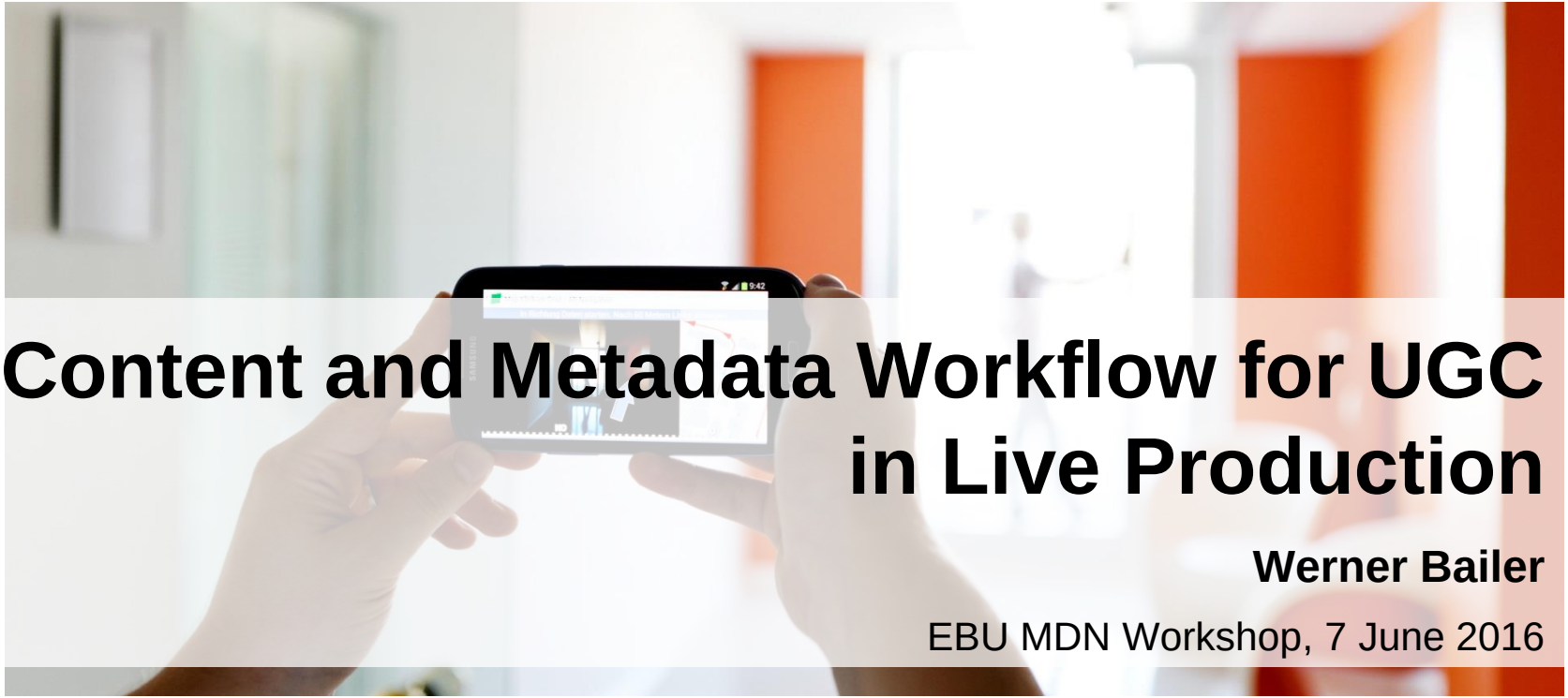
DIGITAL - Institute for Information and Communication Technologies



Content and Metadata Workflow for UGC in Live Production

Werner Bailer

EBU MDN Workshop, 7 June 2016

Outline

- Motivation & Context
- System overview
- Metadata capture, extraction and representation
- Demo
- Conclusion

Motivation

- User generated content (UGC) can complement professional content
  - Unplanned/unexpected events
  - Additional views from events spreading large areas (festivals, sports events)
  - Better convey feeling of being in the crowd
  - Quality of capture devices is improving

JOANNEUM RESEARCH DIGITAL

Motivation

- Live UGC is becoming available
  - Apps such as Meerkat, Periscope
  - Streams are provided to viewers “as is”
  - No integration with editing systems
  - Quality is sometimes questionable

Context

- ICoSOLE: Immersive Coverage of Spatially Outspread Live Events
  - www.icosole.eu
  - 2013-2016

Context

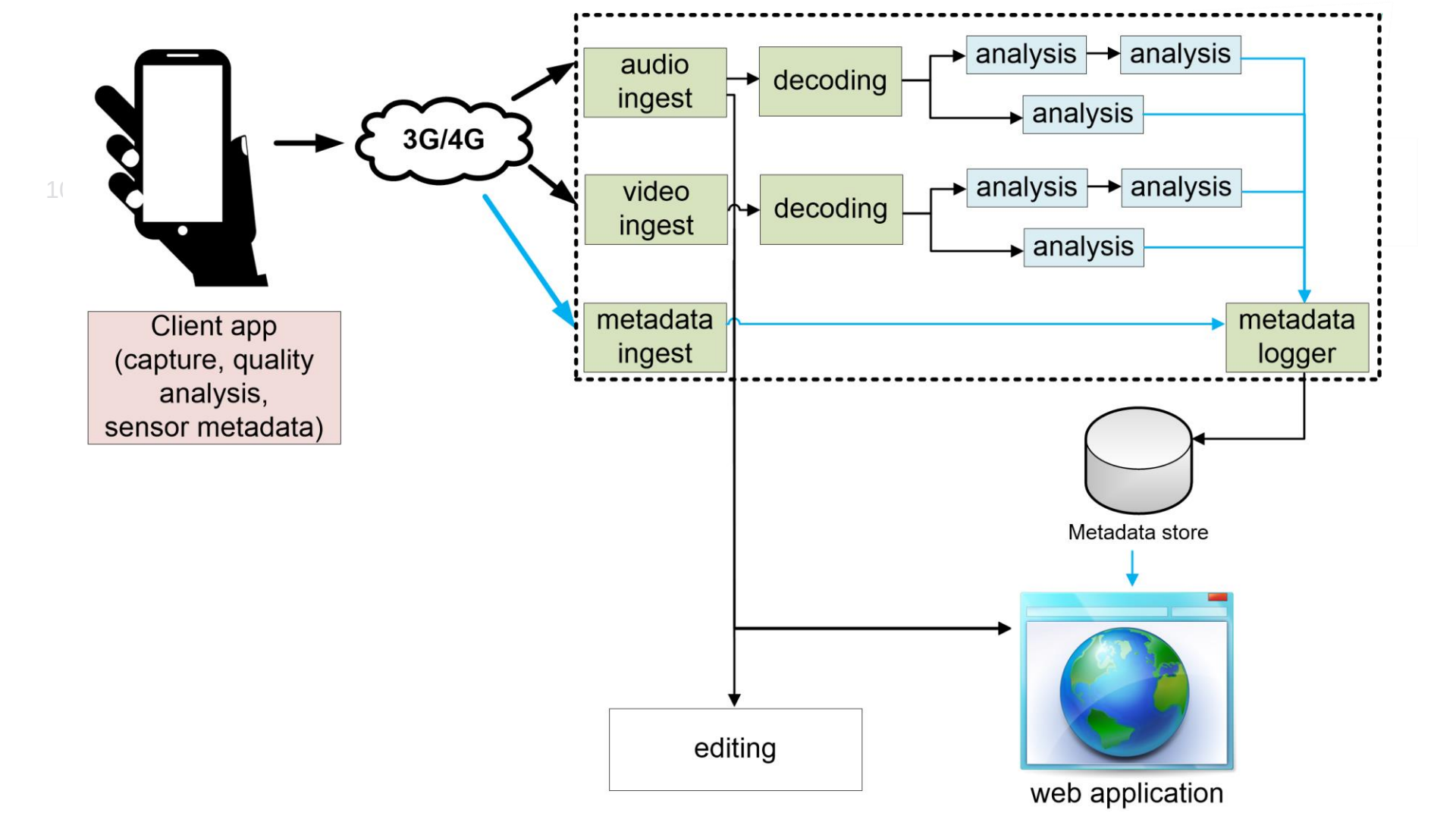
- Enable a more comprehensive and immersive experience of live events
  - both for on-site and remote media consumers
  - deliver content to broadcast, web, second screen and mobile
- Many events are spread over large areas
  - only high-profile events can be fully covered by professional capture
  - seamlessly integrate professional and user generated content

R&D areas

- Content capture
- Video content fusion
- Video content selection/filtering
- Professional production tools
- Content distribution
- Audio production/delivery/presentation

System overview

- Capture app for mobile devices
- Ingest and content analysis
- Metadata storage
- Visualisation and interfaces to editing, content management, …

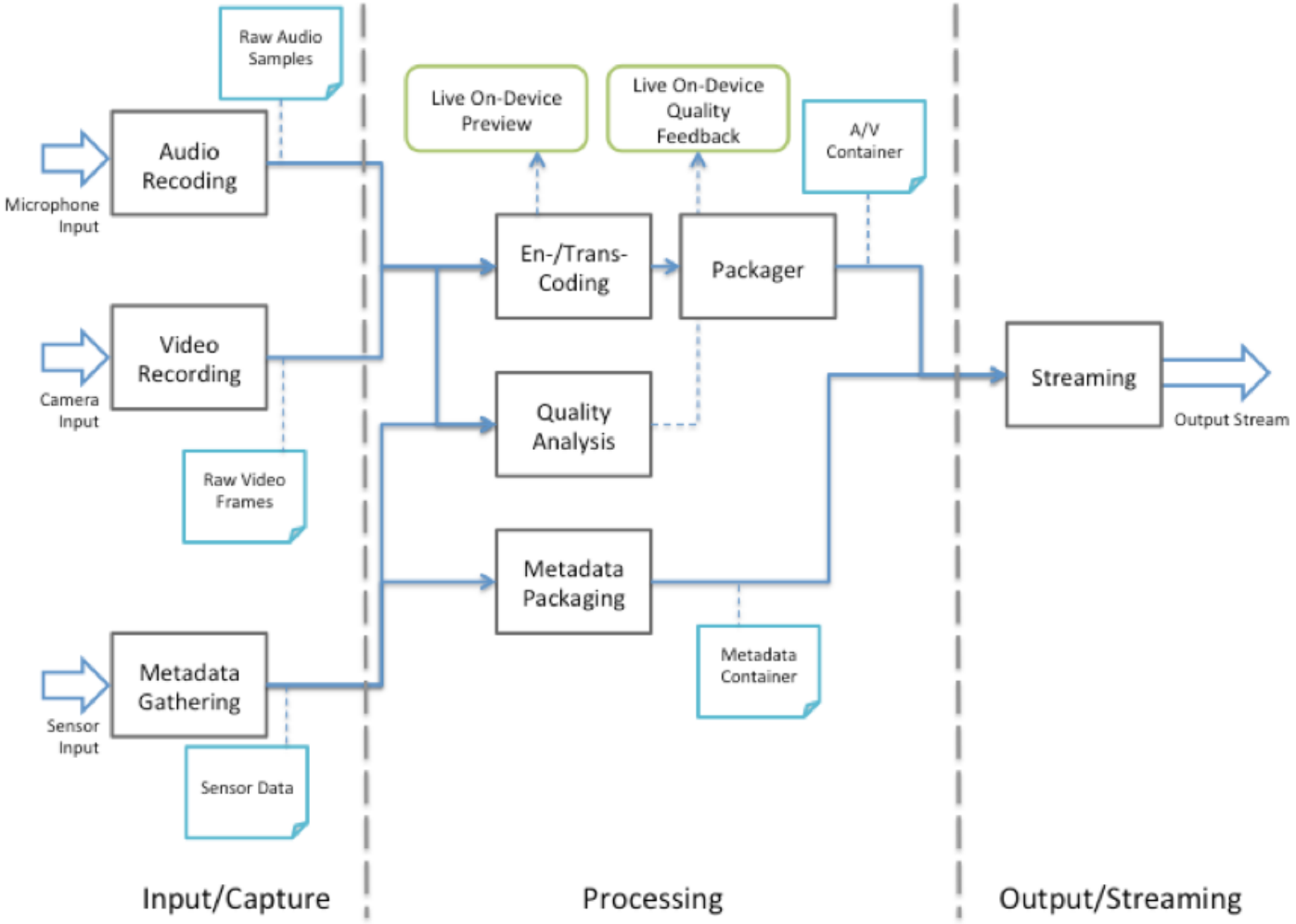

Client app
(capture, quality
analysis,
sensor metadata)
3G/4G
audio
ingest
decoding
analysis
analysis
analysis
video
ingest
decoding
analysis
analysis
analysis
metadata
ingest
metadata
logger
Metadata store
editing
web application

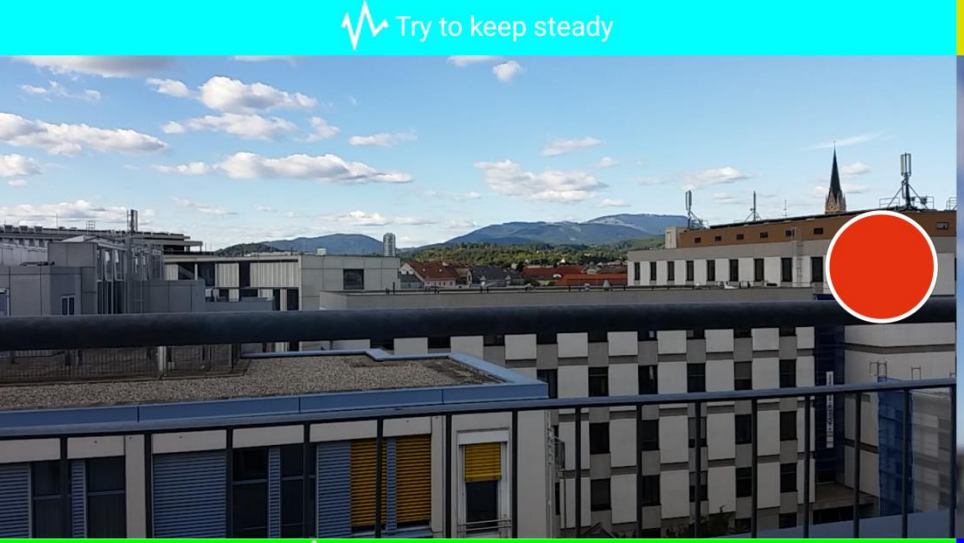
Capture app

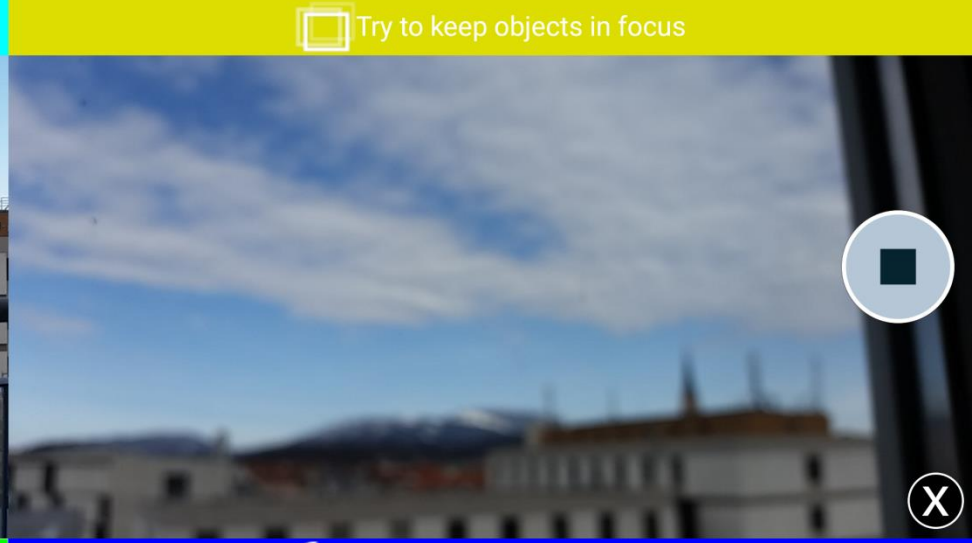
- Video & audio streams
  - using HW encoder of mobile device
- Sensors
  - location, accelerometer, gyroscope, magnetic field, orientation, rotation, ambient light, proximity and pressure
- Wrapping and streaming (RTP)
  - ISOBMFF container
  - streams for video, audio, metadata

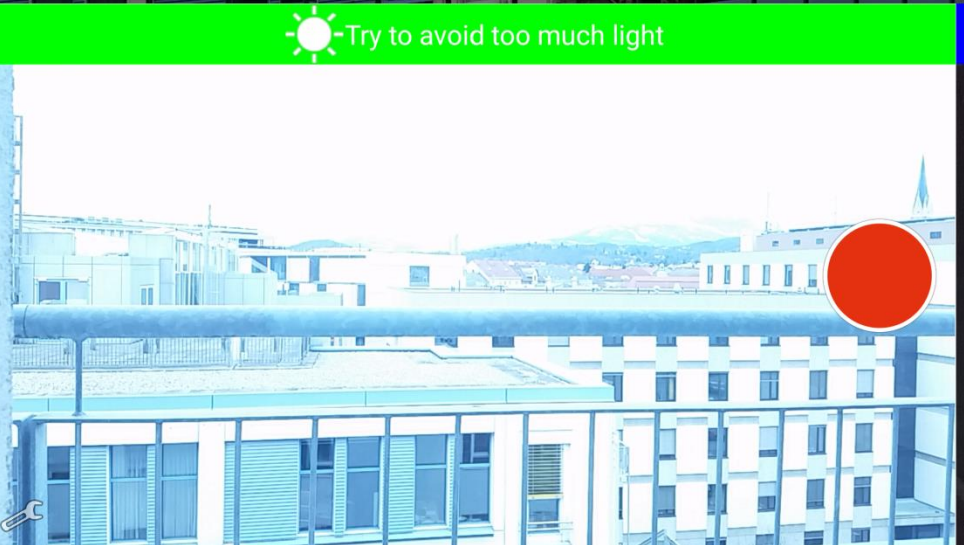
Capture app

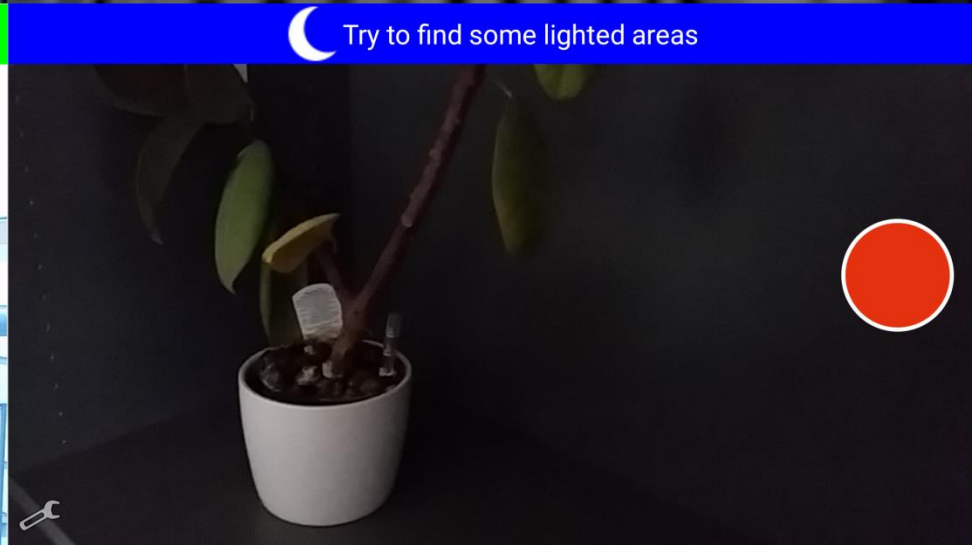
- Visual quality analysis on mobile device
  - low-complexity versions of algorithms
  - detect problems originating before encoding and transmission
  - noise, sharpness, over-/underexposure
- Analysis of sensor data
  - e.g., unsteadiness
- Provide feedback to user about quality issues
  - option to stop streaming if user does not react
- Send quality metadata to server


Raw Audio Samples
Live On-Device Preview
Live On-Device Quality Feedback
A/V Container
Audio Recoding
Microphone Input
En-/Trans-Coding
Packager
Video Recording
Camera Input
Quality Analysis
Streaming
Output Stream
Raw Video Frames
Metadata Packaging
Metadata Container
Metadata Gathering
Sensor Input
Sensor Data
Input/Capture
Processing
Output/Streaming


Try to keep steady

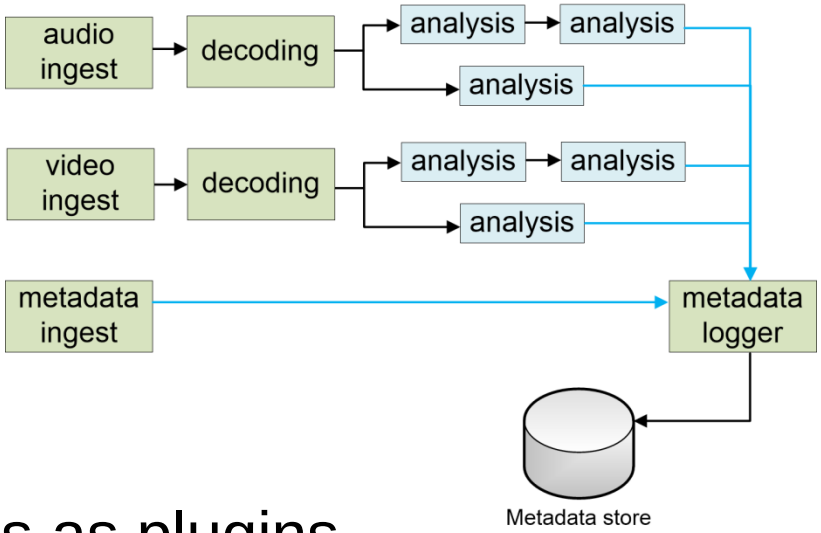

Try to keep objects in focus


Try to avoid too much light


Try to find some lighted areas

Backend system

- Receive, unwrap, decode
- Perform further metadata extraction (real-time)
- Implemented based on Gstreamer
  - wrapping analysis algorithms as plugins



Metadata extraction

- Quality analysis
  - more complex versions of quality analysis algorithms (sharpness, noise) for more precise results
  - macroblocking, (block) dropouts
  - audio QA: clipping (further modules under development)
  - determine overall quality score from individual metrics
- Visual similarity matching (work in progress)
  - match incoming stream against location images, other (professional) streams to determine overlap in view
  - visual similarity score for segments wrt. other streams

Synchronisation

- Backend server is synced with master clock (e.g. PTP)
- Synchronisation protocol to measure offset and network delay of mobile device
  - repeated on initialisation and during capture
  - measured times are sent with metadata stream
- Backend system corrects time stamps in stream headers
- Issues
  - high framerate jitter on mobile phones under full load occur
  - jitter in network delays

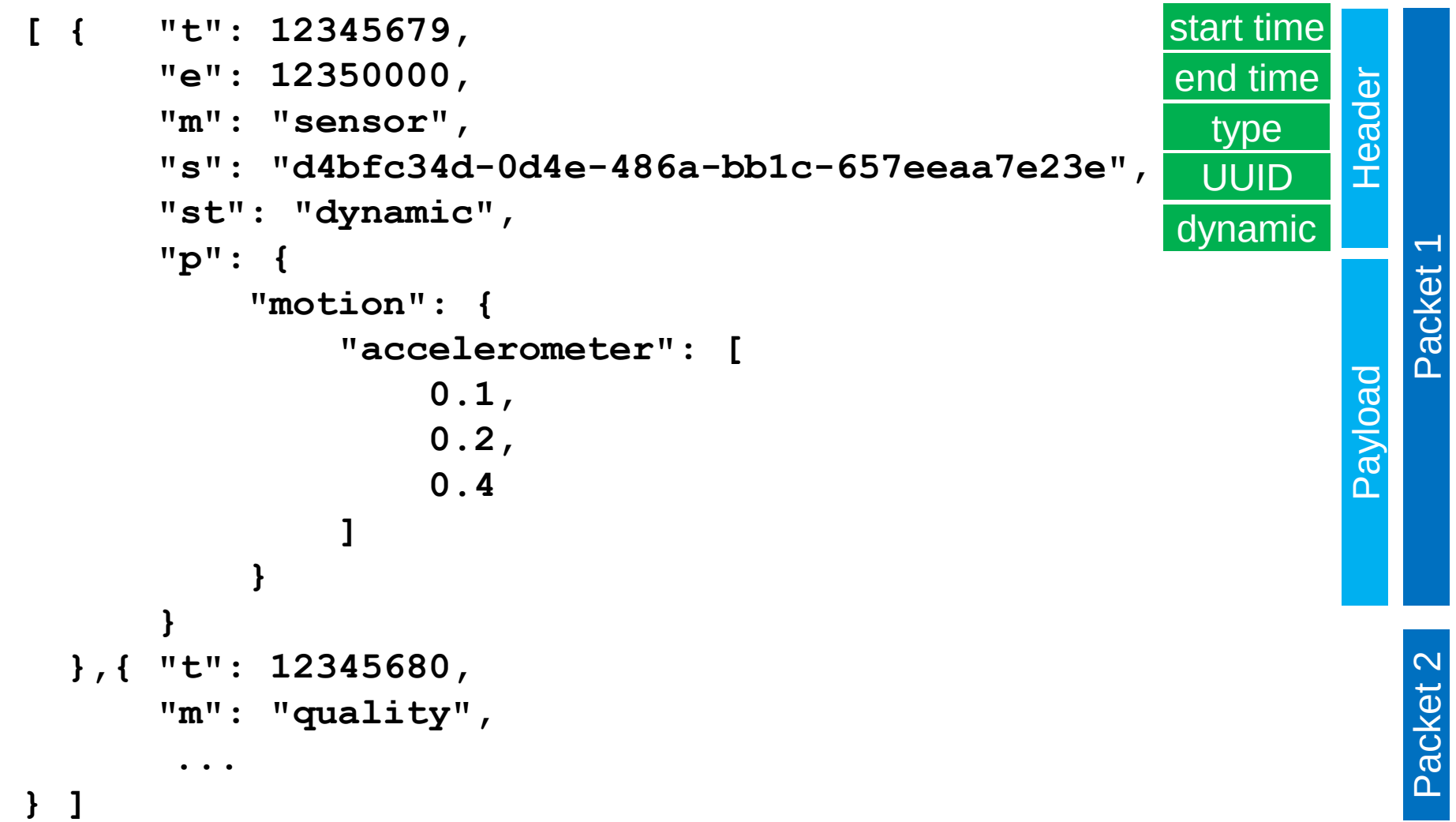
Metadata handling

- Metadata sources
  - static information from devices (type, capabilities, …)
  - sensor data
  - extracted metadata (from mobile device and backend)
- Metadata is treated as a stream like a/v essence
  - sent as UDP packets between machines
  - any component can connect to the metadata stream
- Store time indexed metadata

Metadata representation

- Time-indexed packets
- JSON
  - compactness
  - easy integration in web applications
- Specified using JSON Schema
  - being aware of some issues [1]
  - following a more restricted syntax with conversion rules from/to XML schema [2]

[1] http://dret.typepad.com/dretblog/2016/05/json-schema-why-and-how.html
[2] http://www.xml.com/pub/a/2006/05/31/converting-between-xml-and-json.html


[{
"t": 12345679,
"e": 12350000,
"m": "sensor",
"s": "d4bfc34d-0d4e-486a-bb1c-657eeaa7e23e",
"st": "dynamic",
"p": {
"motion": {
"accelerometer": [
0.1,
0.2,
0.4
]
}
}
},{ "t": 12345680,
"m": "quality",
...
}]
start time
end time
type
UUID
dynamic
Header
Payload
Packet 1
Packet 2

Metadata representation

- Representation of payload
  - JSON representation of existing specification (e.g. XML Schema)
  - MPEG-7 (e.g., persons, objects, regions)
  - EBU QC output model
  - EBU ADM

Metadata storage

- Indexing real-time metadata
  - Redis: fast, in memory database
  - store metadata for recent time window (typically hours)
  - older metadata is moved to relational DB (MySQL)
- Metadata store interface
  - querying for metadata by time, type, source, … using a REST interface
  - transparent where data is stored (only response time is different)

Content filtering/selection

- Discard content below minimum quality requirements
- Create ranked list of available streams
  - overall quality score
  - uniqueness of location, tags, visual content
  - currently rule-based
  - ongoing work: learning from selection made by production team (includes e.g. preferences for trusted users)

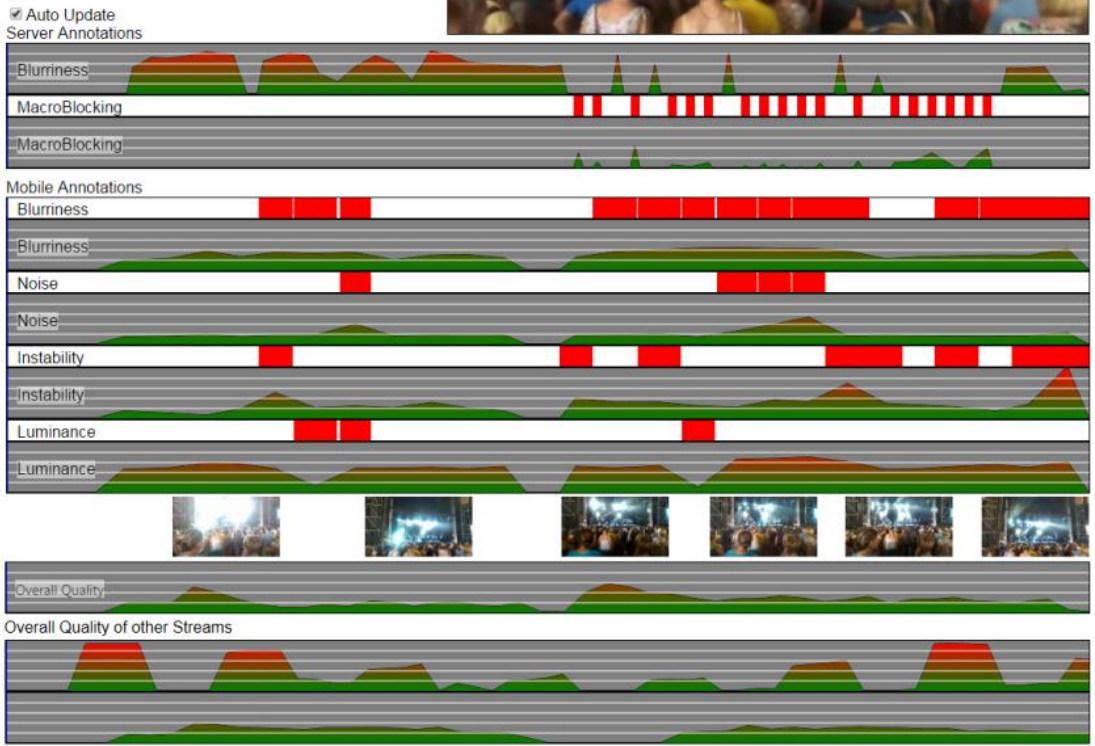
Visualisation

- HTML5 based viewer
- query new metadata from store and update plots
  - includes sensor data, and metadata extracted on phone and in backend system
- video player
  - directly use RTP stream: lower latency, browser plugin required
  - serve as HTTP stream or via WebRTC: additional plugins doing conversions in Gstreamer chain needed

Integration

- Available streams can be provided to further components
  - live mixing, editing
  - content storage
- Discovery of streams via metadata store REST interface
- Easy integration for tools that can ingest RTP streams
  - possibly header extensions need to be inserted (device ID, timecode)
  - ongoing integration tests with BBC‘s IP Studio infrastructure, Tools On Air‘s mixing engine

Demo


Auto Update
Server Annotations
Blurriness
MacroBlocking
MacroBlocking
Mobile Annotations
Blurriness
Blurriness
Noise
Noise
Instability
Instability
Luminance
Luminance
Overall Quality
Overall Quality of other Streams

Conclusion

- Capture and processing of live UGC
- Metadata chain collecting metadata generated/extracted at different stages
- Gstreamer-based backend system
- Interface to further production chain

Werner Bailer
JOANNEUM RESEARCH – DIGITAL

werner.bailer@joanneum.at
http://www.joanneum.at/en/digital

This work has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 610370, ICoSOLE.