A Container for the Whole Dental Case: why the oral cavity needs a new interchange format, and what Gaussian splatting does, and does not, add to it

Luis Garbayo Fernández[a,b]

In collaboration with [a]ANFAIA & [b]Histora

August 29, 2026

Abstract

A complete dental case is today three unrelated artefacts — an intraoral scan, a cone-beam CT series and written reports — each in its own coordinate system, each opened by different software, with the relations between them living in a clinician's memory or inside a proprietary tool. Those relations do not travel with the data, so they are rebuilt every time a case changes hands.

This paper argues that the missing piece is not a better mesh format or a better volume format, but a *composition* format: one that references native formats intact and declares, explicitly and auditably, the spatial and temporal relations among them, the provenance of each datum, and the boundary between what was measured and what a model inferred. We describe such a format, UOS (Unified Oral Scene), specified separately in the UOS Format Specification v0.2, and report what building and measuring it taught us.

We give particular attention to 3D Gaussian splatting, the representation the project adopted, and to a result that constrains how it should be used: **splatting adds no resolution and cannot**. What it adds is the ability to allocate representational capacity where the data actually carries information, which pays only when the modalities being combined have different resolutions — which is precisely the dental case. We report a photometric ceiling that no amount of budget, resolution or spatial concentration moves, and the one thing that does move it: decomposition into several specialised fields, worth +2.08 to +7.41 dB of PSNR across three real CBCT volumes.

We also report three lines of work that did not reach their objective, each with the route the measurement itself identifies. Per-tooth segmentation on the intraoral scan produces correct FDI codes in correct anatomical order and *wrong boundaries*, with 11 of 14 crowns rejectable on anatomy alone; a better classifier on the benchmark performed worse on the real task; and a sharpness criterion set before running was not met. These are included because a format whose purpose is to separate measurement from inference is only worth having if its authors apply that separation to their own results.

Keywords: dental informatics, 3D Gaussian splatting, CBCT, intraoral scanning, interoperability, digital twin, provenance, DICOM, glTF

1. Introduction

1.1. The case that arrives in pieces

Consider what a dental practice actually holds after a single patient visit that involved imaging. There is one or two STL files from the intraoral scanner, a directory of several hundred DICOM slices from the cone-beam CT, a handful of clinical photographs in whatever the camera wrote, and a PDF report. Four artefacts, four programs, four coordinate systems, and no machine-readable statement of how any of them relate to any other.

The relations are not absent, somebody established them. Someone looked at the CBCT beside the scan and knew which surface was which; someone read "distal caries on 24" and knew which of the fourteen crowns on screen that named. But that knowledge lives in a person or inside a vendor's software, and **it does not travel with the data**. When the case goes to the

laboratory, to a second opinion, or to a reviewer six months later, the relations are reconstructed from scratch — or, more often, not reconstructed at all, and the recipient works with whichever artefact their software opens.

1.2. Why this is not solved by picking a better format

The instinct is to look for an existing container. Each candidate fails on a different axis, and the failures are instructive because they are not deficiencies: they are consequences of what each format was designed for.

Format	Why it does not carry a dental case
DICOM	Models acquisitions superbly and is the reason CBCT interoperates at all. It has no primitive for a Gaussian field, no natural web transport, and its model of "study" does not extend to a surface scan plus a photograph plus a report as one composed object.
glTF	The right scene graph, mature tooling, an extension mechanism that works. No volumes, no clinical metadata, no provenance, no notion of a regulatory boundary.
OpenUSD	Composition is precisely what USD is for, and it solves layering better than we do. It is foreign to the clinical ecosystem: no DICOM path, no FHIR path, and an implementation burden a dental viewer cannot justify.
FHIR	The right vocabulary for clinical facts and the right target for a practice-management connector. **FHIR Release 4 has no resource for dental geometry**, and `Media` is defined for photo, video and audio. Meshes end up as `DocumentReference`, which is honest but is not interoperability.

The observation that organises this paper is that none of these formats is missing *data*. They are missing the *relations*, and the relations are what a second reader needs and what no format currently obliges anyone to state.

1.3. What is actually lost, and where

It is worth being concrete about the cost of the fragmentation, because "the data is not integrated" is the kind of statement everyone agrees with and nobody acts on.

The clinician who acquired the case has the relations in their head and does not need them written down, which is exactly why they never get written down. The cost lands on the second reader. A colleague opening the same case a week later sees a grey arch in one program and a stack of slices in another, and must decide whether a finding described in prose belongs to the crown they are looking at. In practice they re-derive what they can and take the rest on trust.

What reaches a dental laboratory is usually an STL and a written instruction. The STL is geometry and nothing else: no colour, no units declared, no statement of which tooth is which, no indication of what was measured versus what the scanner interpolated across a gap. Everything the clinic knew that is not geometry has to be re-expressed in prose, and prose does not survive a workflow with three handoffs.

Longitudinal questions — has this margin receded, has this bone level dropped — are arithmetic if two visits share a coordinate frame and a research project if they do not. Today they do not, so the comparison is either done by eye or skipped. This is the loss we find hardest to justify, because both acquisitions already exist and the only missing element is a transform with an error bar.

A case archived as four loose artefacts is a case whose interpretation depends on software that may not exist in five years. Formats outlive vendors only when they are open and self-describing; a folder of files plus a convention held in one company's product is neither.

1.4. Why now

Three things changed recently enough that the problem is worth attacking today rather than five years ago.

1. **Multimodal acquisition became ordinary.** A practice that had a CBCT and an intraoral scanner in different buildings a decade ago now has both, routinely, on the same patient in the same week. The relation between them stopped being an exotic research problem and became a daily one.
2. **A representation appeared that can hold surfaces and volumes in one substrate.** 3D Gaussian splatting is fast, differentiable, and web-renderable, and it gives the project a common primitive for things that were previously meshes or voxels and never both. Whether that unification actually helps is an empirical question, and §5 is our answer to it.
3. **The regulatory boundary hardened.** As dental software incorporates models, the distinction between a measurement and an inference stops being an epistemic nicety and becomes the line that determines what may be distributed where. A format designed before that pressure has no place to put the distinction; one designed now has no excuse.

1.5. The shape of the answer

Our claim is that the missing artefact is a *composition* format, and that composition is a different problem from encoding. Encoding asks how to store a mesh or a volume well, and it is largely solved. Composition asks what else must be written down so that the stored things remain usable together: which coordinate frame each lives in, what transform relates them and how well it fits, which visit each belongs to, which of them a model produced, and what any of it is worth.

Stated that way, the design follows from a single discipline, which we adopt throughout and state here so that the rest of the paper can be read against it:

Two things that are not the same must never look the same

An automatic alignment nobody reviewed and one signed by a clinician. A colour sampled from the patient's photographs and one from a fallback gradient. A tooth boundary drawn by a network and one drawn by a person. A DICOM series held inside a container and one merely identified by hash.

Each pair is easy to store identically and expensive to confuse. Where the format can enforce the distinction it does; where it cannot, it is required to *name* it. Most of what looks like bookkeeping in [1] — a residual on every registration, a verification field defaulting to "nobody", a per-value declaration of how a clinical figure was extracted — is this one rule applied.

1.6. Contributions

1. A statement of the interoperability problem in dentistry as one of composition and provenance rather than encoding (§2).
2. A container format that references native formats intact and declares their relations explicitly, with a published schema and a validator. The normative document is the *UOS Format Specification v0.2* [1]; this paper motivates and evaluates it rather than restating it.
3. An empirical characterisation of what 3D Gaussian splatting contributes to volumetric and surface dental data (§5).
4. A statement of the mathematics the implementation rests on, including a derivation of the capacity ceiling of the volumetric field from the rasteriser's numerics (§4).
5. Measured results across registration, reversibility, volumetric reconstruction, per-tooth colour, gingival boundary detection and segmentation — including three lines of work that

did not reach their objective, each reported with the route the measurement identifies (§6, §7).

2. Problem statement

We separate the problem into five gaps. Each is stated as what a recipient cannot currently do.

2.1. G1: The relations do not travel

A CBCT and an intraoral scan of the same mouth are in different coordinate systems. Aligning them is a computation someone performs, with a residual error and a method. Today that alignment is either redone by each recipient or is baked irreversibly into a merged artefact. Neither preserves *how well it fits*, which is the number a clinician needs in order to decide whether to trust a measurement taken across the two.

Two failure modes follow, and they are not symmetric. If the alignment is *recomputed* by each recipient, the case has as many geometries as it has readers, and two clinicians measuring the same distance get different answers with no way to discover why. If instead it is *baked* into a merged artefact, there is one geometry but the residual is gone, so a measurement taken across the seam carries an error nobody can quote. The second failure is the more dangerous, because it looks like success.

What makes this tractable is that the transform is small, cheap to store and already computed. Nothing about the state of the art prevents shipping it; there is simply nowhere in the current formats to put it, and no convention that obliges anyone to.

2.2. G2: Measurement and inference are indistinguishable

Once a model has segmented, labelled or completed something, the output looks like the input. A tooth boundary drawn by a network and one drawn by a clinician are the same polygon. In a field moving towards regulated software as a medical device, a format in which those cannot be told apart — or in which the inferred parts cannot be *removed* for distribution in a jurisdiction where the module is not cleared — forces re-export at best and misattribution at worst.

The distinction is not only regulatory. An inference is reproducible only if the model, its version and the weights that produced it are recorded; without them, "the segmentation said 24 was absent" is an assertion with no author. And a case that mixes the two cannot be used as training data without contaminating the next model with the previous one's output — a failure mode that is silent, compounding, and invisible in any metric computed on the mixed corpus.

The requirement, then, is stronger than labelling. Inference has to be *separable*: a recipient must be able to remove it and be left with a complete, valid case rather than a mutilated one.

2.3. G3: Automatic and verified look the same

A registration computed by an algorithm nobody has reviewed and one signed by a clinician carry very different weight and are usually stored identically. This generalises: any pipeline that produces plausible output silently is a pipeline whose output eventually gets trusted for the wrong reason.

The general form is worth naming because it recurs throughout this project. A pipeline that produces plausible output on failure is worse than one that produces obvious garbage: the garbage is caught in the first review, the plausible output is caught after somebody has acted on it. Our own experience supplies three instances — a registration that never tripped the rule written to flag it, a mesh painted with a gradient that looked like tooth colour, and a container that silently carried 216 MB of patient originals it declared it did not. Each was found by measurement rather than by inspection, which is the argument for building the check into the format rather than into anyone's diligence.

2.4. G4: Modalities beyond geometry have nowhere to go

Dentistry is not only surfaces and densities. Shade is colour under a stated illuminant; periodontal status is a set of per-site measurements over time; and there is a growing body of work on the oral microbiome, where the datum is neither a mesh nor a volume but a *per-site* or *per-region* quantity attached to anatomy. A container that can only hold triangles and voxels forces every such datum into a PDF, which is where clinical information goes to become unqueryable.

Three properties separate a datum that is merely *attached* to a case from one that is usable. It must be **addressable by anatomy** (bound to a tooth or a site in a shared vocabulary, not to a page number). It must be **declared**, so that a reader without the schema knows it is skipping something rather than not seeing it. And it must carry its own **units and provenance**, because a number whose scale is implicit is a number that will eventually be compared against one on a different scale.

The oral microbiome is the sharpest example precisely because it is nothing like the rest of the case: no geometry, no image, a per-site quantity from a laboratory process with its own error model, arriving weeks after the scan. If the format can carry that without a new version of itself, the extension mechanism is real; if it cannot, the mechanism was decoration.

2.5. G5: The representation question

Surfaces and volumes are captured at very different resolutions by very different physics. A representation that treats them uniformly wastes capacity on the coarse one and starves the fine one. Whether a unified representation helps here is an empirical question, and §5 answers it.

The question matters more than it may appear, because the appealing answer and the correct one differ. A representation fitted continuously to noisy samples produces output that looks cleaner than its input, and it is easy to present that as recovered detail. Whether it is recovered detail or the representation's own smoothness prior is decidable only by asking whether a feature below the acquisition's resolving power becomes visible — a question with a yes-or-no answer that most demonstrations do not pose.

2.6. What these five have in common

None of the five is a limitation of the underlying data, and none is waiting on a research result. The transform in G1 is already computed; the model identity in G2 is already known to whoever ran it; the operator in G3 is known at the moment of execution; the microbial assay in G4 already has units and a method. In every case the information exists at the moment of production and is discarded before delivery, because nothing in the pipeline obliges anyone to write it down.

That is what makes this a format problem rather than a modelling one. The remaining question, G5, is the only genuinely empirical one, and it is the only one this paper answers with measurements rather than with design.

The design commitment that follows from G1–G3

Two things that are not the same must never look the same. Where the format cannot enforce a distinction, it must at minimum *name* it. Every design decision in [1] that looks like extra bookkeeping — a residual on every registration, a verification field that defaults to "nobody", a per-value declaration of whether a clinical figure was pattern-matched or model-inferred — is this commitment applied.

---

3. Approach

3.1. The case as a graph

The formal object UOS stores is a labelled graph. Let F be the set of coordinate frames named in a case and let one distinguished frame $f_0 \in F$, conventionally the intraoral scanner's, be canonical. Every asset a declares exactly one frame $\phi(a) \in F$. Every registration is an edge carrying a rigid transform

$$T_{ij} \;=\; \begin{pmatrix} R_{ij} & t_{ij} \\ \mathbf{0}^{\top} & 1 \end{pmatrix} \;\in\; SE(3), \qquad R_{ij} \in SO(3),\; t_{ij} \in \mathbb{R}^3,$$

mapping homogeneous points from frame i into frame j, in millimetres, right-handed. Placing an asset in the canonical frame is composition along any path to f_0,

$$T_{i\to 0} \;=\; T_{i_k 0}\, T_{i_{k-1} i_k} \cdots T_{i i_1},$$

which is well defined only if such a path exists. That is the connectivity requirement the format makes normative: every frame named by an asset must be reachable from f_0 [1]. An asset in an unreachable frame is not a partially-placed asset: it is one whose position is undefined, and a viewer that draws it anyway is inventing a pose.

Why the graph is stored rather than collapsed

It would be simpler to resolve every $T_{i\to 0}$ once and store assets pre-transformed. That discards two things. First, the composition path: an error propagates along it, and $T_{i\to 0}$ obtained through two hops is not as trustworthy as one obtained through one. Second, and decisively, the *residual*. A collapsed geometry has no place to record that this edge fits to 0.666 mm and that one was drawn by hand. The graph is kept because the annotations live on the edges, not on the nodes.

3.2. Registration as an estimator with a reported error

The CBCT-to-scanner edge is obtained by point-to-plane ICP over the extracted bone–enamel surface. Given source points $\{p_k\}$ and target points $\{q_k\}$ with target normals $\{n_k\}$, the estimate minimises

$$\hat{T} \;=\; \arg\min_{T \in SE(3)} \sum_{k} \Big((T p_k - q_k) \cdot n_k \Big)^2,$$

and the number the container ships alongside it is the root-mean-square residual

$$\varepsilon_{\mathrm{RMS}} \;=\; \sqrt{\tfrac{1}{K} \textstyle\sum_k \| \hat{T} p_k - q_k \|^2},$$

measured on the reference case at $\varepsilon_{\mathrm{RMS}} = 0.666$ mm. This single scalar is what makes a cross-modal measurement quotable: a crown-to-root length taken across the seam inherits it, and a clinician deciding whether a 1 mm difference is real needs to know that the alignment itself moves by two thirds of that.

The container additionally records *who* produced the estimate. An edge whose operator is a machine and whose `verified_by` is empty is provisional, and both validator and viewer are required to present it as such, which is the direct expression of G3.

3.3. Composition, not encoding

Everything above is metadata about payloads the format does not re-encode. Acquired originals are referenced by content address $H(\cdot) = \text{SHA-256}$ and not carried; a directory asset such as a DICOM series additionally carries a per-file digest list, and its own identifier is defined over names and hashes rather than concatenated bytes,

$$H_{\text{series}} \;=\; H\Big(\big\|_{k \in \text{sort}} \left(\text{name}_k \,\|\, \texttt{NUL} \,\|\, H_k \,\|\, \texttt{LF}\right)\Big),$$

so that renaming a slice changes the identifier, which is correct since slice ordering is clinical data, and verification costs nothing beyond the manifest. The consequence is that a recipient can prove, file by file, that a series they hold is the one a case was built from, without the container ever having held it.

3.4. How a new modality enters

G4 deserves a concrete answer, because "extensible" is easy to claim. Suppose a study wants to attach microbial load per periodontal site to a case. Nothing about that is geometry, and no existing dental format has a place for it. In UOS it requires four declarations and no change to the format:

1. An **asset** carrying the measurements, with its own hash, media type and the visit it belongs to.
2. A **frame**, here the canonical one since sites are named anatomically rather than positioned, so that the connectivity rule is satisfied.
3. An **extension** entry declaring the schema of what was added, listed in `extensions_used` so that a reader without it still opens the case and can say what it left unread.
4. A **descriptor** stating, per column, the unit, whether the value is measured or derived, and the vocabulary it uses, so that a consumer is never left inferring semantics from a column name.

The same four steps admit shade measurements, periodontal charting, or a time series from an intraoral sensor. What the format supplies is not a schema for each; it is the obligation to declare one, and the guarantee that a reader who does not know it degrades predictably instead of silently.

What this does not solve

Declaring a modality is not validating it. UOS makes microbial load carryable, addressable by anatomy, and traceable to whoever measured it. Whether that measurement means anything clinically is outside a container format, and a format that implied otherwise would be making exactly the category error it exists to prevent.

---

4. Mathematical formulation

This section states the model the project implements, because two of its results, the capacity ceiling of §5 and the calibration that lifts it, are consequences of the algebra rather than of any particular dataset.

Notation. The model borrows from two literatures whose conventions collide on one letter, and we keep both rather than renaming, so that a reader arriving from either finds the symbol where they expect it. Lowercase σ_i is the *density* of primitive i, the quantity that replaces opacity in the volumetric model. Uppercase Σ_i is its *covariance matrix*, and uppercase Σ without a subscript is an ordinary summation. Standing alone, as in a "3σ cutoff" or a separation of

"3.4–4.3 σ", σ is a *standard deviation* and has nothing to do with density. Where a formula uses more than one of these, the surrounding text says which.

Symbol	Meaning
$c_i,\ \Sigma_i$	Centre and covariance of Gaussian primitive i
$R_i,\ S_i$	Its rotation and scale factors, $\Sigma_i = R_i S_i S_i^\top R_i^\top$
α_i	Opacity of primitive i, photometric model only, $\alpha_i \in [0, 1]$
σ_i	Density of primitive i, volumetric model only, $\sigma_i \geq 0$
$\mu(x)$	Linear attenuation coefficient at x
$\tau(r)$	Line integral of μ along ray r; $\tau_{\max} \approx 9.21$ is its expressible ceiling
$k,\ k_\ell$	Global and per-layer calibration scalars applied to σ
T_{ij}	Rigid transform from frame i to frame j, $T_{ij} \in SE(3)$
$\varepsilon_{\mathrm{RMS}}$	Root-mean-square residual of a registration, in mm

4.1. The primitive

A scene is a set of N anisotropic Gaussians. Primitive i has a centre $c_i \in \mathbb{R}^3$ and a covariance factorised into rotation and scale so that it stays positive semi-definite under gradient descent,

$$\Sigma_i \;=\; R_i S_i S_i^\top R_i^\top, \qquad R_i \in SO(3), \quad S_i = \mathrm{diag}(s_{i1}, s_{i2}, s_{i3}),$$

with R_i parameterised by a unit quaternion and $s_{ij} > 0$. Its influence at a point is

$$G(x \mid c_i, \Sigma_i) \;=\; \exp\Big(-\tfrac{1}{2}(x - c_i)^\top \Sigma_i^{-1} (x - c_i) \Big).$$

Rendering truncates at a Mahalanobis radius of 3, the convention the Khronos extension fixes [4]. This is the "3σ" of the standard-deviation sense above, not a density: a primitive is treated as compactly supported on the ellipsoid $(x - c_i)^\top \Sigma_i^{-1} (x - c_i) \leq 9$.

4.2. Two different physics on the same primitive

Photometric splatting: appearance, the original formulation [2]. Each primitive carries an opacity $\alpha_i \in [0, 1]$ and a view-dependent colour in spherical harmonics. Pixels are formed by front-to-back alpha compositing over the depth-sorted primitives covering them,

$$C \;=\; \sum_{i=1}^{N} c_i \, \alpha_i \prod_{j<i} (1 - \alpha_j),$$

which is *order-dependent*: the same set of primitives composited in a different order gives a different pixel. The diffuse colour is recovered from the zeroth harmonic through the basis constant $Y_0 = 1/(2\sqrt{\pi}) \approx 0.2820947918$ and the training bias,

$$\mathrm{RGB} \;=\; Y_0 \cdot \mathrm{SH}_{0,0} \;+\; \tfrac{1}{2}.$$

Volumetric splatting: attenuation, the clinical primitive. X-ray attenuation is isotropic and unbounded, so opacity is replaced by a density $\sigma_i \geq 0$ and the harmonics are discarded entirely. The field models the linear attenuation coefficient

$$\mu(x) \;=\; \sum_{i=1}^{N} \sigma_i \, G(x \mid c_i, \Sigma_i),$$

and a detector value follows the Beer–Lambert law along the ray r,

$$\tau(r) \;=\; \int_r \mu(x)\,\mathrm{d}s, \qquad I \;=\; I_0\, e^{-\tau(r)}.$$

Because τ is an integral, this composition is **order-independent**: the same primitives in any order give the same value, which is a materially different object from alpha compositing, and it is why the container refuses to let the two share a vocabulary without a descriptor saying which is which [1].

4.3. Where the capacity ceiling comes from

The clinical field is fitted with a rasteriser built for the photometric model, and that borrowing has an algebraic price which we can state exactly.

Note that the field is expressed in σ but rendered through a pipeline whose numerics are defined on α; the ceiling below is exactly the cost of that translation. The rasteriser culls any primitive whose effective opacity falls below the 8-bit quantum, $\alpha_{\min} \approx 1/255$. A culled primitive contributes nothing *and receives no gradient*: it is dead without notice. At the other end, accumulated alpha saturates in `float32` at $\alpha_{\mathrm{acc}} = 0.9999$, which through $\tau = -\ln(1-\alpha_{\mathrm{acc}})$ caps the expressible line integral at

$$\tau_{\max} \;=\; -\ln(1-0.9999) \;\approx\; 9.21.$$

The usable dynamic range of the field is therefore the interval $[\,\alpha_{\min},\, \tau_{\max}\,]$, and it is set by the renderer's numerics, not by the anatomy. With σ derived directly from Hounsfield units, hundreds of primitives per ray saturate that ceiling: the model pins at **19 dB** with predictions capped at 0.087 against a ground truth of 1.54.

The correction, and why it is legitimate. A single global scalar $k > 0$ rescales the field, $\sigma_i' = k\,\sigma_i$, calibrated so that the mean of the rendered projection matches the mean of the digitally reconstructed radiograph. This is a choice of *units*, not of anatomy: it is a similarity on the range of μ and leaves the *shape* the CBCT measured untouched, so no structure is created or removed. With it, reconstruction goes from **19 dB to 43 dB**.

4.4. Additive decomposition into tissue layers

Because τ is a line integral, it is additive over any partition of the field. If $\{P_1, \ldots, P_L\}$ partitions the primitives disjointly and exhaustively, then

$$\tau(r) \;=\; \sum_{\ell=1}^{L} \tau_\ell(r), \qquad \tau_\ell(r) \;=\; \int_r \sum_{i \in P_\ell} \sigma_i\, G(x \mid c_i, \Sigma_i)\,\mathrm{d}s,$$

so the layers must reproduce the complete radiograph exactly. We verified this before comparing anything, measuring a maximum additivity error of 3.6×10^{-7} on normalised τ, which is what makes it legitimate to score a single field and a five-layer decomposition *on the same image*, the only honest comparison.

The consequence is not merely bookkeeping. Each layer ℓ receives its own calibration k_ℓ, so a layer trained against the radiograph of one tissue attenuates far less and survives the cull $\alpha_{\min}$ that killed it inside the global field. This is the mechanism behind the gain reported in §5, and it is measured rather than supposed: **16 times more live primitives** across five layers than in the single field, at a seeding budget which, given to one field, does nothing.

4.5. Appearance sampled onto a surface

Recovering a per-vertex colour from a fitted field, the operation behind the improved mesh, is a weighted mean over the primitives that actually cover a vertex v, weighted as the rasteriser weights them. For the $k = 32$ nearest primitives within the 3σ support,

$$w_i(v) \;=\; \alpha_i \cdot G(v \mid c_i, \Sigma_i), \qquad \mathrm{RGB}(v) \;=\; \frac{\sum_i w_i(v)\,\mathrm{RGB}_i}{\sum_i w_i(v)},$$

with a vertex marked unmeasured when $\sum_i w_i(v) = 0$. Using the nearest centre instead would give the colour *at* a point rather than the colour *seen* at it, which differs wherever primitives overlap: that is to say, everywhere.

4.6. Thresholding the crown boundary

Where a crown ends is ill-posed on geometry: a vertex on the gingival margin and one on the cervical neck are locally the same surface. It is well posed on colour. Otsu's method selects the threshold $t^\star$ on the CIELAB $a^\star$ channel maximising between-class variance,

$$t^\star = \arg\max_t \; \omega_0(t)\,\omega_1(t)\,\big(\mu_0(t) - \mu_1(t)\big)^2,$$

and we report the separation achieved as the standardised distance between the two class means, where σ_0 and σ_1 are the within-class standard deviations and not densities, $d = |\mu_0 - \mu_1| / \sqrt{(\sigma_0^2 + \sigma_1^2)/2}$, measured at 3.4–4.3 σ on real clinical photographs. No model, no training, no labels.

4.7. Reporting metric

Reconstruction quality is peak signal-to-noise ratio on held-out views, $\mathrm{PSNR} = 10 \log_{10}(M^2/\mathrm{MSE})$ with M the dynamic range. Every figure below is computed on views excluded from fitting (252 views per case, one in eight retained), and every comparison between models is made on the same image.

5. What Gaussian splatting contributes

3D Gaussian splatting [2] is fast, differentiable and web-renderable, and it offers a single substrate for things that were previously either meshes or voxels. We tested that appeal rather than assuming it, and the results below delimit where it pays.

5.1. Resolving power is set by the data

The first question was whether splatting could recover detail the acquisition did not capture. That is the informal hope that an optimised continuous representation "cleans up" a noisy surface or sharpens a radiograph.

We simulated a 9×9 mm patch of occlusal surface containing three findings chosen so that each exists on a different physical support: an occlusal fissure of 0.15 mm (geometry), a white spot lesion of 0.60 mm (surface colour), and a dentinal lesion of 1.50 mm under intact enamel (density only). Each modality was simulated through its physical chain: Gaussian PSF, sampling at its own pitch, then noise.

What we measure today: splatting reallocates capacity, it does not create resolution

A finding that exists only as density is invisible to a surface scanner at any reconstruction quality, and a fissure below the CBCT's sampling pitch is not recovered by fitting Gaussians to the CBCT. What splatting contributes is the ability to *allocate capacity where the data has it*, and that pays exactly when the modalities being combined have different resolving powers.

That condition holds in dentistry, where an intraoral scanner resolves surface detail an order of magnitude finer than a CBCT resolves internal structure, and it is why the representation is worth adopting here. The result is therefore not a limit on the method so much as a statement of what the method is for: composition across heterogeneous sources. Any sharpening claim has to come from adding a modality that carries the detail, not from re-fitting one that does not.

5.2. The capacity ceiling, and what moves it

Fitting a single field to a real CBCT, we varied the seeding budget and measured PSNR on held-out views.

Seeded primitives	Occupancy covered	Live primitives	PSNR (dB)
20 000	61.6 %	7 685	43.07
40 000	76.6 %	9 373	43.16
80 000	86.2 %	11 078	43.13
160 000	93.3 %	10 721	40.55
320 000	97.4 %	8 287	40.18
418 727	98.0 %	8 298	42.44

Twenty times the budget does not improve reconstruction; beyond roughly 10 000 live primitives the optimiser prunes back to a similar working set regardless of what it was given. §4.3 explains why: the usable range is $[\alpha_{\min}, \tau_{\max}] \approx [1/255, 9.21]$, fixed by the rasteriser's numerics, and neither more budget, nor finer resolution, nor spatial concentration widens it.

Where this points: the ceiling is in the renderer, not in the field

The bound is arithmetic and therefore addressable. Three routes follow directly from §4.3, in increasing order of effort.

(i) Calibrate, which we already do. A global scalar on σ moves 19 dB to 43 dB and costs one number. It is a change of units and leaves the measured shape intact.

(ii) Split the range, which we measure below. Several fields, each with its own k_ℓ, each spending the same narrow window on a narrower band of densities. Additivity makes this exact rather than approximate.

(iii) Integrate σ directly, which we have not built. Both routes above work *around* a rasteriser designed for opacity. A rasteriser that accumulates the line integral without an 8-bit cull and without `float32` alpha saturation, accumulating τ additively in higher precision or in log space, has no such window at all, and would make calibration a convenience rather than a necessity. This is a product decision rather than an implementation detail, and it is the single change that would most raise the ceiling for volumetric work.

5.3. Specialised fields: the measured gain

Splitting the volume into disjoint tissue layers, each with its own σ calibration, lifts the ceiling as §4 predicts.

Case	1 field (dB)	5 layers (dB)	Δ	Live, 1 field	Live, 5 layers
A	43.23	47.62	+4.39	11 013	176 865
B	42.02	44.10	+2.08	22 887	185 787
C	37.44	44.85	+7.41	13 672	133 628

With run-to-run noise of about 0.5 dB the gain is signal, and it replicates on all three volumes. It is not a parameter count: a dedicated field seeded at the *same density* as the global one, over the same region, scores 45.02 dB against 25.56 dB for the global field cropped to that region: 19 dB at equal primitive count. Specialisation, not size, is what pays, and the mechanism is

the one derived in §4: 16 times more primitives survive the cull once each layer carries its own calibration.

This also happens to be how a clinician would want to look at the twin (enamel, dentine, bone, separately), so the decomposition is a product feature rather than a fitting trick. It depends on segmentation for the layer names to mean anything; without that, partitioning by density band is the honest version, and it already delivers the gain.

6. Results

All figures below were measured on real clinical data. Where a figure comes from a single reference case, it is a maxillary arch with an intraoral scan, a CBCT, nine photographs and three reports, contributed by the collaborating practice and processed pseudonymised.

6.1. Composition and geometry

Measurement	**Value**	**What it establishes**
CBCT → scanner registration	`icp_surface`, $\varepsilon_{\mathrm{RMS}} = 0.666\,\mathrm{mm}$, unverified	That cross-modal alignment carries a usable residual, and that this one is provisional, which the container states rather than hides.
Mesh reversibility	$4.59 \times 10^{-6}\,\mathrm{mm}$ against a 0.1 mm budget	That the scan is recoverable from what the container carries. The residual is the `float32` of the STL itself.
Per-slice verification	397 slices, 419 entries, 0 errors, 0 warnings	That a DICOM series is verifiable file by file: a missing, altered or extra slice is identified individually rather than as "the series does not match".
Watertight arch export	12 025 closing faces, $11\,936\,\mathrm{mm}^3$, 3.2 s	That the composed scene yields a printable solid the scan alone does not provide.
Container overhead of uncompressed storage	183.2 MB stored vs. 37.3 MB deflated	That random access by HTTP range is paid for, and how much: the `.ply` layers compress to 13–15 %, the `.glb` only to 81 %.

6.2. Volumetric reconstruction

Measurement	**Value**	**What it establishes**
Calibration of σ	$19 \rightarrow 43\,\mathrm{dB}$	That the dominant error was a units mismatch against the rasteriser's window, not a modelling failure.
Additivity of disjoint layers	max. error 3.6×10^{-7} on normalised τ	That layer decompositions can be scored against the same image as the global field, which is what makes the comparison honest.

Measurement	Value	What it establishes
Tissue decomposition	+2.08 to +7.41 dB, three volumes	That specialisation lifts the ceiling reproducibly.
Same-density dedicated field	45.02 dB vs. 25.56 dB	That the gain is specialisation and not primitive count: 19 dB at equal seeding.
Live primitives after culling	16× more across five layers	The mechanism: each layer's own calibration keeps primitives above $\alpha_{\min}$.

6.3. Appearance measured from clinical photographs

An arch rendered from an intraoral scan is grey. The project measures colour per tooth from the patient's own photographs: the median of pixels per crown third (cervical, middle, incisal) taken from the photograph that best sees each crown, expressed in CIELAB.

On the reference case, 108 922 of 112 067 vertices (97.2 %) receive colour from their own tooth. Of the remainder, 97 inherit it from the nearest measured vertex and 3 041 are painted with a fallback gradient that is *not the patient's colour*, a distinction carried explicitly as a per-vertex `measured` flag, because otherwise "I have no colour here" and "I have the wrong colour here" are indistinguishable.

Comparable between teeth, not on an absolute scale

Flash falloff is removed using the patient's own gingiva as an internal reference (measured per-channel slopes 0.649 / 0.577 / 0.677), which makes crowns *comparable to each other*. The absolute level is not established, because a clinical photograph carries no grey reference. The container therefore supports "24 is darker than 23" and refuses to support "24 is an A3.5", and says which is which rather than shipping a number that looks like a shade match.

The route to the absolute scale is a capture-protocol change rather than an algorithmic one: a grey card in frame fixes the illuminant and turns the same pipeline into a calibrated measurement. This is the cheapest open item in the project.

6.4. Gingival boundary without a model

Otsu thresholding on the $a^\star$ channel of the clinical photographs the container already carries separates tooth from gingiva at **3.4–4.3** σ and traces the gingival scallop tooth by tooth. No model, no training, no labels: a threshold.

We report it here rather than among the segmentation work because of what it establishes jointly with §7: the reason geometric segmentation cannot find the crown boundary is that the boundary is not geometric, and the signal that does carry it is already inside the container.

7. Where the current results point

Three lines of work did not reach the outcome they were set up for. We report them with the same weight as the rest, because a format whose purpose is to keep measurement and inference apart earns nothing if its authors report only what worked, and because in each case the measurement identified the next step more precisely than the success would have.

7.1. Per-tooth segmentation on the intraoral scan

What we measure. The pipeline produces FDI codes for all 14 teeth of a maxillary arch *in the correct anatomical order*, and the boundaries are wrong. Eleven of the fourteen crowns can

be rejected on anatomy alone, comparing measured mesiodistal width against anatomical tables, without needing ground truth at all.

Why, and what follows. The boundary a crown ends at is not a geometric feature: across the gingival margin the surface is locally continuous, so no purely geometric criterion has the information to place it. That diagnosis is confirmed by the complementary measurement in §6: the same boundary is separable at 3.4–4.3 σ on the colour channel of photographs the container already carries. The route forward is therefore fusion rather than a better mesh segmenter: constrain the geometric label with the photometric boundary, which is measured, unlabelled and already present. Anatomical width tables give an independent rejection test that is available today and needs no model at all.

What it cost the format. Because tooth labels come from a model they belong to the inference plane, while semantic picking in a foreign viewer is defined over per-primitive metadata in the scene. The container therefore makes a declared compromise: the mesh is partitioned into one primitive per tooth so that any glTF viewer can select one, while the per-vertex labels, the ones with enough resolution to paint and to measure, stay in `derived/` and disappear when it is deleted [1].

7.2. A better classifier that made the real task worse

What we measure. The composite CBCT+scan pipeline delivers each tooth with its root, and half the pieces come out too long: 27–37 mm where a whole tooth measures 20–25. The excess is alveolar bone attached below the apex. The hypothesis was precision, since the first segmenter had recall 0.948 with moderate precision, so a more precise model should trim the bone. The model that won the benchmark lost on the real task, and the defect that motivated training it was unchanged.

Why, and what follows. Voxel-level precision and apex localisation are different objectives, and the benchmark rewards the first. Bone immediately below an apex is, to a voxel classifier, indistinguishable from root: same modality, adjacent densities, no boundary in the signal. The information that separates them is anatomical: roots terminate, alveolar crest has a shape, and both have population statistics. The next step is therefore a truncation criterion fitted on anatomy, evaluated on the quantity that actually matters (tooth length against tables), rather than a further classification run scored on voxel overlap. Choosing the evaluation metric to match the clinical quantity is the part that was wrong, and it is fixable without any new data.

7.3. Sharpness of the appearance field

What we measure. The appearance field reproduces its training views at 34–36 dB PSNR and still looks soft. The project's sharpness measure is the *footprint*, the major axis of the median Gaussian in μm, with a criterion of $< 563\,\mu$m fixed *before* running. Across seedings from 113 000 to 1 090 000 primitives the criterion is not met, and more primitives do not move it.

Why, and what follows. High PSNR with low sharpness is the signature of a fit that is faithful to its training views and limited by them: the optimiser has no gradient towards detail that the views do not resolve. Consistently with §5, adding primitives cannot manufacture that detail. Three routes follow, and they are ordered by cost rather than by promise: raise the resolving power of the input views, which is a capture change; regularise the footprint directly, adding a penalty on the median major axis so the optimiser trades reconstruction error against the quantity we actually care about; or seed at a density matched to the view resolution rather than to the mesh, so that capacity is not spent below the level the data can support. The first is the only one that adds information; the other two allocate existing information better, which is the same conclusion §5 reaches for the volumetric case.

Why these three are in the paper

Each was expected to work, and reporting only the results that did would describe a pipeline that does not exist. More practically, each is load-bearing for a design decision: the segmentation boundary is why the format has a removable inference plane at all, and the sharpness measurement is why appearance layers are declared *reconstructed* rather than *measured*. A criterion fixed before running and not met is a result; the alternative, renegotiating it afterwards, would have produced a paper with three more successes and nothing to act on.

8. Discussion

8.1. What becomes possible

- **Longitudinal comparison as subtraction.** Two visits in one frame turn "how much bone was lost in six months" into a difference between two containers of the same case, rather than a new study.
- **An auditable corpus.** A set of containers declares, per asset, what is measured and what is inferred and by which model at which weights hash, so a dataset can be filtered on those distinctions without opening the payloads.
- **An improved scan back out.** What the container returns is not the file that went in: the same geometry with per-vertex colour measured from the patient's own photographs and a per-vertex FDI code, exportable to the formats a laboratory opens.
- **Modalities that are not geometry.** Microbial load, shade or periodontal charting (§3.4): carried, addressable by anatomy, and declared.

8.2. On denoising and enhancement

The hope that motivated part of this work, that fitting a continuous field to a noisy radiograph or a rough surface would clean it, is not supported by our measurements, and we think the distinction is worth stating carefully because it is easy to demonstrate convincingly and wrongly.

A fitted field *does* produce visually smoother output than the voxels it was fitted to. That smoothness is the representation's prior, not recovered signal. Under a metric that rewards plausibility it looks like enhancement; under one that asks whether a 0.15 mm fissure below the sampling pitch is now visible, it is not. Our position is that the honest use of splatting here is composition and capacity allocation across modalities of differing resolution, and that any denoising claim requires a task-level evaluation against ground truth that this project has not done.

8.3. Regulatory posture

Nothing produced by a model is presented as clinical fact. Inference ships as `investigational`, isolated in a removable directory, at a declared regulatory layer, with the model and weights that produced it named. This is not a compliance claim; it is the minimum structure under which such a claim could later be made by someone who has done the clinical validation, and under which the same case can be distributed where the inference module is not cleared.

9. Limitations

- **One clinical case.** The composition figures come from a single real case. The Gaussian-field results come from three CBCT volumes of a licensed cohort that cannot be redistributed. Neither is a clinical study, and no claim here is validated on patient outcomes.
- **No distributable fixture.** The measured container holds a real patient's case and is not published, so an implementer cannot yet check a reader against anything but their own output.

- **Signals are unimplemented.** The asset class exists; no writer emits one, so the time-series path of §3.4 is designed but untested.
- **Signatures are unspecified.** What is missing is not code but the decision of which key signs and where it lives; signing with an invented key would produce a container that appears signed.
- **The media type is unregistered** and the format is not a standard. It is a proposal with a published schema and a validator, and it remains one until a second, independent implementation opens a container written by the first.

10. Future work

The work that follows is not a list of features we would like to build. It is what clinicians asked for once they saw the twin, plus what the measurements above say it would take to give it to them. We separate the two deliberately: a request is a requirement, and whether it is currently reachable is a measurement.

10.1. The measurement clinicians actually want

Validating whether a vertical gingiva-to-enamel dimension of the kind shown in commercial 2D demos is reproducible on a twin turned out to reveal that it is not one measurement but **three**, and that they differ in what they need and in what they are worth.

	Measurement	**Needs**	**Status**
1	Gingival margin → incisal edge	Scanner only	**Available.** Everything happens inside a single mesh at scanner accuracy, tens of μm, with no registration involved. This is the well-conditioned rung and it is implemented.
2	Cemento-enamel junction → alveolar crest	CBCT only	Blocked on locating the CEJ; see §10.2.
3	Gingival margin → alveolar crest	Scanner *and* CBCT	**The one with no 2D equivalent**, and therefore the one that justifies a fused twin at all. Blocked on registration accuracy.

Rung 3 is where the argument of this paper meets clinical practice. It cannot be taken from a radiograph, because the gingival margin is soft tissue and does not appear on one; and it cannot be taken from a scan, because the alveolar crest is under the gingiva. It exists only in the composition, which is precisely what the container is for.

And it does not yet work, in a way worth stating

Computed today across the two modalities, the margin-to-crest distance comes out **negative** on part of the arch: the crest is placed above the gingival margin, which is anatomically impossible. A negative distance is a useful failure, because it is self-evidently wrong rather than plausibly wrong, and it bounds the error: the registration residual and the crest extraction together exceed the quantity being measured. Dedicated per-tooth registration measures 0.45 mm and its positive control does not yet close.

The immediate work is therefore not a new measurement but the two things it rests on: tightening the alignment where the measurement is taken, and extracting the crest with a criterion that is anatomical rather than density-thresholded. Neither needs new data.

10.2. What is measurably out of reach, and what replaces it

Absolute gingival recession is defined against the cemento-enamel junction, and we measured that the CEJ is not recoverable from either of our modalities:

- **Not from CBCT by density.** Restorative metal dominates the neighbourhood, and cervical enamel is thinner than the voxel.
- **Not from CBCT by shape.** The detected point moves **6.6 mm** depending on the threshold chosen, which is not a measurement but a choice. This route was measured and discarded.
- **Not from the intraoral scan.** The cervical step lies below the triangulation noise of the mesh. This was checked on an arch *with* visible recession, so the failure is not an absence of the feature.

What is solid is the gingival margin itself, detected in 11 of 11 sections across the three arches tested. So the honest product is not absolute recession but **how far that margin has moved between two moments**, which is also what a digital twin exists to provide: not a photograph, a trajectory. Two visits in one frame make that a subtraction (§6); what remains is to establish the clinical threshold at which a displacement is reportable rather than noise.

10.3. Per-tooth motion, and a reference that does not move

Orthodontic follow-up asks a narrower question: did *this* tooth move, and by how much. Per-tooth registration gives a robustly small residual, but the displacement it yields is currently unusable, and the reason is instructive. Displacement is measured against the global arch registration, and that registration itself shifts with the ICP trimming parameter: the median displacement moves from **0.171 to 0.738 mm** on that hyperparameter alone. Every tooth appears to move because the frame moved.

Two consequences shape the work. First, the reference has to be built so that it cannot absorb the signal: a leave-one-out frame, estimated from every tooth except the one being measured, with a reportability threshold derived from the residual rather than assumed. Second, an intuitive shortcut is measurably wrong and should be recorded as such: estimating the transform for three teeth and averaging it for the rest is *worse than doing nothing* in 2 629 of 2 860 comparisons, against simply using the global registration. This matters because it is the kind of simplification a product roadmap adopts for performance reasons without measuring it.

10.4. The oral microbiome as a first-class modality

The clinical interest is not in the bacteria as a curiosity but in *which species are present and at what concentration*, because several are associated with conditions well outside the mouth, and because periodontal disease is the pathway through which that association is thought to act. A twin that carries geometry and density but cannot carry a microbial panel is, from that point of view, describing the least interesting half of the patient.

This is the concrete instance of G4, and it is the reason §3.4 was written as a mechanism rather than as a schema. The datum is a per-site quantity from a laboratory process, arriving days or weeks after the imaging, with its own units, its own detection limit and its own error model. Nothing about it is geometric, and everything about it needs to be attached to anatomy: a concentration means nothing without the site it was sampled at, and a site means nothing without the vocabulary that names it.

Three requirements follow, and only the first is already satisfied. It must be **addressable** by FDI site, which the container supports today. It must be **temporally located**, since a panel taken before treatment and one taken after are the entire point, which the visit mechanism supports but no writer yet exercises. And it must be **declared with its detection limit**, because a reported zero and a concentration below the assay's threshold are different facts, and

a format that stores them identically would repeat, on a new modality, exactly the confusion this project exists to avoid.

10.5. Occlusal force as a measured field over the surface

Clinicians also asked for bite force: not as a single number per patient but distributed over the occlusal surface, so that the arch can be shaded by how much load each region actually takes. Mechanically this is the same shape of problem as per-tooth colour (§6): a scalar quantity sampled at points and interpolated onto a surface the container already carries.

Two properties make it more delicate than colour, and both are worth fixing before any implementation.

- **A false-colour overlay must never be mistakable for measured appearance.** The container already distinguishes colour that came from the patient's photographs from colour that did not, per vertex. A force map painted onto the same surface has to declare itself as a mapped quantity with its own units and scale, not as an appearance layer, or a viewer will eventually show a load map to a clinician who reads it as a shade.
- **The colour map is part of the datum.** A field rendered through an unstated palette is not reproducible: the same measurements under a different scale tell a different clinical story. The scale, its bounds and whether it saturates belong in the layer descriptor alongside the units.

10.6. Audit: who did this, and how

The format already treats data as effectively immutable, traceable and pseudonymous. A case is never edited in place: a new version is written whose manifest points at the hash of the previous one, so the chain is verifiable by whoever receives it and without trusting the sender. What is *not* yet recorded is the clinical act.

Today the container answers "which software produced this and which model, at which version, with which weights" and "has a human verified this alignment". It does not answer the question a clinician actually asks when a case reaches them from elsewhere: **who performed this study, under what protocol, and when**. That is a different axis from software provenance, and it is the one that determines whether a second practitioner can interpret a finding or has to repeat the acquisition.

Why this cannot simply be a name field

The obvious implementation, a professional's name on the acquisition, collides with the property the format works hardest to preserve. A clinician is a person, their identity is personal data, and a container that pseudonymises the patient while naming the practitioner in clear has moved the problem rather than solved it. It is also a problem the format already has vocabulary for: `phi_state` declares what identifiable content a case carries, and it is currently a statement about the patient alone.

So the work is not a field but a design decision with three parts: an actor identifier that can be resolved by an institution and not by a recipient; a protocol reference that names how an acquisition was performed rather than who performed it, since that is what a second reader needs in order to reproduce it; and an extension of the PHI state to say whether professional identities are present, so that the same honesty applied to the patient applies to everybody in the record. Signatures (§9) are the natural carrier for the first, and are themselves unspecified for the same reason: the missing piece is a decision about keys and custody, not code.

10.7. Surgical planning, and why it needs the layers

Graft planning is the case where the tissue decomposition of §5 stops being a reconstruction result and becomes the deliverable: what a surgeon needs to see is bone volume available and

bone volume required, separated from the soft tissue that hides it and from the restorative metal that confounds it. The decomposition already produces those layers, and the container already carries them as separate, individually described Gaussian fields. What is missing is the clinical quantity computed on top of them, with its own error bar, and a validation that it agrees with what a surgeon measures intraoperatively. We consider this the most valuable open item and the one furthest from being closed.

10.8. Platform work

1. **A rasteriser that integrates σ directly**, without the 8-bit cull and the `float32` alpha saturation that fix the ceiling of §4.3. This is the single change that would most raise the ceiling for volumetric work, and it converts calibration from a necessity into a convenience.
2. **A distributable conformance fixture**, synthetic or built from an openly licensed dataset, so that interoperability can be tested rather than asserted.
3. **Anatomical root truncation**, following §7: fit the truncation criterion on anatomy and score it on tooth length rather than voxel overlap.
4. **Fusing the photometric boundary into the geometric segmentation**, using the colour separation of §6, which is already measured and already inside the container.
5. **A grey reference in the capture protocol**, which turns per-tooth colour from a comparative measurement into a calibrated one at no algorithmic cost.
6. **A non-geometric modality end to end**, microbial load or periodontal charting, to test §3.4 against a real dataset rather than a design.
7. **Ratification tracking.** `KHR_gaussian_splatting` is a Khronos Release Candidate; when it ratifies, the container's fallback path for external Gaussian layers retires.

The pattern across all of these

Every clinical item above is blocked on the same two things: the accuracy of a cross-modality alignment, and the ability to locate an anatomical landmark that no single modality resolves. Neither is a limitation of Gaussian splatting, and neither is solved by more primitives. They are the reasons the format records a residual on every registration and refuses to let an inferred landmark pass for a measured one, and they are where the next year of work belongs.

---

11. Conclusion

The problem in dental interoperability is not that the data cannot be encoded. It is that the *relations* between the encodings, spatial, temporal, semantic, and the relation between a datum and the process that produced it, are not written down anywhere that travels. We have described a container that obliges them to be written down, and reported what building it measured.

The representational finding is the one we would most want carried forward: Gaussian splatting is worth adopting in dentistry for capacity allocation across modalities of different resolution, and not for enhancement, because it adds no resolution and cannot. The photometric ceiling is real, budget does not move it, and specialised decomposition does, by up to 7.41 dB at equal primitive count.

And the format's central commitment is a discipline rather than a feature: two things that are not the same must never look the same. A registration nobody verified, a colour that is the patient's versus one from a fallback gradient, a boundary drawn by a model versus one drawn by a clinician. Each of those distinctions costs a field in a manifest and buys a recipient the ability to decide what to trust, which is the only thing that makes a case worth sending at all.

Acknowledgements

The author is grateful to ANFAIA and to HISTORA for their collaboration on this work.

The author thanks **Ismael Faro**, who founded ANFAIA and made this programme exist, for opening the door to work of this kind in Galicia at all, and specifically for proposing that Gaussian splatting be brought to bear on dental data in the first place. The idea that a representation developed for novel-view synthesis might serve as a common substrate for a surface scan and a volume is his, as is the practical framing that made it testable: fit it to real acquisitions, on real hardware, and measure what it does rather than what it promises.

Thanks are due also to **Matías Molinas**, CTO of HISTORA, for accompanying the project throughout and for the technical judgement that shaped several of the decisions defended in these pages, including more than one that turned out to be a decision to measure something rather than to argue about it.

The author is indebted to the two clinicians who founded HISTORA. **Dr. Pedro Guitián**, oral surgeon and founder of Centro Médico Odontológico Guitián, and **Dr. Elena López Alvar**, oral surgeon, set out what a clinician actually needs to measure and, just as usefully, what a plausible-looking number is worth when nobody can say how it was obtained.

Any errors, and every claim that turned out to be measurable rather than merely arguable, remain the author's.

References

[1] L. Garbayo Fernández. *UOS: Unified Oral Scene: Format Specification v0.2*, 2026. The normative document for the container described here: manifest, scenes, volume, provenance, extensions, the validation algorithm and implementation recipes.

[2] B. Kerbl, G. Kopanas, T. Leimkühler and G. Drettakis. *3D Gaussian Splatting for Real-Time Radiance Field Rendering*. ACM Transactions on Graphics, 42(4), SIGGRAPH 2023.

[3] Khronos Group. *glTF 2.0 Specification*. `https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.html`

[4] Khronos Group. *`KHR_gaussian_splatting`*, Release Candidate, 2026. `https://github.com/KhronosGroup/glTF/tree/main/extensions/2.0/Khronos/KHR_gaussian_splatting`

[5] NEMA. *DICOM Standard*, PS3.10 (media storage) and PS3.15 Annex E.1 (attribute confidentiality profiles).

[6] HL7. *FHIR Release 4* (4.0.1), 2019.

[7] ISO 3950:2016. *Dentistry — Designation system for teeth and areas of the oral cavity.*

[8] Apple. *USDZ File Format Specification*. `https://openusd.org/release/spec_usdz.html`