DaTra: Data Transformation Semantics as a Symmetric Monoidal Category on Atlas Presheaves

@qualiascript ChatGPT

Abstract

This work formalizes the semantics of the Datra programming-language project in Lean. Data domains are organized into atlases: structured, eventually stable families of pages encoding how data is partitioned and related. Data transformations are modeled as presheaves on the resulting atlas categories, yielding a topos-theoretic semantics. Stable transformations are then assembled over atlas federations, whose horizontal sum extends by Day convolution to a symmetric monoidal category.

1 Dominions

Definition 1.1 (The Category of Dominions)**.** The **Category of Dominions**, denoted Dom, has as its objects sets whose cardinality is at most ω_0, equivalently sets admitting an injection into ω_0, and as its morphisms set-theoretic functions.

Definition 1.2 (The Domanial Embedding Functor)**.** The **Domanial Embedding Functor**, denoted $\mathsf{Emb} : \mathsf{Dom} \to \mathsf{Set}$, sends each dominion to its underlying set.

2 Domanial Insertions

Definition 2.1 (The Category of Domanial Insertions)**.** The **Category of Domanial Insertions**, denoted DomIns, has dominions as objects and monomorphisms in Dom as morphisms.

Definition 2.2 (The Category of Codomanial Insertions)**.** The **Category of Codomanial Insertions**, denoted $\mathsf{CoDomIns}$, is the opposite category $\mathsf{DomIns}^{\mathrm{op}}$.

3 Chains

Definition 3.1 (Chain)**.** A **Chain** is a totally ordered thin category, indexed skeletally by an ordinal strictly smaller than $\omega_0^{\omega_0}$.

Definition 3.2 (The Spine)**.** The **spine** is the chain with ω_0 objects.

4 Consolidations

Definition 4.1 (The Category of Consolidations)**.** The **Category of Consolidations**, denoted Con, has chains as objects and point-surjective functors between their thin categories as morphisms.

Definition 4.2 (The Category of Coconsolidations)**.** The **Category of Coconsolidations**, denoted CoCon, is the opposite category $\mathsf{Con}^{\mathrm{op}}$.

Definition 4.3 (Consolidation Transport)**.** The **Consolidation Transport Functor** is denoted ConTra. It is a functor from Con to Set that sends a chain to its object set and a consolidation to its underlying set-theoretic surjection.

5 Folios and Paginations

Definition 5.1 (Folio)**.** A **folio** is a functor $F : S \to \mathsf{CoCon}$, where S is the spine, such that $F(0)$ is the singleton chain.

Definition 5.2 (Page Elements)**.** For a folio F, its **category of page elements**, denoted $\mathrm{El}(F)$, has as objects pairs (m, k), where m is a page and k is an object of the chain $F(m)$. An arrow $(m, k) \to (n, l)$ is an arrow $f : m \to n$ in the opposite spine for which the consolidation induced by F transports k exactly to l. Equivalently,

$$\mathrm{El}(F) = \int \big(\mathsf{ConTra} \circ F^{\mathrm{op}}\big).$$

Definition 5.3 (The Category of Paginations)**.** The **Category of Paginations**, denoted Pag, has as objects a folio Fo together with its associated category of page elements $El = \mathrm{El}(Fo)$. For $W : \mathsf{Pag}$, write W_{Fo} for its folio and W_{El} for its page-element category. A morphism $T : X \to Y$ in Pag is a functor $T : X_{El} \to Y_{El}$.

For illustration, let $X_H := \mathsf{ConTra} \circ X_{Fo}^{\mathrm{op}}$ and define Y_H likewise. If an arrow f of X_{El} has source $x = (m, i)$, then T acts schematically as follows:

$$\begin{tikzcd} x = (m,i) \arrow[r, "f"] \arrow[d, "T"'] & (n, X_H(f)(i)) \arrow[d, "T"] \\ (m', i') \arrow[r, "T(f)"'] & (n', Y_H(T(f))(i')). \end{tikzcd}$$

6 Atlases

Definition 6.1 (The Category of Atlases)**.** The **Category of Atlases**, denoted Atl, has as objects pairs (Pa, Da), where $Pa : \mathsf{Pag}$ and $Da : Pa_{El} \to \mathsf{DomIns}$ is a functor, equivalently a presheaf $Da : Pa_{El}^{\mathrm{op}} \to \mathsf{CoDomIns}$. For the folio Pa_{Fo}, a page m, and distinct objects k, k' of the chain $Pa_{Fo}(m)$, the pullback in Da of the arrows induced by $(m \to 0)(k)$ and $(m \to 0)(k')$ is empty.

For $X : \mathsf{Atl}$, write X_{Pa} and X_{Da} for its two components, and abbreviate $X_{Fo} := (X_{Pa})_{Fo}$ and $X_{El} := (X_{Pa})_{El}$.

For every $n \in \mathbb{N}$, the **nth atlas page** of X is the chain $X_{Fo}(n)$ on its full spine. An object k of this chain is a **page cell**, written (n, k), and the dominion carried by that cell is $X_{Da}(n, k)$.

There is an integer w such that, for every $w' > w$, the consolidation from page w' to page w induced by Pa_{Fo} is the identity and, for every object k of $Pa_{Fo}(w)$, the corresponding arrow $Da((w' \to w)(k))$ is the identity. The least such integer is the **cardinality of the atlas**, denoted $|A|$.

A morphism $T : X \to Y$ is a pair (T_{Pa}, T_{Da}), where $T_{Pa} : X_{El} \to Y_{El}$ is a functor and $T_{Da} : X_{Da} \Rightarrow Y_{Da} \circ T_{Pa}$ is a natural transformation. Thus, for every $f : x \to y$ in X_{El}, the following square commutes:

$$\begin{tikzcd} X_{Da}(x) \arrow[r, "X_{Da}(f)"] \arrow[d, "(T_{Da})_x"'] & X_{Da}(y) \arrow[d, "(T_{Da})_y"] \\ Y_{Da}(T_{Pa}(x)) \arrow[r, "Y_{Da}(T_{Pa}(f))"'] & Y_{Da}(T_{Pa}(y)). \end{tikzcd}$$

Finally, if $x = |X|$, $x' > x$, and r is an object of the final chain $X_{Fo}(|X| - 1)$, then $T_{Pa}(x', r) = T_{Pa}(x, r)$ and $(T_{Da})_{(x',r)} = (T_{Da})_{(x,r)}$.

Definition 6.2 (Extent of an Atlas)**.** The **extent** of A is $\mathsf{Ex}(A) = A_{Da}(0,0)$. Since objects of DomIns are dominions, $\mathsf{Ex}(A) : \mathsf{Dom}$ for every $A : \mathsf{Atl}$.

Definition 6.3 (Territory of an Atlas)**.** Let M be the objects of the final chain $A_{Fo}(|A|-1)$. The **territory** is the indexed family $\mathsf{Ter}(A) : M \to \mathsf{Dom}$ given by $\mathsf{Ter}(A)(k) = A_{Da}(|A|-1, k)$.

Definition 6.4 (nth Region of an Atlas)**.** The n**th region** of A is $\mathsf{Ter}(A)(n)$, with indexing beginning at 0.

7 Transposals and Traversals

Definition 7.1 (The Category of Atlas Transposals)**.** The **Category of Atlas Transposals**, denoted $\mathsf{AtlTrap}$, is the wide subcategory of Atl whose morphisms $F : X \to Y$ have point-injective object maps F_{Pa}.

Definition 7.2 (The Category of Atlas Traversals)**.** The **Category of Atlas Traversals**, denoted $\mathsf{AtlTrav}$, is the wide subcategory of $\mathsf{AtlTrap}$ whose morphisms satisfy the following conditions. For $F : X \to Y$, let $x = (m, i)$ and $y = (m', j)$, put $p = \min(m, m')$, write $F_{Pa}(x) = (n, i')$ and $F_{Pa}(y) = (n', j')$, and put $p' = \min(n, n')$. Let $c^X_{m,p}$ and $c^X_{m',p}$ denote the consolidations induced by the folio X_{Fo}, and define the analogous maps for Y. If

$$c^X_{m,p}(i) < c^X_{m',p}(j),$$

then

$$c^Y_{n,p'}(i') < c^Y_{n',p'}(j').$$

Furthermore, if $q \in |X|$ and $t \in X_{Da}(q)$ is in the image of a final region—that is, there exist $k \in |\mathsf{Ter}(X)|$ and $l \in \mathsf{Ter}(X)(k)$ whose image under $X_{Da}((|X|-1) \to q)(k)$ is t—then this property is preserved by F.

Definition 7.3 (The Category of Stable Atlas Traversals)**.** The **Category of Stable Atlas Traversals**, denoted $\mathsf{StaAtlTrav}$, is the wide subcategory of $\mathsf{AtlTrav}$ whose morphisms $F : X \to Y$ satisfy $F_{Pa}(0,0) = (0,0)$.

8 Maps and Cartography

Definition 8.1 (The Category of Atlas Maps)**.** The **Category of Atlas Maps**, denoted AtlMap, is the full subcategory of atlases for which every element of the extent is in the image of a final region.

Definition 8.2 (The Atlas Map Inclusion Functor)**.** The **Atlas Map Inclusion Functor** $\mathsf{AtlMapInc} : \mathsf{AtlMap} \to \mathsf{Atl}$ is the canonical inclusion.

Definition 8.3 (The Category of Atlas Traversal Maps)**.** The **Category of Atlas Traversal Maps**, denoted $\mathsf{AtlTravMap}$, is the full subcategory of $\mathsf{AtlTrav}$ whose objects are Atlas Maps.

Definition 8.4 (The Atlas Traversal Map Inclusion Functor)**.** The canonical inclusion is denoted $\mathsf{AtlTravMapInc} : \mathsf{AtlTravMap} \to \mathsf{AtlTrav}$.

Lemma 8.5 (Cartography Lemma)**.** *The inclusion* $\mathsf{AtlTravMapInc}$ *has a right adjoint, denoted the* ***Charting Functor*** $\mathsf{Chr} : \mathsf{AtlTrav} \to \mathsf{AtlTravMap}$*.*

Proof sketch. *On an object* $X : \mathsf{AtlTrav}$*, the functor* Chr *sends* X *to the universal arrow from* $\mathsf{AtlTravMapInc}$ *to* X*. Write* $X' = \mathsf{AtlTravMapInc}(\mathsf{Chr}(X))$*, together with* $H : X' \to X$*, with* X' *terminal among objects having this property. The solution is the DaTra map satisfying* $\mathsf{Ter}(X') \cong \mathsf{Ter}(X)$*. It is unique because* $\mathsf{AtlTrav}$ *preserves orders and terminal because* H_{Pa} *is faithful.*

9 Coalitions

Definition 9.1 (The Coalizing Functor)**.** The **Coalizing Functor**, denoted

$$\mathsf{Coa} : \mathsf{StaAtlTrav} \longrightarrow \mathsf{DomIns},$$

is the extent of the charted atlas: $\mathsf{Coa} = \mathsf{Ex} \circ \mathsf{Chr}$.

Definition 9.2 (The Coalition of an Atlas)**.** The **coalition** of an atlas X is the extent of its chart, regarding X as an object of the wide category of stable traversals.

Definition 9.3 (The Domanial Inclusion Functor)**.** The **Domanial Inclusion Functor** sends a dominion to the atlas of cardinality 1 whose coalition is that dominion. Write $\mathsf{DomInc} : \mathsf{DomIns} \to \mathsf{StaAtlTrav}$ for this functor. It is left adjoint to Coa: for $X : \mathsf{DomIns}$ and $Y : \mathsf{StaAtlTrav}$, naturally in X and Y,

$$\mathrm{Hom}_{\mathsf{StaAtlTrav}}(\mathsf{DomInc}(X), Y) \cong \mathrm{Hom}_{\mathsf{DomIns}}(X, \mathsf{Coa}(Y)),$$

and $\mathsf{Coa}(\mathsf{DomInc}(X)) \cong X$.

10 Data Transformations

Definition 10.1 (The Category of Data Transformations)**.** The **Category of Data Transformations**, also called the category of **Data Transformation Sets** or **DaTra Sets**, is the presheaf category

$$\mathsf{DaTra} = [\mathsf{Atl}^{\mathrm{op}}, \mathsf{Set}].$$

As a presheaf category, it is a topos.

Definition 10.2 (Navigation)**.** A **navigation** of $D : \mathsf{DaTra}$ is a monomorphism $\mathsf{Nav} : \mathsf{Yo}(A) \to D$ for some atlas A.

Definition 10.3 (Expedition)**.** An **expedition** is a navigation represented by an Atlas Map.

Definition 10.4 (The Category of Data Transposals)**.** The **Category of Data Transposals**, denoted DaTrap, is the presheaf category $[\mathsf{AtlTrap}^{\mathrm{op}}, \mathsf{Set}]$.

Definition 10.5 (The Category of Data Traversals)**.** The **Category of Data Traversals**, denoted DaTrav, is the presheaf category $[\mathsf{AtlTrav}^{\mathrm{op}}, \mathsf{Set}]$.

Definition 10.6 (Data Transformation Maps)**.** The **Category of Data Transformation Maps**, denoted $\mathsf{DaTraMap}$, is the full subcategory of DaTra Sets all of whose navigations are expeditions.

11 Atlas Federations

Definition 11.1 (Atlas Merge)**.** The **Atlas Merge** is the object operation

$$\mathsf{AtlMerge} : \mathrm{Ob}(\mathsf{Atl}) \times \mathrm{Ob}(\mathsf{Atl}) \longrightarrow \mathrm{Ob}(\mathsf{Atl}).$$

For atlases X and Y, the extent of $\mathsf{AtlMerge}(X, Y)$ is the disjoint union of their extents. Both atlases are evaluated on their full spines: page $n+1$ of the merge is the tagged, componentwise combination of page n of X and page n of Y.

Definition 11.2 (Atlas Federation)**.** An **Atlas Federation** is a countable tagged family of atlas objects. The category of Atlas Federations is denoted AtlFed. A morphism consists of a map of tags and, at each source tag, a stable atlas traversal to its selected target tag.

Definition 11.3 (The Empty Atlas)**.** The **Empty Atlas** is $\mathsf{AtlI} = \mathsf{DomInc}(\varnothing)$, also viewed as an Atlas Federation.

Definition 11.4 (The Atlas Horizontal Sum Bifunctor)**.** The **Atlas Horizontal Sum Bifunctor**, denoted $\mathsf{AtlHorSum} : \mathsf{AtlFed} \times \mathsf{AtlFed} \to \mathsf{AtlFed}$, has object action

$$\mathsf{AtlHorSum}(X, X') = \mathsf{AtlMerge}(X, X').$$

Here AtlFed is the category of Atlas Federations; its empty federation represents AtlI. Given stable traversals $F : X \to Y$ and $F' : X' \to Y'$, horizontal sum preserves the left and right tags and applies F and F' componentwise. Stability fixes the common origin, so this arrow action is again a stable atlas traversal.

Lemma 11.5 (Horizontal Lemma)**.** *The Atlas Horizontal Sum Bifunctor* $\mathsf{AtlHorSum}$ *exists.*

Proof. *Identity arrows act identically on every tag, composition holds componentwise, and every component arrow is required to lie in* $\mathsf{StaAtlTrav}$*. The federation structure supplies the associator, unitors, and braider by reassociation, deletion of the empty tag family, and tag swapping, respectively.*

Definition 11.6 (The Atlas Braider)**.** The **Atlas Braider** $\mathsf{AtlBrd}_{F,F'} : \mathsf{AtlHorSum}(F, F') \to \mathsf{AtlHorSum}(F', F)$ swaps the left and right component tags. Hence $\mathsf{AtlBrd}_{F,F'} \circ \mathsf{AtlBrd}_{F',F} = \mathrm{Id}$.

Definition 11.7 (The Atlas Associator)**.** The **Atlas Associator**, denoted

$$\mathsf{AtlAsoc}_{F,F',F''} : \mathsf{AtlHorSum}(\mathsf{AtlHorSum}(F, F'), F'') \longrightarrow \mathsf{AtlHorSum}(F, \mathsf{AtlHorSum}(F', F'')),$$

flattens the nested federation and retags its three sides using the canonical reassociation $((x + y) + z) \leftrightarrow (x + (y + z))$. Its data component is the matching reassociation of the three extent summands.

Definition 11.8 (The Atlas Unitors)**.** The **Atlas Left Unitor** $\mathsf{Lu} : \mathsf{AtlHorSum}(\mathsf{AtlI}, F) \to F$ and the **Atlas Right Unitor** $\mathsf{Ru} : \mathsf{AtlHorSum}(F, \mathsf{AtlI}) \to F$ remove the empty tag family.

12 Stable Data Transversals

Definition 12.1 (Stable Data Transversals)**.** The **Category of Stable Data Transversals**, denoted $\mathsf{StaDaTrav}$, is the presheaf category on Atlas Federations whose component arrows lie in $\mathsf{StaAtlTrav}$.

Definition 12.2 (The Empty Map)**.** The **Empty Map**, denoted I, is the Day unit represented by the empty Atlas Federation:

$$I = \mathsf{Yo}(\mathsf{AtlI}).$$

Definition 12.3 (The Horizontal Sum Bifunctor)**.** The **Horizontal Sum Bifunctor**, denoted $\mathsf{HorSum} : \mathsf{StaDaTrav} \times \mathsf{StaDaTrav} \to \mathsf{StaDaTrav}$, or in infix notation by $\boxplus : \mathsf{StaDaTrav} \times \mathsf{StaDaTrav} \to \mathsf{StaDaTrav}$, is the Day convolution extension of stable atlas horizontal sum. Because the indexing morphisms are arrows of $\mathsf{StaAtlTrav}$, its result and its arrow action land in $\mathsf{StaDaTrav}$ by construction.

Definition 12.4 (The Stable Data Transversal Braider)**.** The **Stable Data Transversal Braider** $\mathsf{Brd}_{F,F'} : F \boxplus F' \to F' \boxplus F$ is the Day convolution extension of AtlBrd.

Definition 12.5 (The Stable Data Transversal Associator)**.** The **Stable Data Transversal Associator**

$$\mathsf{Asoc}_{F,F',F''} : (F \boxplus F') \boxplus F'' \longrightarrow F \boxplus (F' \boxplus F'')$$

is the Day convolution extension of $\mathsf{AtlAsoc}$.

Definition 12.6 (The Stable Data Transversal Unitors)**.** The **Stable Data Transversal Left Unitor** $\mathsf{Lu} : I \boxplus F \to F$ and the **Stable Data Transversal Right Unitor** $\mathsf{Ru} : F \boxplus I \to F$ are the Day convolution extensions of AtlLu and AtlRu.

Definition 12.7 (Stable Data Transversals Monoidal Category)**.** The **Stable Data Transversals Monoidal Category**, denoted $\mathsf{StaDaTravMon}$, is

$$\mathsf{StaDaTravMon} = (\mathsf{StaDaTrav}, \boxplus, I).$$

Its tensor is Day convolution on the Atlas Federation. The braider, associator, and unitors are induced by retagging that form, so all coherence maps remain within stable data transversals.