EXTENDED FORMAL SPECIFICATIONS OF 3D SPATIAL DATA TYPES

- TECHNICAL REPORT -

André Borrmann
Computation Civil Engineering
Technische Universität München

INTRODUCTION

Starting point for the development of a spatial query language is the formal definition of the semantics of the available spatial data types and the operators working on them. Such a system of spatial data types is also referred to as "spatial algebra". It captures the fundamental abstractions of spatial entities and their relationships.
Our spatial type system consists of the four types *Point*, *Line*, *Surface* and *Body*. The incorporation of types with lower dimensionality also allows for the utilization of the language in the context of dimensionally reduced models which are widely used in civil engineering. The definition of the spatial types is not limited to non-curved (plane) entities.
The data types are formally defined using point set theory and point set topology (Gaal, 1963). This methodology is well established thanks to the efforts of the GIS research community. Straight-forward application of pure point set theory would imply that all points of a dimensionally reduced entity, such as a point, a line or a surface, would belong to the boundary of this entity. Because this is not suitable for the specification of topological relationships, we take in general another approach here:
We define a dimensionally reduced geometric object (Line, Surface) as being composed of mappings from objects of lower dimensionality (1D-Intervall -> 3D-Line ; 2D-Region -> 3D-Surface) while preserving the concept of boundary / interior / exterior during this mapping. This finally leads to more powerful definitions that can be used for the specification of topological relationships.
The GIS research community distinguishes between simple and complex spatial types. Complex spatial data types can be understood as multi-component data types, i.e. a *Complex Point* can consist of an arbitrary number of points, a *Complex Line* can consist of several curves and a *Complex Body* may have a number of unconnected parts. Though simple data types reflect intuitive understanding, they do not incorporate the properties of a closed type system, where the geometric set operations *union*, *intersection*, and *difference* never result in an object outside of the type system. Therefore and because of the existence of bodies with holes and cavities within buildings, we decided to use complex spatial objects.
In the main, our definitions for spatial data types in 3D space are aligned to the model proposed by Schneider and Weinrich (2004). One exception is the denomination of the data types: instead of *point3D*, *line3D* and *volume* we use the terms *Point*, *Line*, *Surface* and *Body*, as proposed in (Zlatanova, 2000). Furthermore, the type *relief* is omitted, because it is not needed for the application domain considered here.
The topological notions of *boundary* (∂A)*, interior*(A°)*, exterior* (A^-) and *closure* ($\overline{A}$) are given for each spatial data type. They are required for the formal specification of topological relationships. Schneider and Weinrich provide both a structured and an unstructured definition for each type. Whereas the unstructured definition defines a type as a point set satisfying particular conditions, it provides no specification for the boundary, interior or exterior of this type. This is provided by the structured definition which models a type as

being composed of mappings from objects of lower dimensionality. At the same time, the structured definitions must not be more restrictive than the unstructured definition, i.e. all of the point sets classified as belonging to a certain type by its unstructured definition must also be constructible by means of the structured definition. Because specifications for boundary, interior and exterior of each type are required, we will provide only the structured definition and omit the unstructured one.

Schneider et al. enforce a *uniqueness constraint* in their structured definitions. It allows an entity to be composed in only one specific way, e.g. if a Line object can be rendered by a single curve it should not be possible to create the same Line by two curves. We consider the uniqueness constraint not necessary for our purposes and do not apply it in our definitions.

POINT

A value of type *Point* is defined as a finite set of isolated geometric points in the 3D space:

$$Point = \{P \subset \mathbb{R}^3 \mid P \text{ is finite}\}$$

By definition, a *Point* $P = \{p_1,\dots p_n\}$ has no *boundary, i.e.* $\partial P = \varnothing$, and all points belong to the *interior*, which is equal to the *closure:* $P = P^\circ = \overline{P}$. The *exterior* of P is $P^- = \mathbb{R}^3 - P$.

LINE

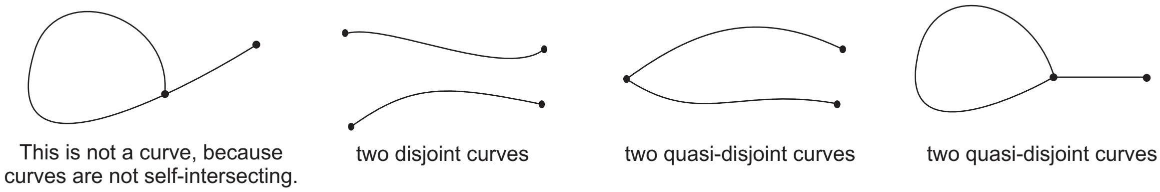
A *Line* is an arbitrary collection of 3D curves. It is defined as the union of the images of a finite number of continuous mappings from 1D to 3D space. In order to be able to define the boundary of a *Line* we have to investigate its components. A *Line* is composed of several *curves.* A *curve* results from a single mapping f_i and is by definition *non-self-intersecting.* It is defined as:

$$\begin{aligned} curve = \{ f([0,1]) \mid\ & (i)\ f:[0,1] \to \mathbb{R}^3 \text{ is a continous mapping } \wedge \\ & (ii)\ \forall a,b \in \left]0,1\right[: a \neq b \Rightarrow f(a) \neq f(b)\ \wedge \\ & (iii)\ \forall a \in \{0,1\}\ \forall b \in \left]0,1\right[: f(a) \neq f(b) \} \end{aligned}$$

The mappings $f(0)$ and $f(1)$ are called the *end points* of the curve. Condition (ii) avoids two interior points of the interval [0,1] being mapped to a single point of the curve, thereby prohibiting self-intersection and degenerated curves consisting of only one point. Condition (iii) allows loops $f(0) = f(1)$, but forbids equality of different interior points and equality of an interior point with an end point. If $f(0) = f(1)$, the curve forms a loop and hence has by definition no endpoints.

Two curves c_1, c_2 are called *quasi-disjoint*, if their interiors do not intersect. They are allowed to meet in the endpoints[1]. Formally written:

$$\begin{aligned} \text{quasi-disjoint}(c_1,c_2) \Leftrightarrow \forall a,b \in \left]0,1\right[:\ & f_1(a) \neq f_2(b) \wedge \\ & f_1(0) \neq f_2(b) \wedge f_1(1) \neq f_2(b) \wedge \\ & f_1(a) \neq f_2(0) \wedge f_1(a) \neq f_2(1) \end{aligned}$$

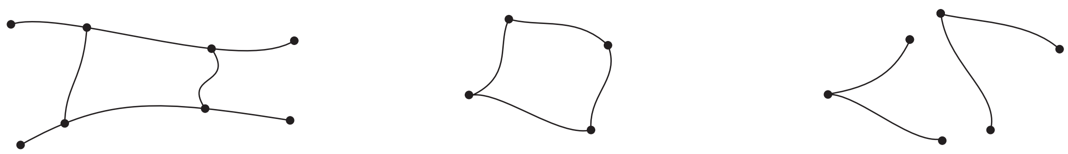


[1] Schneider et al. claim that two quasi-disjoint curves do not form loops in order to fulfill the uniqueness constraint.

The definition of *quasi-disjoint* curves is used to introduce the concept of a *block*[2] which specifies a connected component of a *Line*:

$$block = \Big\{ \bigcup_{i=1}^{m} c_i \mid (i)\;\; m \in \mathbb{N}, \forall 1 \le i \le m : c_i \in curve$$
$$(ii)\;\; \forall 1 \le i < j \le m : c_i \text{ and } c_j \text{ are quasi-disjoint}$$
$$(iii)\;\; m > 1 \Rightarrow \forall 1 \le i \le m : \exists 1 \le j \le m, i \ne j : c_i \text{ and } \mathrm{c}_j \text{ meet} \Big\}$$

Condition (ii) ensures that all curves are mutually quasi-disjoint. Thereby we avoid that an end point of one curve coincides with an interior point of another curve. Condition (iii) ensures that each curve is connected to at least one other curve.



a block consisting of 8 curves — a block consisting of 4 curves — one *Line* consisting of 2 blocks

A *Line* can now be defined as the union of a number of disjoint *blocks:*

$$Line = \Big\{ \bigcup_{i=1}^{m} b_i \mid \;(i)\;\;\; n \in \mathbb{N}, \forall 1 \le i \le m : b_i \in block$$
$$(ii)\;\;\; m \in \mathbb{N}, \forall 1 \le i \le j \le m : b_i \text{ and } b_i \text{ are disjoint} \;\Big\}$$

Condition (ii) ensures that a curve in one block can not intersect with any curve in another block.

We are now able to define the *interior*, *exterior*, *boundary* and *closure* of a *Line*. The *boundary* of a *Line* is the set of the end points of all quasi-disjoint curves it is composed of, minus those end points that are shared by several curves. These shared points belong to the *interior* of a *Line*.

The *closure* of a *Line L* is the set of all points of *L* including the end points. For the *interior* we obtain $L^\circ = \overline{L} - \partial L = L - \partial L$, and for the *exterior* we get $L^- = \mathbb{R}^3 – L$.

SURFACE

Since the definition of the *Surface* type is based on mappings of 2D regions to 3D space, the definition of the *Region2D* type as provided in (Schneider, 2006) is needed first. A *Region2D* is embedded into the two-dimensional Euclidean space $\mathbb{R}^2$ and modeled as a special infinite point set.

The concept of the *neighborhood* of a point is used to define the *interior*, *exterior* and *closure* of the *Region2D* type.

Assuming the existence of a Euclidean distance function in 2D

$$d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} \text{ with } d(p,q) = d((x_1, y_1),(x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

we define the *neighborhood* as follows:

Let $q \in \mathbb{R}^2$ and $\varepsilon \in \mathbb{R}^+$. The set $N_\varepsilon(q) = \{p \in \mathbb{R}^2 \mid d(p,q) \le \varepsilon\}$ is called the *(closed) neighborhood* of radius ε and center *q*.

Let $X \subseteq \mathbb{R}^2$ and $q \in \mathbb{R}^2$. *q* is an *interior point* of *X* if there is a neighborhood $N_\varepsilon(q)$ such that $N_\varepsilon(q) \subseteq X$. *q* is an *exterior point* of *X* if there is a neighborhood $N_\varepsilon(q)$ such that $N_\varepsilon(q) \cap X = \varnothing$. *q* is a *boundary point* of *X* if *q* is neither an interior nor an exterior point of X. *q* is a *closure point* of *X* if *q* is either an interior or a boundary point of *X*. The set of all *interior / exterior / boundary / closure points* of X is called the *interior / exterior / boundary / closure* of *X*.

[2] Schneider et al. claim that in a block an end point is either the end point of only one single curve or *by more than two* curves in order to fulfill the uniqueness constraint.

A single Region2D object. It may have holes and consist of several components.

Geometric anomalies that are excluded: Isolated or dangling point and lines feautures and missing point and line features.

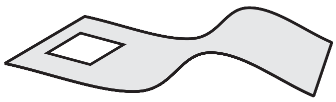
Schneider et al. use the notion of *regular closed* point sets for the definition of the *Region2D* type in order to avoid geometric anomalies: a set of points $X \subseteq \mathbb{R}^2$ is *regular closed* if, and only if, $X = \overline{X^{\circ}}$. The *interior* operation excludes point sets with dangling points, dangling lines and boundary parts. The *closure* operation excludes points sets containing cuts and punctures while re-establishing the boundary that was excluded by the *interior* operation.
Specifications for *bounded* and *connected* sets are another requirement for the definition of the Region2D type. Two point sets $X, Y \subseteq \mathbb{R}^2$ are *separated* if, and only if, $X \cap \overline{Y} = \emptyset = \overline{X} \cap Y$. A point set $X \subseteq \mathbb{R}^2$ is *connected* if, and only if, it is not the union of non-empty separated sets. Let $q = (x, y) \in \mathbb{R}^2$ and $\|q\| = \sqrt{x^2 + y^2}$. A set $X \subseteq \mathbb{R}^2$ is *bounded* if there is a number $r \in \mathbb{R}^+$ such that $\|q\| < r$ for all $q \in X$. The type *Region2D* can now be defined as:

$$Region2D = \{R \subseteq \mathbb{R}^2 \mid R \text{ is regular closed and bounded, the number of connected sets of } R \text{ is finite }\}$$

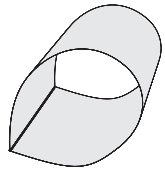
This definition models complex regions as possibly consisting of several components and possibly possessing holes.
Based on this definition, the Surface data type can accordingly be defined as the union of the images of a finite number of continuous mappings from a 2D region to 3D space. For a structured definition, we define the surface object as being composed of a number of so called superficies. A superficies is considered a non-self-intersecting surface component resulting from one single mapping s. It may possess holes.
Let *connected_Region2D* $\subset$ *Region2D* be all single-component 2D regions possibly possessing holes. Then the set of superficies is defined as:

$$\begin{aligned} superficies = \{s(R) \mid\ & (i)\ \ R \in connected\_region2D, \\ & \text{(ii)}\ \ s : R \to \mathbb{R}^3 \text{ is a continous mapping } \wedge \\ & \text{(iii)}\ \ \forall (x_1, y_1),\ (x_2, y_2) \in R^{\circ} : (x_1, y_1) \neq (x_2, y_2) \Rightarrow s((x_1, y_1)) \neq s((x_2, y_2))\ \wedge \\ & \text{(iv)}\ \ \forall (x_1, y_1) \in \partial R\ \ \forall (x_1, y_1) \in R^{\circ} : s((x_1, y_1)) \neq s((x_2, y_2))\} \end{aligned}$$

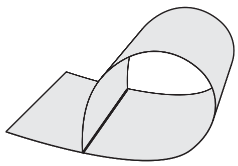
Condition (iii) avoids two interior points of *R* being mapped to a single point in the superficies, thereby prohibiting self-intersection and degenerated superficies consisting of only one point or being a line. Condition (iv) forbids boundary points of *R* being mapped to the same point as an interior point.



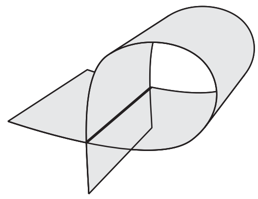
This is a superficies.



This is a partially closed superficies.
The "seam" belongs to the interior.



This is not a superficies, because boundary points of the underlying Region2D are mapped to points that coincide with points that result from mappings of interior points.



This is not a superficies, because condition (i) is violated: different interior points of the underlying Region2D are mapped to the same points in IR³.

The definition allows superficies to be closed or partially closed. A superficies S is *closed (partially closed)* if sets B_1 and B_2 exist with $\partial R = B_1 \cup B_2$ ($\partial R \supset B_1 \cup B_2$) and $B_1 \cap B_2 = \varnothing$, so that $s(B_1) = s(B_2)$. In this case, the set $I(s)=s(B_1)$ is part of the interior of S; otherwise $I(S) = \varnothing$.

Given a superficies S, its *boundary* is $\partial S = s(\partial R) - I(S)$ and its *interior* is $S^\circ = s(R^\circ) \cup I(S)$. Hence, its closure is $\overline{S} = \partial(S) \cup S^\circ$, and its *exterior* is $S^- = \mathbb{R}^3 - \overline{S}$.

Let T be the set of all superficies over $\mathbb{R}^3$. Two superficies $S_1, S_2 \in T$ are called *quasi-disjoint* if their interiors do not intersect. They may share a common boundary. Formally written:

$$\text{quasi-disjoint}(S_1, S_2) \Leftrightarrow s_1(R_1^\circ) \cap s_2(R_2^\circ) = \varnothing \wedge$$
$$s_1(\partial R_1) \cap s_2(R_2^\circ) = \varnothing \wedge s_1(R_1^\circ) \cap s_2(\partial R_2) = \varnothing$$

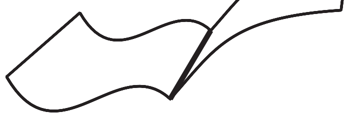
The definition allows that in case one or both of the superficies is (partially) closed they may meet at the "neat". Two superficies *meet* if they are quasi-disjoint and $s_1(\partial R_1) \cap s_2(\partial R_2) \neq \varnothing$.

We use the definition of *quasi-disjoint* superficies to introduce the concept of a *superficies block*, which is a connected component of a surface. A superficies block is defined[3] as

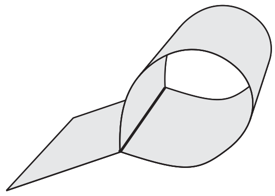
$$superficies\_block = \Big\{ \bigcup_{i=1}^{m} S_i \mid (i)\ m \in \mathbb{N}, \forall 1 \le i \le m : S_i \in T \ \wedge$$
$$\text{(ii) } \forall 1 \le i < j \le m : S_i \text{ and } S_j \text{ are quasi-disjoint } \wedge$$
$$\text{(iii) } m > 1 \Rightarrow \forall 1 \le i \le m : \exists 1 \le j \le m, i \neq j : S_i \text{ and } S_j \text{ meet} \Big\}$$

Condition (ii) ensures that all superficies are mutually quasi-disjoint. Thereby we avoid that the boundary of one superficies coincides with the interior of another superficies, except for the seam of a (partially) closes superficies. Condition (iii) ensures that each superficies is connected to at least one other superficies.

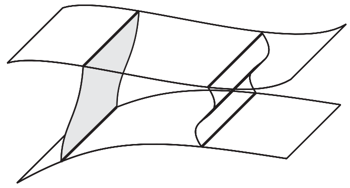
[3] Schneider et al. claim that et al. claim that in a surface block, a boundary curve belongs either to exactly one superficies or is shared or *by more than two* superficies in order to fulfill the uniqueness constraint.



two quasi-disjoint surfaces



a superficies block consisting of two superficies. superficies may meet at the 'seam' of a closed superficies,



a Surface consisting of 8 superficies

Two surface blocks $c_1, c_2 \in superficies\_block$ are disjoint if, and only if, $c_1 \cap c_2 = \varnothing$. We are now able to provide a structured definition for the spatial data type *Surface*:

$$Surface = \left\{ \bigcup_{i=1}^{n} c_i \mid (i)\ n \in N, \forall 1 \le i \le n : c_i \in superficies\_block \right.$$

$$\left. (ii)\ \forall 1 \le i \le j \le n : c_i \text{ and } c_j \text{ are disjoint} \right\}$$

Condition (ii) ensures that superficies in one block may not intersect with superficies in any other block.

In order to define the boundary of a *Surface*, we use the auxiliary set *L(S)*. Let *S* be a surface object with superficies $S_1, \ldots, S_m$, and let *L(S)* be the set of all curves that form the boundary of more than one superficies[4]:

$$L(S) = \left\{ \bigcup_{i=1}^{m} l_i \mid\ m \in \mathbb{N}, \forall 1 \le i \le m : l \in curve\ \wedge \right.$$

$$\left. l \subseteq \bigcup_{i=1}^{m} \partial(S_i) \wedge card\left(\{ S_i \mid (1 \le i \le m \wedge l \subseteq \partial S_i \} \right) \ge 2 \right\}$$

Then the *boundary* of S is $\partial S = \bigcup_{i=1}^{m} \partial S_i - L(S)$, and the *interior* of S is $S^\circ = \bigcup_{i=1}^{m} \underline{S_i}^\circ \cup L(S)$. Hence, the *closure* of S is $\overline{S} = \partial S \cup S^\circ$, and the *exterior* of S is $S^- = \mathbb{R}^3 - \overline{S}$.

BODY

Bodies are embedded into the three-dimensional Euclidean space $\mathbb{R}^3$ and modeled as special infinite point sets. It is possible for a *Body* to consist of several components and it may possess cavities. The notion of the *neighborhood* of a point is employed to define the *interior*, *exterior* and *closure* of the *Body* type, the definition of which corresponds to that given above for 2D space.

Let $X \in \mathbb{R}^3$ and $q \in \mathbb{R}^3$. q is an *interior point* of X if there is a neighborhood $N_\varepsilon(q)$ such that $N_\varepsilon(q) \subseteq X$. q is an *exterior point* of X if there is a neighborhood $N_\varepsilon(q)$ such that $N_\varepsilon(q) \cap X = \varnothing$. q is a *boundary point* of X if q is neither an interior nor an exterior point of X. q is a closure point of X if q is either an interior or a boundary point of X. The set of all *interior / exterior / boundary / closure* points of X is called the *interior / exterior / boundary / closure* of X.

For similar reasons as for the definition of *Region2D*, the definition of the type *Body* is based on the notion of *regular closed point sets*:

A set of points $X \subseteq \mathbb{R}^3$ is *regular closed* if, and only if, $X = \overline{X^\circ}$.

Here, the *interior* operation excludes point sets containing isolated or dangling point, line, and surface features. The *closure* operation excludes point sets with punctures, cuts or stripes and reestablishes the boundary that was removed by the *interior* operation. By definition, closed neighborhoods are regular closed sets.

Another essential is a specification for *bounded* and *connected* sets in 3D. Two point sets $X, Y \subseteq \mathbb{R}^3$ are *separated* if, and only if, $X \cap \overline{Y} = \varnothing = \overline{X} \cap Y$. A point set $X \subseteq \mathbb{R}^3$ is *connected*

---

[4] Schneider et al. use a completely different definition for L(S).

if, and only if, it is not the union of non-empty separated sets. Let $q=(x, y, z)\in\mathbb{R}^3$, and $\|q\| = \sqrt{x^2+y^2+z^2}$. A set $X\subseteq\mathbb{R}^3$ is *bounded* if there is a number $r\in\mathbb{R}^+$ such that $\|q\| < r$ for all $q\in X$. The spatial data type *Body* can now be defined as

$$Body = \{B \subset \mathbb{R}^3 \mid B \text{ is regular closed and bounded, the number of connected sets of B is finite } \}$$

CORRESPONDING CONVEX OBJECT

The specification of topological predicates that can be used to query relationships of concave objects, such as *surround* and *encompass*, is based on the concept of "corresponding convex objects".

First, a definition of convexity is needed. In general, a point set in Euclidean space $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$ is convex if it contains all the line segments connecting any pair of its points. In order to gain a more powerful concept of convexity for dimensionally reduced entities, we use a slightly different approach here. We define an object as being convex, if it is convex in it is original domain space, i.e. that of a *Line* in $\mathbb{R}$, that of a *Surface* in $\mathbb{R}^2$ and that of a Body in $\mathbb{R}^3$.

The operation *convex* returns the convex object corresponding to a given object. The corresponding convex object is defined as follows:

Body

The convex *Body* B_{convex} corresponding to a *Body B* is defined as the union of the all convex sets corresponding to the connected sets in *B*.

For each connected set B_i in *B* there is a corresponding convex set that is the smallest set that

(i) contains all points of B and

(ii) is a convex set in $\mathbb{R}^3$.

Surface

The definition of the convex *Surface* S_{convex} corresponding to the *Surface S* is based on the definition of the convex set corresponding to the *Region2D* object underlying the *Surface S*.

Let *R* be a Region2D object. For each connected set R_i in *R* there is a corresponding convex set that is the smallest set that

(i) contains all points of R_i and

(ii) is a convex set in $\mathbb{R}^2$.

The convex *Region* R_{convex} corresponding to a *Region2D R* is defined as the union of all convex sets corresponding to the connected sets in *R*.

The convex *Surface* S_{convex} corresponding to the *Surface S* results from the convex *Region2D* R_{convex} by applying the same mappings f_i to R_{convex} as applied to *R* to yield *S*.

Line

A Line is the union of mappings of the interval [0;1] (1D space) into 3D space. The interval [0;1] is convex in $\mathbb{R}$. Therefore, each Line is convex.

REFERENCES

Clementini, E., and Felice, P. D. (1996). "A model for representing topological relationships between complex geometric features in spatial databases." *Information Sciences*, 90(1-4), 121-136.

Egenhofer, M., and Franzosa, R. (1991). "Pointset topological spatial relations." *International Journal of Geographical Information Systems*, 5(2), 161-174.

Gaal, S. (1963). *Point Set Topology.* Academic Press, New York, USA.

Melton, J. (2003). *Advanced SQL:1999. Understanding Object-Relational and Other Advanced Features*, Morgan Kaufmann, San Francisco, USA.

Open Geospatial Consortium (OGC) (2005). “OpenGIS Implementation Specification for Geographic information”. Document 06-126.

Schneider, M., and Weinrich, B. E. (2004). "An abstract model of three-dimensional spatial data types." *12th ACM Intern. Worksh. on Geographic Information Systems*. Washington, DC, USA.

Schneider, M., and Behr, T. (2006). "Topological relationships between complex spatial objects." *ACM Transactions on Database Systems*. In Press.

Zlatanova, S. (2000). "On 3D Topological Relationships." *11th Int. Workshop on Database and Expert Systems Applications*, Greenwich, London, UK.