CLARIAH Video Annotation Interoperability Expert Meeting

Collaborative notes

Day 1, Thursday - Morning session

- Marijn introduces the day, motivation: many different tools for different users and uses. Since scholars work across media types, the idea is to increase interoperability of these tools. For CLARIAH annotation is a top priority in the next five years (CLARIAH Plus)
- Idea of having three meetings, first one identifying elements of the different models, aligning with IIIF framework.
- Introductions:
    - Marijn, expertise on IR
    - Aristotelis, human-computer interaction and music, interfaces for musical expression
    - Heizo, center for music and computing, professor for audiovisual media, film analysis education, protocols. Movie Pulse (app), formal analyzes (in Excel)
    - Hugo Huurdeman, information science and human-computer interaction. XIMPEL.net project, for interactive video.
    - Joscha Jaeger, web video, annotation, metadata formats for describing videos, IIIF expert, part of the working group. Developer of Frame Trail.
    - Jaap Blom, CLARIAH Media suite main developer, web-based annotation, experience with IIIF
    - Jeanine Evers, chair of the REFI initiative, cultural anthropologist, expert on the QDA commercial tools.
    - Liliana
    - Gaudenz, developer of VIAN,
    - John Bell, Media Ecology project,
    - Hans Slootjes,
- Marijn's presentation:
    - Concept of annotation, something to discuss in this meeting
    - Differences between desktop annotation tools and browser-based
    - Interoperability challenges
        - Different understanding of what counts as annotation
        - Import/export formats
        - Desktop vs. Browser-based tools
        - Offline vs. online AV materials

        - Tier-based vs. standoff annotation
        - Manual vs. automatic annotation
    - Desktop tools
        - ELAN, VIAN
    - Browser-based tools
        - FrameTrail, Waldorf.js, CLARIAH Media Suite
- Different tools are good at different tasks. We try to support tasks for video annotation, this is the focus
- Our goal is to make mappings between the tools and their semantics (crosswalks), how to align
- W3C, defined as recommendation in 2017, meant to standardize all annotations on the web (images, text, …). Every element added in semantic web terms. Drawbacks: modelling of tiers,
- Concept of fragment (Web annotation recommends “media fragments”, but this has a lot of limitations, especially when encoding video in film studies), e.g
    - it’s hard to represent/select polygons
    - how to represent aligned sources, e.g., textual sources
- There can be multiple targets within one annotation (we will run into these issues, elements in the model for which there is no definition in the W3C model), we need to find out these
- Hyphothesis: not focusing on video annotation at the moment
- John Bell uses web annotation as a way to exchange data between the different applications in the Media Ecology Project

- ELAN
    - See his presentation for general description of the tool and user communities
    - Supports multiple tiers:
        - Tier types with the possibility of attaching a template with certain properties
    - Difference between transcribing (textual representation) and coding (limited set of categories)
    - Concept of “tier” is not clear for qualitative expert (Evers), she argues this is also coding, we conclude that tiers are a high-level category (e.g., in Nvivo) or a family of codes (in Atlas.ti). Han comments tiers are not the codes, Jeanine insists they are also codes
    - Time slots contain annotations, so it’s possible to put multiple annotations in a time slot (different from WA). Annotations have a timeslot ID
    - Vocabularies/codes are either contained within each document or referenced externally
    - Annotations on annotations are possible
    - How to translate the linguistic types to W3C will be a challenge (important part of ELAN)

- Han is not familiar with W3C, his first experiment converting does not make much sense (first try)
- Challenges for the conversion from ELAN/EAF to Web annotation data model:
  - Encoding of tier names, type, typology
  - Controlled vocabularies
  - Losing information after exporting/importing
  - Lexicon links, references to external resources

  When will interoperability based on the Web annotation data model be considered a success? How much data loss is still acceptable?
- What round trips would make sense?
  (ask Han, he described a couple of them…)

- VIAN
  - Intention to have a web site where all the annotations are presented, and that people can add their own projects, crowdsourcing support, and semi-automatic annotation
  - Tools by Pajarola group, written in Python, embeds VLC
  - VIAN projects are serialized as JSON files, a lot of application specific info
  - Numeric and binary data is serialized in to a SQLite database (initially, scholars used FileMaker for annotating/coding the fragments made in ELAN)
  - The format/model was not made with exchange in mind
  - Types of annotations
    - Temporal segments (similar as in ELAN, with tiers -visual, audio- but no hierarchies in tiers)
      - Start + end
      - Annotation body (text, so TextualBody in W3C WA)
    - Visual annotations (they also have duration and can be automated: move over time), they have start and end, type (rectangle, ellipse, text, image, free hand), they are grouped in the so-called “Annotation layer)
      - Vector graphics, shown as box within a shot/frame
      - Can be animated (using e.g. tracking algorithms)
    - Screenshots
    - Keywords/tags (happens in different GUI/system)
    - Linking to external objects (e.g. PDFs)
      - Drag-and-drop on segment, files are copied to project dir
  - Data model (shown as UML diagram)
    - Project can have multiple experiments
    - Classification objects is part of experiment
    - Experiment = enrichment

- Supports all kinds of shapes (& “free hand”). Not yet a standardised way to express them (currently vectors)
- Screenshot is also a type of annotation = markers. It has a time code + a title (+ possibly an image mask?)
- Different vocabularies available , each vocabulary has a distinct concept (e.g., a vocabulary for background), these are called “classification objects” for the main entities being analyzed
- Question: to what extent do the classification objects resemble tiers
  - Answer: can have different classifications on the same segmentation (so, sort of yes, sort of no)
- Question by Niels: are these vocabularies published? E.g., see project Kinopoetics
  - Answer: Not yet, but will be
- Questions: are terms in vocabulary formally defined? E.g. how is “blue” defined?
  - Answer: for VIAN it’s a set of words, but they refer to terms in DB with proper definitions
- Question (Heizo): how do you give access to research results?
  - Answer: the projects can be visualized in the latest stage, e.g., as a movie with data (there will be a web app), the movie is not shown (cinemetrics style says Heizo)
- External media objects: can be attached to a Segment or a Visual Annotation by drag-and=drop a file on it. Files are copied into the project directory
- “An annotation is a kind of overlay on top of the movie”

- **FEI: A format for scholarly film analysis**
  - Background: MEI music encoding initiative, became the de-facto standard for musicology and musical digital editions
  - Based on projects, MEI file appears, with annotation software
  - MEI affords tool pipelines
  - Option to annotate scans
  - Three aspects: metadata, score / symbolic representation, and annotated scans
  - There is also a presentation software for this
  - Heizo: motivation that he (media studies) may have:
    - Heizo analyzed more than 300 films in Excel sheets (motifs)
    - Aim to combine findings from: Cinemetrics, Barbara’s work, Movie-pulse.com (his work). Also, how to preserve research data
    - Need for a standard for: preservation, combination, best practice, tool chain
    - Standards for film description?
      - Shot length (american standard vs European standard)

                - meta(data) for editions and inside metadata of the shots
            - Currently in media studies data is shared in pdfs, via protocols, and excel tables
    - Question: is web annotation the way to go?
        - Answer (John Bell): usually this goes on top
        - Interoperability issues are in layers
        - Namespaces as part of the interoperability problem

- **Waldorf.js**
    - Waldorf (frontend, jQuery plugin) and statler (Ruby on rails, backend)
    - For ease of use: plugin in HTML page, it wrap HTML5 video tags into annotation interface, config allows specifying which Statler server to communicate with
    - Intended to be a component of a series of other components
    - Use cases/projects
        - Machine vision search, automatic identification of objects and actions, feedback loop: users agreeing or not on annotation (machine learning)
        - Three monks
            - Call backs built in (java script) advantage of W3C annotation model here
        - Workflow: Waldorf uses videos from Internet Archive, vocabularies from onomy.org, Statler feeds back into Internet Archive
    - Features
        - Time- and geometry-delimited annotations
        - Suggestion of tags loaded from external identifiers
        - Everything identified by URIs
    - FOAF for creators
    - Time codes: decimal seconds from the beginning of the clip (problematic since browsers begin time…)
    - WA issue: marking machine-generated tag as incorrect
        - solution: overload “tag” semantics by serializing more data into it
        - Body is text, can parse the body of the annotation (putting all there easier instead of rewriting the application)

- **Media Suite**
    - W3C compliant
    - Motivation: only bookmarking (no others so far)
    - IIIF meant to be for interoperability
    - You add multiple bodies to an annotation (comment, classification, link, information card)
        - Information card: set of key/value
        - Template per use-case/domain

    - The target describes the identity of the annotated resource
        - Source
        - Asset ID
        - Type
        - Selector: custom (non-W3C), it allows to query later annotations, give me all annotations that are related to media objects from this collection, he added nested PID selector
    - Resource= television program, movie… this abstract resource can have multiple representations or media objects (e.g., tracks), and these things can be segmented
    - Jaap added the Asset ID: should be the persistent ID to find the content across tools
    - Body
        - Contains the actual contents of what you are annotating
        - Multiple bodies per annotation
    - Bookmark group = Just an annotation
        - An annotation with one body, which has a classification (that is the body)
        - Multiple bodies in an annotation can have different purposes (different to motivation)
    - In the MS you can overlap annotations, a single tier in the MS
    - Good example for tier challenge: Chronoviz
    - Persistent identifiers are very important for interoperability and playback (but many online resources have no persistent identifiers and esp. problematic for dealing with offline materials in Web Annotation)

- **FrameTrail**
    - Web application to experience, manage, edit interactive video (in the browser)
    - Basis: video file / time span
    - Add / edit:
        - Overlays: are part of the video
        - Annotations (documents/ comments) are supplementary information
        - Code snippets, events & actions
    - File based, open format
    - Use cases
        - Journalism
        - Online learning
        - Interactive storytelling
        - Scientific publishing
        - Parliamentary debates
        - Psychology learning

- Flat data model, only annotations
- Tiers are simulated at the interface level, overlapping
- Link to documents, it is an annotation, but the body includes where to find the document
- The W3C data model is good, we can use it and adapt it
- But there is no standard for hypervideo (used W3C and added custom fields)
- Why not IIIF Presentation API?
    - File-based data storage hard to manage (when doing everything file-based
- Abstract timing concept of IIIF canvas, he used this in a player for BFI
- Video files are synchronized, with only a IIIF Json manifest
- The annotation part is solved, the agreements are missing (types to share, to get into a spec), but on a technical level is solved, problems are for expressing Collections (pieces, e.g., a video in 3 CDs); management of the information around the contents, how to present it. This is where IIIF comes very handy
- IIIF AV is handy for contextualisation of annotations

Day 1, Thursday - Afternoon session

REFI

- Qualitative Data Analysis software interoperability
- Exchange format co-developed with the developers/maintainers of the various (commercial) QDA software packages: Atlas.ti, Nvivo, Transana, Quirkos, QDA miner, Dedose,
    - Round tripping is not an aim
- Coordination group (standards expert, power users)
- 8 software packages, highly international group
    - All desktop-based
    - Projects are local files (resources, annotations)
- Code exchange: released March 2018
    - *Code* can refer to other code
    - *Codes* can be grouped (set)
- Project exchange: planned release March 2019
    - Features unique to single tool is not useful for exchange (although it is for round trips)
    - Some features were dropped from the XML schema, e.g. multi-selection (from multiple synchronized resources)
    - *GUID* can be used for round tripping, e.g. export only shared-feature elements, compare imported project GUID with local project GUID and merge
    - *Coding* connects code to resource
    - *Variables* (text, boolean, integer, ...)
    - For textual resources, use only plain text, as with rich format it's unclear what selection should include about layout information
- Issues
    - Terminology: e.g. code is feature that enables you to label
    - Exchange is losing things: specificity in own tool and in other tools
    - Organizing information (documents, communication channels)
    - Organizing interactions between participants (meetings, conf. calls)
    - Live meetings are extremely effective and essential

General issues:

- What should be expressed as annotations and what should not?
- Selectors (temporal fragments, spatial fragments, polygons, …)
- Tiers: implicit semantics in tiers
- Resource identifiers: offline and online, persistence, resolvability
- Namespaces: defining, document, sharing and aligning

- Exports are all related to a project (ID); each tool importing will use this ID as well. In the case of a roundtrip, a tool simply can merge the data based on the project ID.

Interoperability challenges:

- Different understanding of what counts as annotation
- Import/export formats
- Desktop vs. Browser-based tools
- Offline vs. online AV materials
- Tier-based vs. standoff annotation
- Manual vs. automatic annotation

Question (Han): in our endeavour we look for an application profile, as a subset of the standard. Marijn says this doesn't apply since W3C is already so generic.

- There was a previous initiative for an exchange format (Qdex), but this was top down, not adopted (only by Atlas.ti back then)

DISCUSSION

Aspects that have to be discussed for interoperability (based on W3C) which are dealt with differently between the tools:

1) Selector/target: what resource we are referring to, selector=temporal segments, espacial segments, ...
2) Project/collection/hermeneutic unit (context):
    a) Collection: an annotation has a dialogical relation with other annotations
        i) Marijn: a way for a user to organize all his annotations, a collection has an identifier which groups all the identifiers of the annotations.
        ii) It's very close to the concept of tier, since it groups annotations, while project is everything that is not annotation
        iii) Marijn likes the IIIF solution of the manifest (Joscha asks if this is where the project information should be stored)

    Tier:

    b) Jaap says this is grouping within a target, multiple target vs one target with multiple dimensions
3) Provenance/context
    a) Where, by whom, how annotations were made
    b) John says it's important to consider the level of granularity of this provenance
4) Body/motivation:
    a) How the body looks like/contains: code, string, URI to an external term from a vocabulary

GROUP 1. Selector/target

- Getting familiar with the model
- What are the classes
    - Class is meant in an RDF style, a hierarchical way of modelling things
    - In RDF we create the class of a body and the class of a target and the class of an annotation
    - Concept of Type in the model
- Issue of how to represent fragments in the URI
- Two problems:
    - Limitations of media fragments model
    - Inclusion of other selectors in combination with media-fragments
- We should start by defining the different modalities we want to refer to
    - Time interval
    - Frame
    - Regions in a frame
    - Sound/Music
- The selector approach has a very technical way to approach modality
- Synchronization between two media fragments (same video, different time codes): this is done with the body, and with an annotation of the annotation.
    - Claim of synchronicity: it creates abstract units, not at the structural level
    - Then clarification of what this correspondence actually means: this is another annotation
- What to encode in the URI? How to render it/process it? There are extensions to the model in this aspect in Github, but the media fragments is not active (2012 last update)
- Use case: transform FEI to Web Annotation, inform the development of FEI by Web Annotation, e.g. model scenes with “isScene” predicate
- Distinction between form and content in computer science kind of works since there are media types, but this is artificial
- Structural elements (semantic): credits,.... This is a different problem, how to bring this information into the annotation. What we deal with here is
- Limitation: moving polygons are too big for IRIs
- Semantics should not be put in the IDs (e.g., chapter 1)
- What are the means we have to communicate the semantics in the W3C model?
- One of our main recommendation is not to extend the model, this is not feasible, but for this and this situation, **generate/extract a new resource** that you can use with the selector (e.g. of the polygon), and relate the two resources (something like animated svg), something SMILE attempted long ago, and now IIIF
- Another approach would be to use the selector URI to refer to a file describing the selection without the limitations of an URI
- Summary options, Day 2:

- In case of a very complicated selector (e.g. describing a moving object/region over time, i.e. a “tween”) it could be hosted & referenced via URI. This can be especially useful in case this same selector is referenced in a lot of annotations.
-

GROUP 2. Project/collection/hermeneutic unit (context)/ TIER

Tool	Collections	Creator/Person	Annotations	Relations
Waldorf.js	N/A	FOAF	W3C	N/A
ELAN	Document, Tiers	Annotator, Participant	Text, URL	Linked Annotations
VIAN	Projects Segmentations Annotation Layers Screenshot Groups	Contributor	Text, Documents, Geometry,	Linked Annotations
CLARIAH	Projects, Datasets	User	Text URL Key/Value Obj	N/A
FEI	N/A	List of Users	Text, URL, Images, Values	N/A

Collection Example ELAN Document:
- Collection Node for Document
- Collection Node for each Tier
    - Separate Node to represent Tier Hierarchy

Summary Day 2:

- There is no mechanism for specifying the relations between individual annotations (dialogic structure as Niels referred to it)
- Difficult to agree on terminology: what is a project? How is it related to tier of collection and should it be part of an annotation exchange format?
- Marijn suggests concept of project as everything that is not an annotation. Is there a way to make a clean separation between what is annotation and what is not?

GROUP 3. Provenance/Context

GROUP 4. Body/Motivation

- Use case: segmenting and coding Desmet film
- Discussing body
    - Motivation in IIIF vs Web Annotation M
    - Perspective of application (how to show it) vs. user motivation
- Tools:
    - CLARIAH Media Suite
    - FrameTrail
        - Looking at their annotation structure
        - Hard to do a direct crosswalk due to differences in structure
- Goal: map bodies
- ToDo:
    - Compare the different annotation types per tool and see how they align to motivations

Day 2, Friday - Morning session

Input from Joscha (email):

- Regarding the crosswalks, I'm pretty sure after yesterday that Web Annotation alone is not sufficient as an exchange format. In order to reflect hierarchical structures like scenes or tiers, something like a IIIF manifest is needed. That way the annotation can exist in a "flat" Form and it's up to the client / application to respect / reconstruct the structure (which would reside in the structures - > ranges properties of the manifest).
- Re: IIIF: in case a decision is made towards IIIF, it might make sense to try and invite Tom Crane from Digirati for the follow-up meeting.
- A really good read for people new to IIIF and it's relation to existing models: https://resources.digirati.com/iiif/an-introduction-to-iiif/wheres-my-model.html

Discussion

- See notes above (in each group)
- Exchanging textual annotations should be straightforward
- For classifications the vocabulary info might be problematic
- For the WA-based tools, what fields do they add beyond the WA spec?

Practical exercise

The idea is to compare, per tool, how each of them will deal with the following cases when exporting/importing: translate this to your particular tool and keep track of the issues that arise. Do this in three parts:

(1) Minimal Example: Comment/text exchange
(2) Slight more complex: Classification
(3) Even more complex: Tier-based

What fields do WA-based tools add beyond the WA spec

Basic format:


```
[
      {
      "@context": "http://www.w3.org/ns/anno.jsonld",
      "id": 26,
      "type": "Annotation",
      "motivation": "highlighting",
      "creator": {
            "type": "Person",
            "nickname": "John Bell",
            "email": "2bb2a925eb5ac9fd97fc4c041fabc050f194392d"
      },
      "body": [
            {
                  "type": "TextualBody",
                  "value": "This clip was filmed at the Parker
Center...",
                  "format": "text/plain",
                  "language": "en",
                  "purpose": "describing"
            }
      ],
      "target": {
            "id":
"http://mediaecology.dartmouth.edu/other/KTLAEulaLove.mp4",
            "type": "Video",
            "selector": [
                  {
```

```
                        "type": "FragmentSelector",
                        "conformsTo":
"http://www.w3.org/TR/media-frags/",
                        "value": x"t=12.0,19.0"
                    }
            ]
      }
      }
]
```

Namespaces, properties

Waldorf:

- SVGSelector (which is allowed but not part of mediafragment?)
- Future animated target selector format to be named
- FOAF creator details (name, hashed email)
- Missing: Collections not currently supported but could be added or migrated to tags
- Missing: id for tags is currently optional, should be required?

FrameTrail:

- Joscha uses a frametrail namespace (not documented yet) with application specific types etc.

CLARIAH Media Suite:

- NestedPIDSelector: this is mainly used to track the hierarchical structure of resources, e.g. a film at the level of *work*, at the level of a specific *version* and at the level of a specific video stream. Making an annotation on a segment of the video stream is also an annotation on the film as a work. This helps users to organize annotations at different levels.
- Vocabulary: this is used in annotation bodies that have a *classification* purpose, and identifies the vocabulary from which a classification label is used.
- Key/value pairs in body for metadata card

Day 2, Friday - Afternoon session

Updates from work in pairs/tools:

- WA-ELAN
    - This repeats a bit what is in the previous file

        - Option to consider each motivation as a tier (very complicated, it's an option but not the best idea)
        - Selecting default tiers (guidelines are required here)
        - Images as targets: would show up in ELAN, but it doesn't make sense, the tool will create a virtual timeline.
        - Images as bodies: this doesn't make sense in ELAN, only text as body
        - Annotations can be at the resource level, not having a selector, but this doesn't make sense in ELAN
        - Motivation/property tiers can be mapped back
        - More interesting stuff: tiers called speaker…
            - Either use annotation collection that correspond to the name of the tier
            - Two bodies within one annotation

- **FEI<->WA**
    - WA requires time ranges, even for specifying a specific point in time
    - Also issue of associating different annotations to the same segment
    - WA is a graph model, FEI is tree model. Difficult to map/align graph models and hierarchical models
    - To express is XML-based in WA, this would be the most common scenario
    - Samvera: http://samvera.org/ (can't remember exact point in which this was mentioned)
    - A previous solution for layered things, say Marijn, was to use annotations referring to annotations
    - Linguistic Annotation Framework (LAF)
    - Need to define what in terms of structure we/tools want to support
    - In musicology, they always end up with an MEI format.
    - What is common use case and interoperability need (we revisited this issue when talking about what levels of complexity in structure we should consider, but decided to focus first on the simple cases listed in this exercise to see what can be offered first).

- VIAN-Waldorf
    - Export from VIAN to WA
        - Issue: identifying vocabulary from which term is chosen, use *id* property for URL to vocabulary with an anchor to the specific term.
        - Tiers in VIAN were used for different types of segmentations (color, music…)
        - We came up with a solution (showed already by Gaudenz/John) for the key-value problem, sell it to W3C
    - Tier issue
        - *AnnotationCollection* with property *label* and value "Segmentation tier"

- CLARIAH Annotation Tool <-> Waldorf.js
    - Property *text* of body type *commenting* is not part of WA spec
    - Metadata card has dictionary format, could be new body type or *type*:*Dataset*
    -
    -

General comments (see also in each file above)

- Part of the application profile we want to end up with is to agree on the format and also in what you don't use, and what you use as extras.
- Introducing RDFs is necessary since this is not inherent in the W3C model, for talking about relations between the distinct components of an annotation
- That is also why is needed to use namespaces
- What to do with bodies referring to external resources that are stored locally?

CONCLUSION

- Wish to wrap up this meeting, make a summary and document recommendations
- ToDos
    - Finish the tables, see if it's possible to combine all done today into a single table
    - Test cases: document the issues
        - Simple annotation as it would come from Waldorf
        - Annotation with classification
        - Non-dependent tiers
        - Hierarchical-dependent tiers
        - Vocabulary exchange
            - using dataset type in the body
        - Refined selectors
    - How do the test cases should look like
        - Come up with a format to describe the test case
        -
    - Document generic solutions in the form of recommendations
        - Body type datasets
        - IIIF manifests
    - John will send around a big Json for testing
    - Next meeting: what are best practices? What elements are required or suggested? What is the application profile? How to systematize?
- **We should probably explicitly define what we mean by annotation, so that this definition can be shared and tested against.**

Onomy.org was created because

- Need for interface to enter manually data with hierarchical relations
-