



ANUSH VENKATARAMAN
FILMSTER NETWORK

BIG DATA LDN
24 SEPTEMBER 2026

Building the Knowledge Graph of Cinema

TURNING FILM INTO DATA INFRASTRUCTURE

A FILM BY
Anush Venkataraman

A film can take a year and a village to make. We put one person's name at the beginning, and the end.

For an audience, this is authorship. But behind this frame is an enormous production system.

There are actors, cinematographers, designers, editors, sound departments, production teams, locations, equipment, suppliers and thousands of people and processes.

There are decisions, dependencies and relationships. Some exist for a single shooting day. Some survive across twenty films.

The credits tell us who did what, but considerably less about how all these people, decisions and relationships came together to make the film.

And that is the question we have been exploring at Filmster Network.

What if everything behind this frame could become connected, evolving, computable knowledge?

Metadata is the new film negative



3

Cinema generates an enormous amount of metadata.

We are very good at recording the obvious things: title, year, director, crew, cast, genre.

But a production contains far more.

People. Relationships. Locations. Assets. Events. Decisions. Time. Context. Skills. Processes. Resources. Outcomes.

Every production generates these through schedules, call sheets, budgets, scripts, production notes, images, video, availability, changes and eventually an impressive archaeological site of spreadsheets.

The problem is that each record usually describes one tiny part of the production.

A credit tells us that somebody worked. A production record tells us where something happened. A schedule tells us when. A journal might explain why something changed.

Individually, these are records.

Together, they contain a latent representation of the production.

That is why we think metadata is the new film negative, and it gets interesting when we develop the relationships inside it.

Start with a filmmaker



4

So where do we begin?

With a filmmaker.

A director creates a project. The project connects the director to people- the cast and crew, processes and assets, which further connects to their events, stills, and locations.

Now follow one of those people into another project.

That introduces another director, another crew, more locations, more processes and another production history, eventually everything becomes connected knowledge.

A filmmaker creates the network. The network accumulates knowledge. And that knowledge eventually returns to a filmmaker.

But this is just a very confident-looking diagram.

We still must understand what those connections mean.

Three ways we see a film

RECORDED

STRUCTURED

FILMSTER-0000-0000-0001
ANUSHVENKATARAMAN
DIRECTOR, CREATIVE PRODUCER
LONDON, UK
COMEDY, ROMANCE

EXPLICIT RELATIONSHIPS

EXPRESSED

UNSTRUCTURED

"I worked with Hugh Grant on this call sheet."
"We had to reschedule the entire filming that week."

EXTRACTED RELATIONSHIPS

VISIBLE

VISUAL



VISUAL EMBEDDINGS

ENTITY • RELATIONSHIP • TIME • CONTEXT

5

That begins with three broad forms of data.

Structured data.

These are things we explicitly record: identities, projects, departments, roles, skills, locations, availability, filmography and known relationships.

Unstructured data.

Production notes, journals, scripts, interviews and documents contain relationships nobody has explicitly entered into a database.

Suppose a filmmaker writes:

"I worked with Hugh Grant on this call sheet."

Neural relation extraction can identify the entities and candidate relationship.

Because the statement is already associated with a filmmaker and a project, we can resolve those identities and examine related records.

Visual data.

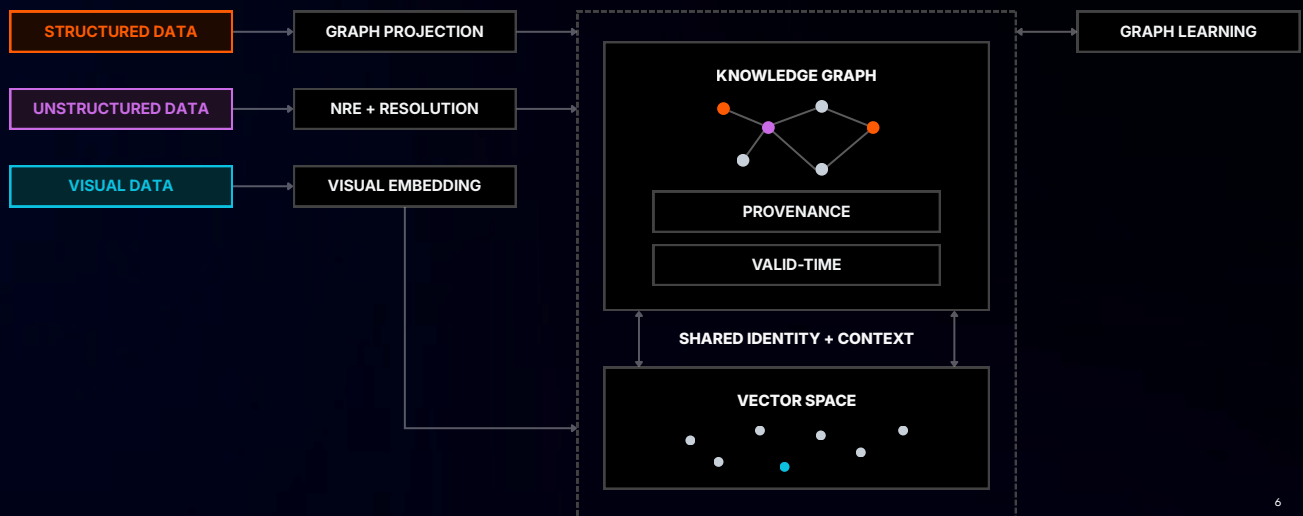
Frames, stills, references and footage contain information that text cannot completely represent. Visual embeddings give us mathematical representations through which we can investigate visual similarity.

So, three ways we see a film:

Recorded. Expressed. Visible.

And underneath, we need common identity, relationships, time and context.

The Knowledge Model



6

Now we start connecting them.

Structured data can be projected into the Knowledge Graph.

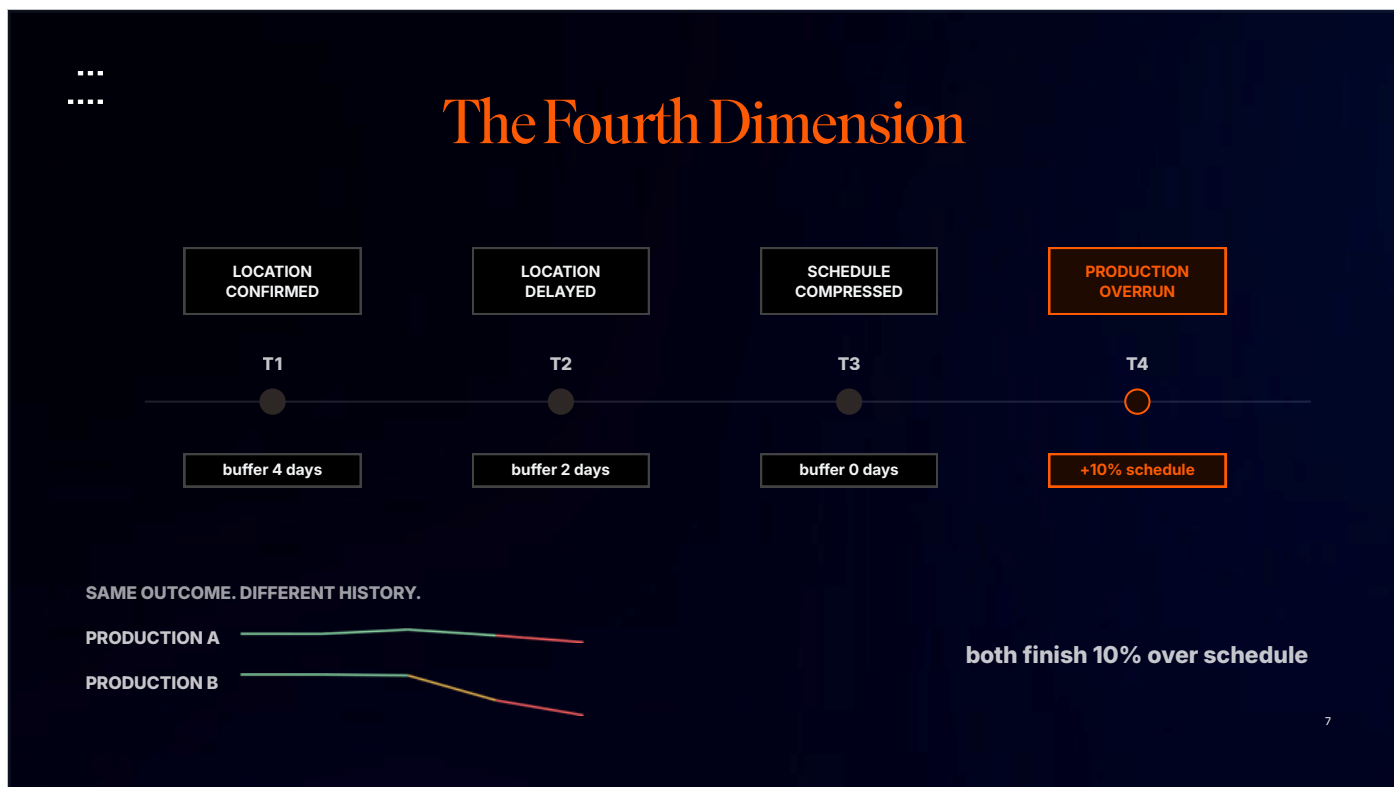
Unstructured data passes through relation extraction and entity resolution.

Visual data is represented through visual embeddings in vector space.

And these do different jobs.

The graph tells us how things are related. Vector space tells us how things are similar. They remain connected through shared identity and context.

Once sufficient relational structure exists, we can learn from the graph itself to investigate relational similarity, discover candidate relationships and identify potentially incomplete knowledge.



There is still something missing.

Filmmaking is constantly changing, and relationships have valid time, because what is true today may not have been true six months ago.

Time!

A location is confirmed. Then it is delayed.

The schedule compresses. The buffer disappears.

Eventually, the production overruns.

These meaningful changes become events.

An event tells us what happened.

A sequence of events allows us to reconstruct what happened at a particular point in time.

A collaboration begins. It may recur on another project. Someone's role changes. Availability changes. Production conditions change.

So rather than continually replacing history with the latest value, we preserve the evolution of that state - more like versioning our code.

Look at the two productions at the bottom.

Both finish ten per cent over schedule.

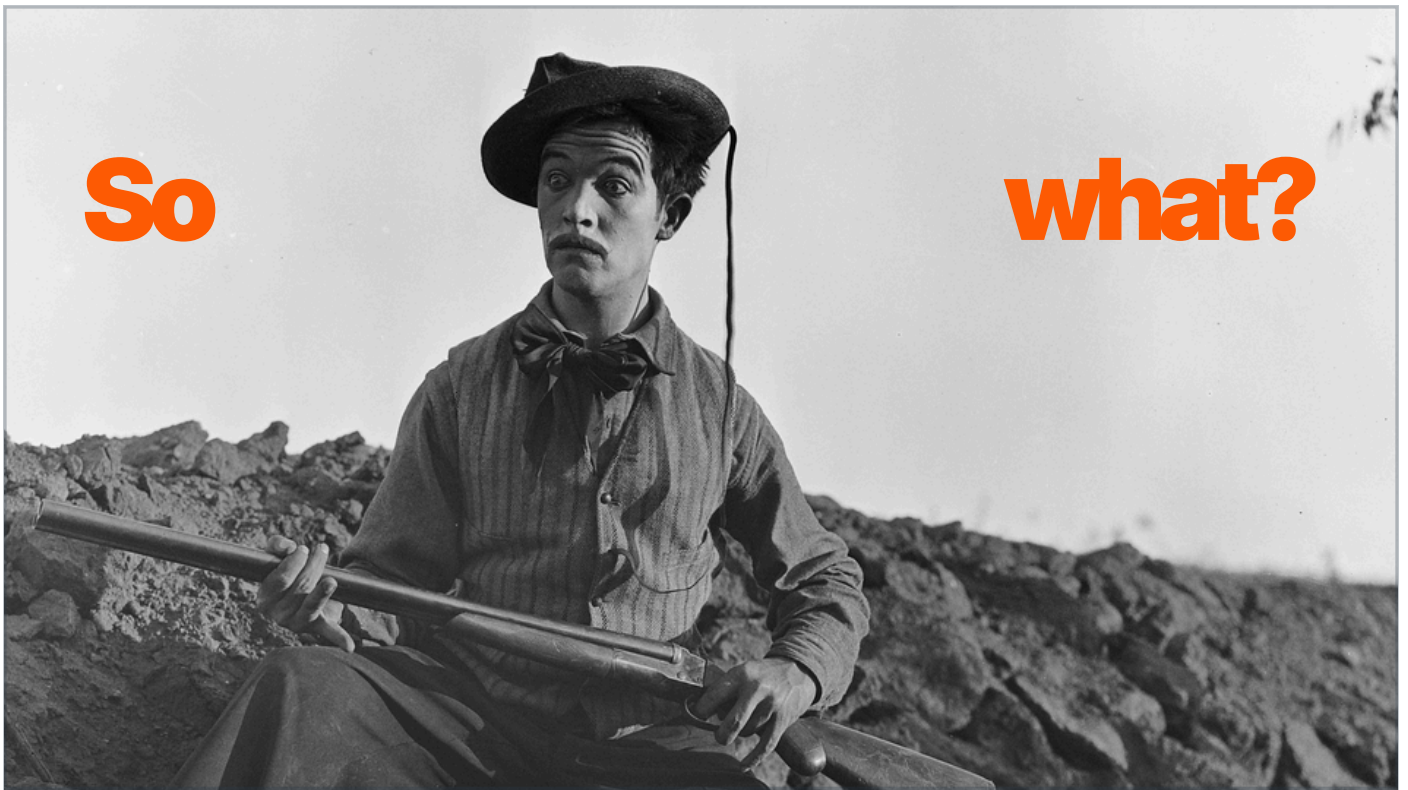
If we only store the outcome, they look identical.

But one may have gradually consumed contingency through deliberate production decisions.

The other may have suffered a rapid loss of capacity after a location problem.

Same outcome. Different history.

The graph tells us what is connected. Time tells us how the system evolved.



So what?

We now have structured data, relation extraction, visual embeddings, a Knowledge Graph, vector space, provenance, graph learning and temporal history.

Technically interesting.

But if an experienced producer can answer the useful part of my question from memory, we probably don't need all of this.

The value begins where individual memory stops being enough.

And that changes when we introduce scale.

A collective memory of cinema



The value begins where individual memory stops being enough.

9

One filmmaker can accumulate extraordinary knowledge across a career.

But there is a physical limit.

Now imagine ten thousand filmmakers contributing production histories.

Different productions that have never interacted may contain comparable states.

Filmmakers who have never met may have similar positions within completely different professional networks.

Projects may encounter similar combinations of constraints and then evolve in completely different ways.

No filmmaker could experience all these productions.

No producer could remember all of them.

No studio executive could simultaneously hold every relationship, event, state and trajectory in their head.

Once thousands of production histories contribute to the same knowledge architecture, we begin to get a collective memory of cinema - sounds Asimovian!

And now the question changes.

What do we do with this collective memory?

Film to Data Infrastructure

Metadata

scene took 9 hours
location changed
buffer consumed

Knowledge

connected across people,
projects, events, assets
and decisions

Intelligence

relevant insight and
evidence for the current
production state



10

This is the transformation between metadata, knowledge and intelligence.

Metadata tells us what happened.

A scene took nine hours. A location changed. A production buffer was consumed.

Useful facts.

Knowledge emerges when those facts become connected across people, projects, events, assets, decisions, relationships and time.

And Intelligence begins when that accumulated knowledge becomes relevant to the production currently being made.

For example,

We have a night exterior. A large cast and crew. Weather exposure. A low production buffer.

Individually, those are conditions.

But across many comparable productions, we may discover that this combination is associated with higher schedule variance.

We can then surface the relevant historical insights to the current production.

And crucially, we can study the productions that recovered successfully as well as the ones that failed.

So, Intelligence is not merely telling filmmakers what happened before.

It is making accumulated production knowledge available in the context where it becomes useful.

But there is an uncomfortable question here.

What makes that intelligence intelligent?

...
...

What makes Intelligence intelligent?



"As soon as it's done cleaning the house, it brings in trash from the street, and starts all over again!"

CONTEXT

STATE

CONSTRAINT

DEPENDENCY

JUDGEMENT

When a measure becomes a target, it ceases to be a good measure.

GOODHART'S LAW

11

Having enormous amounts of knowledge might produce a healthy filmmaking ecosystem in which more films can be made.

But if we begin optimising that number as the target, a system could favour shorter development, smaller crews, lower contingency, familiar formats and repeated collaborators.

Excellent. We made more films. We may also have reduced experimentation, resilience, crew sustainability, creative diversity and opportunities for new talent.

So, intelligence cannot simply mean maximising a metric.

Return to our two productions that both finished ten per cent over schedule.

The final measure says they are the same. Their histories tell us otherwise.

Production Intelligence should understand the conditions around what state the production is in, what constraints exist, what depends on what, what buffers remain, what happened before, and what production data supports the pattern – and how?

We already have those conditions represented in the Knowledge Model.

The graph gives us the relationships and dependencies. The temporal model gives us

the sequence of events and changing states. The production data gives us constraints, resources and available buffers.

So, we can represent the current production as a state, find comparable states and trajectories across previous productions, and examine what happened next.

We could say something closer to:

“Under these conditions, additional locations historically increased these dependencies and risks. Here’s the insight, the current production state, the available contingency.”

And still the filmmaker will want to say- “we’ll film anyway” - but now with caution.

...
...

The story doesn't change. We do.



12

Children understand recursive learning remarkably well.

They will ask for the same story forty-seven times, apparently just to make sure we understand it too.

The story remains the same.

But the person experiencing it changes.

The first time I get an idea, I may think: “Ummm!”

After five projects, I encounter the same idea and recognise a pattern.

After twenty projects, I may finally understand the conditions under which that pattern actually holds.

Professional relationships evolve in the same way.

“A worked with B” may begin as a single production relationship.

Across time it may become repeated collaboration, knowledge exchange or mentorship.

Thus, the interpretation of patterns must evolve too.

Something observed across fifty productions may change after five hundred.

After five thousand, we may discover it only applies under particular conditions.

So, the graph becomes larger and richer as Filmster learns.



And once we follow these relationships far enough, we see something much larger than the film industry.

A film already interacts with all these systems.

We could take something as simple as James Bond and Aston Martin.

At the most basic level:

Aston Martin appears in a Bond film.

True, but spectacularly uninteresting.

A richer Knowledge Model can connect the vehicle to a scene, character, narrative context, visual treatment, geography, cultural association and audience response across place and time.

Now that relationship becomes relevant to marketing, tourism and cultural research.

Follow a filming location and we reach tourism.

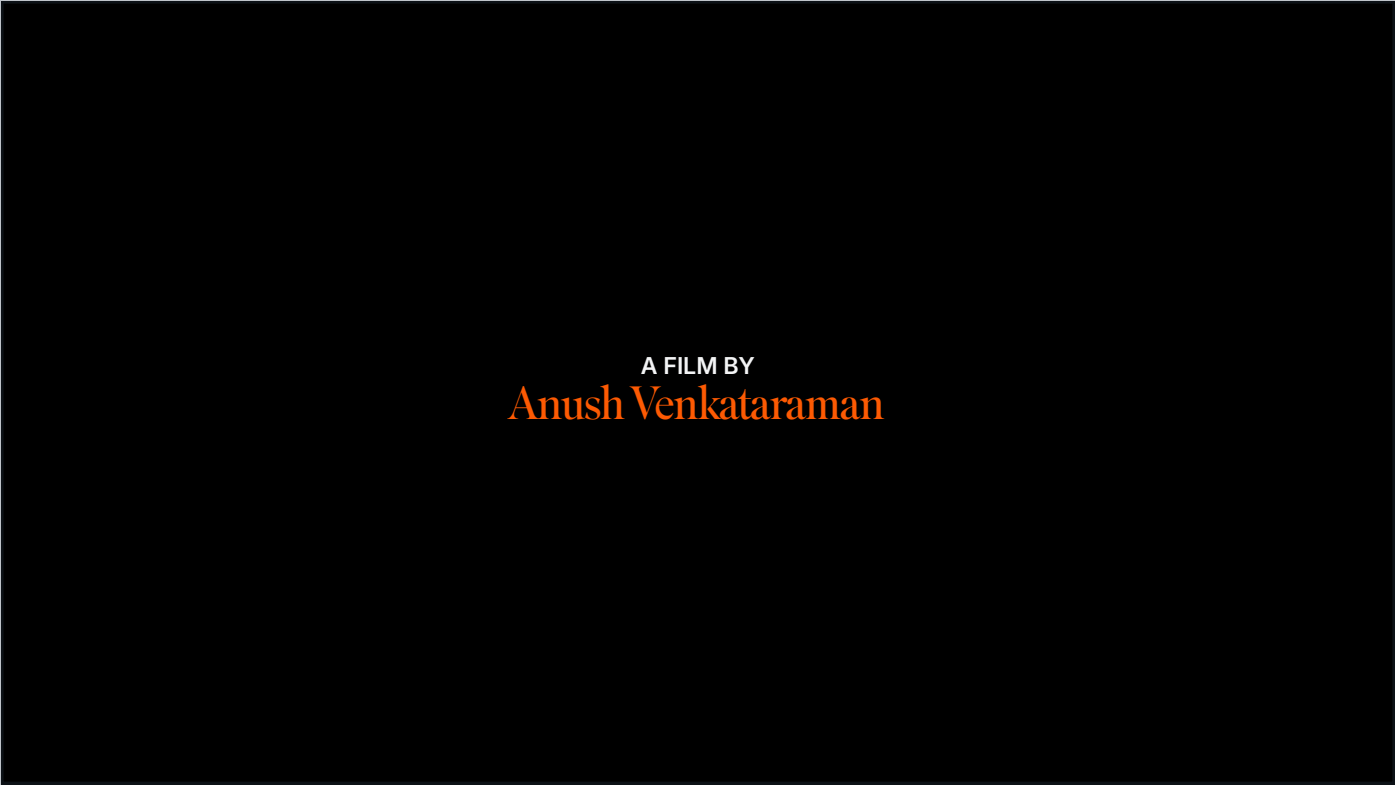
Follow production accommodation and we reach hospitality.

Follow equipment and we reach technology.

Follow workforce demand and we reach education and skills.

Follow physical production and we reach construction, property, transport, finance and insurance.

And this is where our longer research direction begins: exploring how this architecture can evolve into a domain-specific Large Knowledge Model for cinema.



A FILM BY
Anush Venkataraman

And now we return to where we started.

Nothing on this screen has changed.

But hopefully, what you see behind it has.

We simply developed a richer representation of what's already there.

That is what we are building at Filmster Network.

Metadata becomes connected knowledge.

Across thousands of production histories, that knowledge begins to resemble a collective memory.

Production Intelligence makes relevant parts of that accumulated experience visible when filmmakers need them.

And eventually, everything returns to exactly where we started.

The filmmaker.

Technology can expose relationships.

It can preserve history.

It can reveal patterns no individual person could reasonably see.

But it cannot author the film.

Cinema is still a human creative process.

A bad filmmaker with an advanced knowledge model is still a bad filmmaker.

And however sophisticated our knowledge systems become...

we still remain wonderfully inconvenient.

Thank you.

Connect and build with us



ANUSH VENKATARAMAN
FILMSTER NETWORK