Approximately 15 Chairs

Day 1

10:30 am (Session Block A)

Modeling Exhibitions/Events and the relationship between bibliographic descriptions

Attendance: Rob Sanderson, Jonathan Lill (convener), Mary Seem, Sarah Armsler, Rachel Panella, Steven Folsom, Simon Cobb, Joyce Ray, Axel Ermert, Allison Kupietzky, Monika Hagedorn-Saupe, Gwen, Liz Neely

- Genesis of interest at MOMA, objects are well described, with a background in archive often looking at exhibitions as a method of access, and as data itself.
- MOMA.org has surfaced some of this information
- Monika- using EDM at (www.ddb.de), Finding aids are different that Museum treatment, backends can be different (EAD and Museum records) and combined in discovery apps.
- Allison- LIDO adds whole module for events, also have condition reports, implemented in Europeana.
- Liz- considering the depth of the descriptions and blind spots, e.g. when an object is unaccounted for for 50 years, did it result in a sale?
- Steven - how are folks capturing their competency questions around, e.g. transactions
- Liz- get the data linked up with and iterate on the data to support queries.
- Monika- an exhibition can result in higher market prices.
- John - each MOMAexhibition has a URL, lots of interactions are driven either by exhibitions or artists. Objects relationships to exhibitions are way more difficult
- John- museums community hubs, interacting with galleries and community, documenting this
- Monika- Curators have tremendous power,
- Mary- MOMA has people working to make sure that web archiving is keeping the websites functional
- Mary- exhibit brochures/catalog can record information about events, need a BF extension, not note oriented
- Rob - provenance index is tracked at Getty, have conflated exhibitions and auctions
- Monika- exhibition checklists are critical forms of documentation
- John- challenge to turn this information into data because we can't always know for sure that this particular thing is the thing described. Untitles, or even when Numbered and Titled
- Monika- if present, registration numbers can help with this

- Rob- Exhibitions and provenance makes sense to make available as LOD.
- Rob- Exhibition modeling is current focus of a linked.art, want the data to be useful and interesting to surface
- There was a discussion last week about what is an exhibitions, and the relationship with Program/Series, exhibitions, venues.
- John- the difference with other events
- Monika- they have published their exhibition lists, travel exhibitions can be a revenue stream. Used AI to look at pictures of events to find people
- Should we share this information?
- John - MOMA tries to be complete so that they aren't seen as creating a select documentation of their history
- Joan- sometimes there needs to be exemptions
- Monika- arrangement could be an important
- John- traveling shows can show the diffusion of modern art.
- Rob- events and activity based models allow to describe who did want, when, where
- John- Seurat painting was at MOMA at the time of a fire
- Liz- so do you have a URI for the fire as an event? With the object linked to it?
- John - we could, but don't now
- Simon- wouldn't need to link to an event if the event took place at the same time and place at the same time and place location of an object
- Monika- can we create shared exhibition lists?
    - There seemed to be interest.

11:30 am (Session Block B)

Machine Actionable Profiles - Capturing data targets explicit

Attendance: Steven Folsom (convener), David Newbury, Lynette Rayle, Edward Anderson (Rijksmuseum), Michelle Futornick, Ted Lawless, Eric De Ruijter, Beat Estermann, Mary Beth Weber, Eero Hyvönen, May Chan, Amye McCarther, Kevin Ford, Ethan Gruber, Stefano, Kenneth , Kelli Babcock

Facilitator: Steven Folsom, Cornell

Intros

- Steven Folsom: interested in application profiles esp as applied to form configuration in a cataloging tool; and at community level with Program for Cooperative Cataloging where traditionally decisions are made with interoperability as a goal. So far used spreadsheets

to record modelling decisions; would like to move in the direction of more sophisticated ways to do this, e.g. also enabling automatic validation, etc.

- Eric De Ruijter, international institute of social history, netherlands. Strategic interest in linked data. Converted all collection and research data to linked data (ETL), looking at future to create from scratch, with want to edit
- Stefano Cossu, Getty, IIIF Profiles
- Michelle Futornick- Stanford, working on creating profiles for forms, but only understood by one application
- Amye McCarther, archivist, New Museum, interested in interoperability with other museums, interest in Wikibase
- Ted Lawless, JSTOR Labs, interested in form creation and data validation
- Kelli Babcock, U of Toronto digital initiatives, interested in IIIF + Wikidata
- Beat Estermann – Bern University of Applied Sciences, ingesting data into Wikidata (on performing arts and on heritage institutions), wants to make sure the data is consistent and commit to supporting specific queries
- Mary Beth Weber, Rutgers, on a project with application profiles that are labor intensive and not public, looking to implement Exploro (a research profile system)
- Eero Hyvönen – University of Helsinki, applications on top of linked data like Sampo
- Edward Anderson, Rijksmuseum, Netherlands, harmonization data layer for siloed systems
- David Newbury, Getty enterprise sw architect, vocabularies and provenance index work with a lot of human collaborators, how can data profiles help manage this collaboration, limit to scaling the human contribution
- John McQuaid, Frick
- May Chan, U of Toronto, manages traditional metadata services group; relationship between local practice and community decisions
- Lynette Rayle, Cornell, developer on vocabularies but may move into profiles
- Kevin Ford, Library of Congress, id.loc.gov, BIBFRAME activities. Ambivalent about application profiles, why so little success in implementation?
- Ethan Gruber, American Numismatic Society, builds user interfaces and web forms for creating, editing and publishing linked data. All tailored to their own data model, wants to generalize the work, use SHACL
- Kenneth Seals-Nutt, Yale / Science Stories, standards for validation

Community level considerations

- David, projects that accept data from external contributors. The Getty Vocabularies do have an XML format for contributors but only 1% of people use it. How to manage all the data contributions. Challenge: if too strictly specified, people won't follow, if too loosely specified, machines can't use. Getty publishes documentation on how to create data according to their specifications (precisely defined XML schema) but people don't follow it.
- Stefano: isn't this a reconciliation problem? Put work toward making data better from the contributor side? Or once data is received?

Feasibility of application profiles

- Steven: ideally could have a looser, light-threshhold profile and then build on that depending on local needs. All could agree on some core, baseline level
- Kevin: why do we have application profiles? Situation where standard is so flexible (Dublin Core, METS) that there are so many options for how to implement. The profiles attempt to give guidance but invariably become tied to the whoever created them, so they are not as interoperable as hoped. Re MARC: the strength of it is the rules. The MARC rules are an application profile, expressed as rules.
- May: concept of level 1, level 2, level 3 cataloging: could you layer profiles in a similar way? Could you add additional profiles later without having to redo anything
- Steven: ShEx and SHACL could support this, profiles relating to each other
- Stefano: how do you notify users about how compliant some chunk of data is?
- Steven: you can't save data in a system til it's compliant, reports on data validation
- Stefano: on data consumer end? How do I know if the data is compliant?
- David: application profiles traditionally used to say "I have this data and it can be used across systems", this hasn't worked as well because of the needs of the different systems. What would be better is to be able to say: "would your data work well in my system"?

Who is using profiles?

- Mary Beth: profiles used for validation, only internally, tells researchers what they can and can't do (?) *didn't hear this part*
- Stefano: IIIF API gives you compliance level
- David: hasn't seen any IIIF client ask for that info and adjust its behavior according to the answer

Kevin: defining what's possible vs what should and shouldn't be. Debate in BIBFRAME work over what does required mean? "Required" means cataloger should supply? Or that data should have? Successful profiles were successful because applications were designed specifically to use them.

Steven: 2 workflows: converting MARC to RDF vs native creation of RDF, can application profiles help reconcile these

May: different mindset. Creating metadata mindfully for shareability in future, not always mindset

David: back to question of why people don't produce compliant data, not enough incentive. IIIF gives you red / green indication of whether data was successful

Steven: how can profiles be reused if they are in different specifications, how handed off among tools?
Kevin: why would you choose one application profile spec over another?

Steven: majority of use cases are supported by either ShEx or SHACL, different communities have adopted one or the other
David: SPARQL people like SHACL; JSON schema people like ShEx. ShEx is simpler. SHACL is very powerful.

Steven: converting among application profiles
David: can't imagine wanting someone else's application profile because my application has different needs (application writer point of view). From data provider point of view would be nice to know whether my data will work on various tools, a way to check
Beat: which side would the checking happen on? Data provider provides to a system and gets response back of how well the data complies, is this enough? Or is there a need for the data provider to be able to validate the data on their own side?
David: the latter would be lovely but who would be responsible?
Beat: could the feedback be given directly to the people creating the data at the point of creation?
David: local needs, doesn't use a majority of the data, need flexibility to use only part of the data

How do people communicate that you have a profile that exists?

- Beat: standards organization that publishes profiles; system that checks ingested data against profile and lets you know

Kevin: what is motivation for wanting to know if your data is acceptable to a particular system? Whose responsibility is it to make the data conform?
Beat: application should provide information about what data is required and why, what won't work if certain data is supplied? Then data provider is responsible
David: fixing contributed data is unsustainable
Beat: data fixed on Wikidata, what happens to the data in the originating system?

-
-

2:00 pm (Session Block C)

Using Who's on First Gazetteer to model collections

Day 2

11:00 am (Session Block D)

1:30pm (Session Block E)

2:30 pm (Session Block F)

Pragmatic approaches to linked data

Proposer/facilitator: Elizabeth Roke / Emory

For institutions with fewer resources to devote to linked data, how to get started

Elizabeth: One outcome from this session: take “useful resources” doc started at this summit and share it more broadly
Liz (Georgia O’Keefe Museum): American Art Collaborative experience showed how much work is required. Payoff wasn’t clear until the registrar actually saw the results of the work. Visual impact. In this group we understand the benefits conceptually but “seeing is believing”.
Duane: adds that the registrar had been working on it for 2 years and conceptually understood the “why” but the visceral reaction didn’t come until she saw the visual (a skirt in the collection connected to a photo of Georgia O’Keefe wearing the skirt, connection between things in the museum, revealing more about the collection to the public)
Kelli: user stories would sell the work to the institution
Aaron: for novices and even for experienced practitioners, the syntax and notation is awful; simplified notation goes a long way. Example: flickr’s machine tagging project, it was essentially triples but explained it in a simpler way, uptake was huge
Elizabeth: visuals for people to get it. Wikidata for example makes it easy to get data out for simple yet impactful queries like “whose birthday is today?”
Karen: Wikidata visualizations are useful, display in timeline or map; Wikidata visualizations as error detection / uncovering outliers would be appealing as use case
Etienne: Does anyone remember InMagic(?). Text files, but essentially a triple format. User friendly because it’s readable, unlike e.g. RDF/XML
Other low barrier tools for producing / publishing linked data?
Dennis: Arches: doesn’t ask people to learn what linked data is, you can supply csv and it turns it into linked data. Showing people outcomes is what we should focus on, what are you hoping to accomplish with linked data?
Elizabeth: push vendors to create serializations …

Duane: American Art Collaborative: data working group created baseline spreadsheet of what fields were needed, institutions figured out easiest way to get the data out of their own systems
Kelli: do we have a responsibility to create pragmatic solutions? Context: institutions want to adopt IIIF but didn't have IT staff to implement it, no infrastructure to support
Elizabeth: if it's only the well-resourced institutions who can produce linked data, and things are built on top of this data, it's a biased view of the world
Aaron: still a high barrier to use the tooling we have produced
Karen: unreasonable to expect every institution to gear up on their own to do linked data, better to do it on organizational, regional, membership basis. Exercise your purchasing power with vendors, even if you're small
David: hard to talk about linked data as a monolith, you can ask for "virtual reality" or "linked data" but it's a big concept
Florian: Aspects immediately usable and convincing: use a Wikidata identifier for places, easier than coordinates. Doesn't require full-fledged linked data system.
Dennis: IT people buy-in gives foothold.
Aaron: Open Street Map as pragmatic example
Karen: some digital collection websites list separate sites (often separately funded, discrete projects, managed by different staff) and no way to manage or search across each
Elizabeth: we try to model everything
Aaron: we are reluctant to start small and to start imperfectly
Simon: working on small projects doesn't show real benefit of linked data
David: need a reason for RDF; if you have links and URIs you're most of the way there. The full linked data is not worth it unless you commit to AI and reasoning
Elizabeth: from archives perspective, RDF allows capture of knowledge in a way that you can't do now
David: why describe it if you're not going to work with the description / do some computation on it
Karen: what is minimum viable linked data?