michel, 2013)).

Conclusion

We argue that images offer a unique and underused lens through which to understand human engagement with the environment. Digital image datasets such as those on social media, and housed in newspapers and other archives, could help SIA practitioners and others identify issues, predict responses, inform future proposals or reflect on the impacts of past ones. Thanks to progress in computer science and other fields, culturomic tools are on the horizon to leverage digital image datasets. Such tools hold considerable promise as nonreactive research methods to expand the SIA toolkit. Social scientists, archivists, as well as members and scholars of disadvantaged communities or cultural groups must become centrally involved to ensure that such tools are carefully developed and tested, and appropriately and justly applied. Such tools must never replace direct engagement with stakeholders.

References

Agarwal N, Yiliyasi Y. Information quality challenges in social media. International Conference on Information Quality (ICIQ)2010.

Aiden E, Michel J-B. Uncharted: Big Data as a Lens on Human Culture. New York, NY: Riverhead Books; 2013.

Asselin J, Parkins JR. Comparative Case Study as Social Impact Assessment: Possibilities and Limitations for Anticipating Social Change in the Far North. Social Indicators Research. 2009;94:483.

Avery C, Holmlund M. Better off Forgetting: Essays on Archives, Public Policy, and Collective Memory. Toronto, Ontario,: University of Toronto Press; 2010. p. 242.

Bachmann-Medick D. Cultural Turns: New Orientations in the Study of Culture. Gottingen, Germany: Walter de Gruyter Gmbh & Co; 2015.

Barry SJ. Using Social Media to Discover Public Values, Interests, and Perceptions about Cattle Grazing on Park Lands. Environ Manage. 2013;53:454-64.

Beilin R, Bohnet IC. Culture-production-place and nature: the landscapes of somewhere. Sustainability Science. 2015;10:195-205.

Belting H. An Anthropology of Images: Picture, Medium, Body. Princeton: Princeton University Press; 2011.

Bennett NJ, Roth R. The Conservation Social Sciences: What?, How?, and Why? Vancouver, BC: Canadian Wildlife Federation and Institute for Resources, Environment and Society, University of British Columbia; 2015.

Bennett T, Joyce P. Material Powers: Cultural Studies, History and the Material Turn. Abingdon, UK: Routledgde; 2010.

Blandford A, Taylor D, Smit M. Examining the role of information in the civic engagement of youth. Proceedings of the Association for Information Science and Technology. 2015;52:1-9.

Boyd D, Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society. 2012;15:662-79.

Braat LC, de Groot R. The ecosystem services agenda:bridging the worlds of natural science and economics, conservation and development, and public and private policy. Ecosystem Services. 2012;1:4-15.

Brehm JM, Eisenhauer BW, Stedman RC. Environmental Concern: Examining the Role of Place Meaning and Place Attachment. Society & Natural Resources. 2013;26:522-38.

Brown G, Fagerholm N. Empirical PPGIS/PGIS mapping of ecosystem services: A review and evaluation. Ecosystem Services. 2015;13:119-33.

Chan KMA, Satterfield T, Goldstein J. Rethinking ecosystem services to better address and navigate cultural values. Ecological Economics. 2012;74:8-18.

Checkoway B, Allison T, Montoya C. Youth participation in public policy at the municipal level. Children and Youth Services Review. 2005;27:1149-62.

Chen Y. Public Discourse on Twitter and Instagram: The Mactaquac Dam, New Brunswick, And The Site C Dam Proposal, British Columbia. Halifax, NS, and http://energytransitions.ca/wp-content/uploads/2015/08/Twitter-Instagram-Mactaquac-Site-C-Report.pdf: Dalhousie University; 2015. p. 41.

Chen Y, Parkins JR, Sherren K. Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. Landscape and Urban Planning. in press.

Chen Y, Parkins JR, Sherren K. Understanding young people's perception and use of hydroelectric energy landscapes via social media. Landscape Research. in review.

Cioffi-Revilla C. Computational social science. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;2:259-71.

Connelly NA, Brown TL, Decker DJ. Factors Affecting Response Rates to Natural Resource - Focused Mail Surveys: Empirical Evidence of Declining Rates Over Time. Society & Natural Resources. 2003;16:541-9.

Cresswell T. Déjà vu all over again: Spatial science, quantitative revolutions and the culture of numbers. Dialogues in Human Geography. 2014;4:54-8.

da Silva Vieira R, Antunes P. Using photo-surveys to inform participatory urban planning processes: Lessons from practice. Land Use Policy. 2014;38:497-508.

de Loë RC. Dam the news: Newspapers and the Oldman River Dam project in Alberta. Journal of Environmental Management. 1999;55:219-37.

Delli Carpini MX. Gen.com: Youth, Civic Engagement, and the New Information Environment. Political Communication. 2000;17:341-9.

Di Minin E, Tenkanen H, Toivonen T. Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science. 2015;3.

Diesner J, Kim J, Pak S. Computational impact assessment of social justice documentaries. Journal of Electronic Publishing. 2014;17.

Diesner J, Rezapour R. Social Computing for Impact Assessment of Social Change Projects. In: Agarwal N, Xu K, Osgood N, editors. Social Computing, Behavioral-Cultural Modeling, and Prediction: 8th International Conference, SBP 2015, Washington, DC, USA, March 31-April 3, 2015 Proceedings. Cham: Springer International Publishing; 2015. p. 34-43.

Esteves AM, Franks D, Vanclay F. Social impact assessment: the state of the art. Impact Assessment and Project Appraisal. 2012;30:34-42.

Fang H, Gupta S, Iandola F, Srivastava RK, Deng L, Dollár P, et al. From captions to visual concepts and back. Proc CVPR IEEE. 2015:1473--82.

Gill AM, Reed MG. The reimaging of a Canadian resource town: postproductivism in a North American context. Applied Geographic Studies. 1997;1:129-47.

Ginosar S, Rakelly K, Sachs S, Yin B, Efros AA. A Century of Portraits: A Visual Historical Record of American High School Yearbooks. The IEEE International Conference on Computer Vision (ICCV) Workshops. Santiago, Chile: Computer Vision Location; 2015. p. 1-7.

Glenn J. Once Upon an Oldman: Special Interest Politics and the Oldman River Dam. Vancouver, BC: UBC Press; 1999.

Goodchild MF. Citizens as sensors: the world of volunteered geography. GeoJournal. 2007;69:211-21.

Graham C, Laurier E, O'Brien V, Rouncefield M. New visual technologies: shifting boundaries, shared moments. Visual Studies. 2011;26:87-91.

Halkes P. Aspiring to the Landscape: On Painting and the Subject of Nature. Toronto, Canada: University of Toronto Press; 2006.

Hanna P, Vanclay F, Langdon EJ, Arts J. Conceptualizing social protest and the significance of protest actions to large projects. The Extractive Industries and Society. 2016;3:217-39.

Harper D. Visual Sociology. London and New York: Routledge; 2012.

Hausmann A, Toivonen T, Slotow R, Tenkanen H, Moilanen A, Heikinheimo V, et al. Social Media Data can be used to Understand Tourists´ Preferences for Nature-based Experiences in Protected Areas. Conservation Letters. 2017:n/a-n/a.

Karami S, Karami E, Buys L, Drogemuller R. System dynamic simulation: A new method in social impact assessment (SIA). Environmental Impact Assessment Review. 2017;62:25-34.

Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. Proc CVPR IEEE. 2015:3128--37.

Kim H, Park JW. Topics on aesthetic data visualization: viewpoints, interpretation, and alternative senses. SIGGRAPH Asia 2013 Art Gallery. Hong Kong, Hong Kong: ACM; 2013. p. 1-7.

Kopperoinen L, Luque S, Tenerelli P, Zulian G, Viinikka A. Mapping Cultural Ecosystem Services. In: Burkhard B, Maes J, editors. Mapping Ecosystem Services. Sofia, Bulgaria: Pensoft Publishers; 2017. p. 197-209.

Ladle RJ, Correia RA, Do Y, Joo GJ, Malhado A, Proulx R, et al. Conservation culturomics. Front Ecol Environ. 2016;14:269-75.

Mah A. Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise. Environmental Sociology. 2017;3:122-33.

Martínez-Harms MJ, Balvanera P. Methods for mapping ecosystem service supply: a review. International Journal of Biodiversity Science, Ecosystem Services & Management. 2012;8:17-25.

Michel J-B, Shen YK, Aiden AP, Veres A, Gray MK, Pickett JP, et al. Quantitative Analysis of Culture Using Millions of Digitized Books. Science. 2011;331:176-82.

Moore SA, Roth RE, Rosenfeld H, Nost E, Vincent K, Rafi Arefin M, et al. Undisciplining environmental justice research with visual storytelling. Geoforum. 2017.

Musacchio LR. Key concepts and research priorities for landscape sustainability. Landscape Ecology. 2013;28:995-8.

National Public Relations, Corporate Research Associates. Mactaquac Public Engagement Program: What Was Said - Final Report. Fredericton, NB, and online at http://www.nbpower.com/media/689752/what_was_said_report_mactaquac.pdf: Energie NB Power; 2016.

Oteros-Rozas E, Martín-López B, Fagerholm N, Bieling C, Plieninger T. Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. Ecological Indicators.

Parkins JR, Mitchell RE. Social impact assessment: A review of academic and practitioner perspectives and emerging approaches. In: Hanna KS, editor. Environmental Impact Assessment: Practice and Participation. 3rd ed. Toronto: Oxford University Press; 2016. p. 122-40.

Pearsall H, Pierce J. A spoiled well (of data): addressing the procedural injustice of contemporary environmental justice research through collaborative qualitative data gathering. Local Environment. 2017;22:388-92.

Pink S. Interdisciplinary agendas in visual research: re-situating visual anthropology. Visual Stud. 2003;18:179-92.

Pink S. Amateur photographic practice, collective representation and the constitution of place. Visual Studies. 2011;26:92-101.

Plieninger T, Dijks S, Oteros-Rozas E, Bieling C. Assessing, mapping, and quantifying cultural ecosystem services at community level. Land Use Policy. 2013;33:118-29.

Reed MG, George C. Where in the world is environmental justice? Progress in Human Geography. 2011;35:835-42.

Rezapour R, Diesner J. Classification and Detection of Micro-Level Impact of Issue-Focused Documentary Films based on Reviews. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. Portland, Oregon, USA: ACM; 2017. p. 1419-31.

Rugna JD, Chareyron G, Branchet B. Tourist behavior analysis through geotagged photographies: A method to identify the country of origin. Computational Intelligence and Informatics (CINTI), 2012 IEEE 13th International Symposium on2012. p. 347-51.

Sherren K, Beckley TM, Greenland-Smith S, Comeau L. Path dependency meets uncertainty: How provincial discourses aligned with local preferences against the prospect of dam removal in New Brunswick, Canada. Water Alternatives. in review.

Sherren K, Smit M, Holmlund M, Parkins JR, Chen Y. Conservation culturomics should include images and a wider range of scholars. Frontiers in Ecology and the Environment. 2017;15.

Shove E, Watson M, Hand M, Ingram J. The Design of Everyday Life. Oxford, UK: Berg; 2007.

Si Z. A theoretical framework for social impact analysis with special reference to population relocation at the Mactaquac Dam Project of the Saint John River. Halifax, NS: Dalhousie University; 1993.

Sinclair AJ, Diduck AP. Reconceptualizing public participation in environmental assessment as EA civics. Environmental Impact Assessment Review. 2017;62:174-82.

Smaldone D, Harris C, Sanyal N. The role of time in developing place meanings. Journal of Leisure Research. 2008;40:479.

Stedman R, Connelly N, Heberlein T, Brown T, Decker D, Allred S. The End of the (Research) World as we know It? Understanding and Coping with Declining Response Rates to Mail Surveys International Symposium for Society and Resource Management. Michigan Tech, Houghton, MI2016.

Stepchenkova S, Zhan F. Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. Tourism Management. 2013;36:590–601.

Truth and Reconciliation Commission of Canada. Honouring the Truth, Reconciling for the Future. Winnipeg, MB, and online at www.trc.ca: Truth and Reconciliation Commission of Canada; 2015.

United Nations. United Nations Declaration of the Rights of Indigenous Peoples. Online at http://www.un.org/esa/socdev/unpfii/documents/DRIPS_en.pdf: United Nations; 2008.

van Zanten BT, Van Berkel DB, Meentemeyer RK, Smith JW, Tieskens KF, Verburg PH. Continental-scale quantification of landscape values using social media data. Proceedings of the National Academy of Sciences. 2016;113:12974-9.

Vanclay F. Conceptualising social impacts. Environmental Impact Assessment Review. 2002;22:183-211.

Vanclay F, Esteves AM, Aucamp I, Franks DM. Social Impact Assessment: Guidance for Assessing and Managing the Social Impacts of Projects. Online at http://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf: International Association for Impact Assessment; 2015.

Viégas FB, Wattenberg M. Artistic Data Visualization: Beyond Visual Analytics. In: Schuler D, editor. Online Communities and Social Computing: Second International Conference, OCSC 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 182-91.

Vorkinn M, Riese H. Environmental Concern in a Local Context. Environment and Behavior. 2001;33:249-63.

Walker P, Fortmann L. Whose landscape? A political ecology of the 'exurban' Sierra. cultural geographies. 2003;10:469-91.

Wu J. Landscape sustainability science: ecosystem services and human well-being in changing landscapes. Landscape Ecology. 2013;28:999-1023.

Yzaguirre A, Warren R, Smit M. Detecting environmental disasters in digital news archives. Big Data (Big Data), 2015 IEEE International Conference on2015. p. 2027-35.