

Working Group I: Tool Interoperability

Mike Cochran (Co-chair), Jeff Good (Co-chair)
Steve Miller, Dan Loehr, Briony Williams, Shane
Stephens, Imelda Udoh

Themes

- Resolving data model incompatibilities
- Typology of annotations
 - String, controlled vocabulary
 - Time-aligned, aligned to another annotation
- Basic interoperation system requirements
- Sample interoperation system models

Model incompatibility

- Concrete example
 - ELAN allows you to associate an annotation tier with a speaker
 - Praat doesn't
- If I convert ELAN files to do work with them in Praat—and then go back—how do I make sure I don't lose that extra information?

Diversity is fine

- There will be many tools
- They will do different things
- They will not all annotate for the same information
- We probably can't—and shouldn't—expect to use the exact same data models
- But how do accept this, but keep the data

Opaque annotations

- We can begin by making a notional distinction between two types of annotation with respect to a given program
 - Transparent annotations: Annotations which the program's data model allows it manipulate
 - Opaque annotations: Annotations which the program's data model does not allow it manipulate—because it came from another program
- Basic desiderata: Opaque annotations should be treated nicely

A problem

- What if one annotation depends on another?
- What if the dependent annotation is opaque, but the other one is transparent and a tool deletes the transparent annotation?
- When annotation conflicts arise
 - The data should not be lost
 - There should be straightforward methods for resolving inconsistencies in a satisfactory manner

More concretely

- In ELAN some annotations are linked directly to times in an audio file
- Others, though—for example, translations—are linked to other annotations
- What if I delete the annotation a translation was linked to in Praat?
- Ideally: After finishing with Praat, some tool would tell me I have a “stray” annotation floating around and ask me what to do with it?

Interlinear example

l'homme	
l'	homme
the	man

Annotation typology

- Annotation constituents
 - Annotation class, e.g., time-aligned, symbolically aligned, subdivision, whole resource annotation
 - Annotation content, e.g., string, controlled vocabulary
 - Annotation metadata, e.g., annotation groups, provenance

Annotation relationships

- Part-whole relations
 - Easier
 - Simple part-whole relationships
- Cross-cutting
 - Harder
 - No simple part-whole relationships

Mediation models

- File-based mediation: Tools import/export from/to interlingua
 - Interlingua: A common exchange format
- Process-based mediation: Separate tool mediates conversion
- An alternative: Tools import/export all other tools' formats (an n-squared problem)
- Process-based mediation may or may not use an interlingua

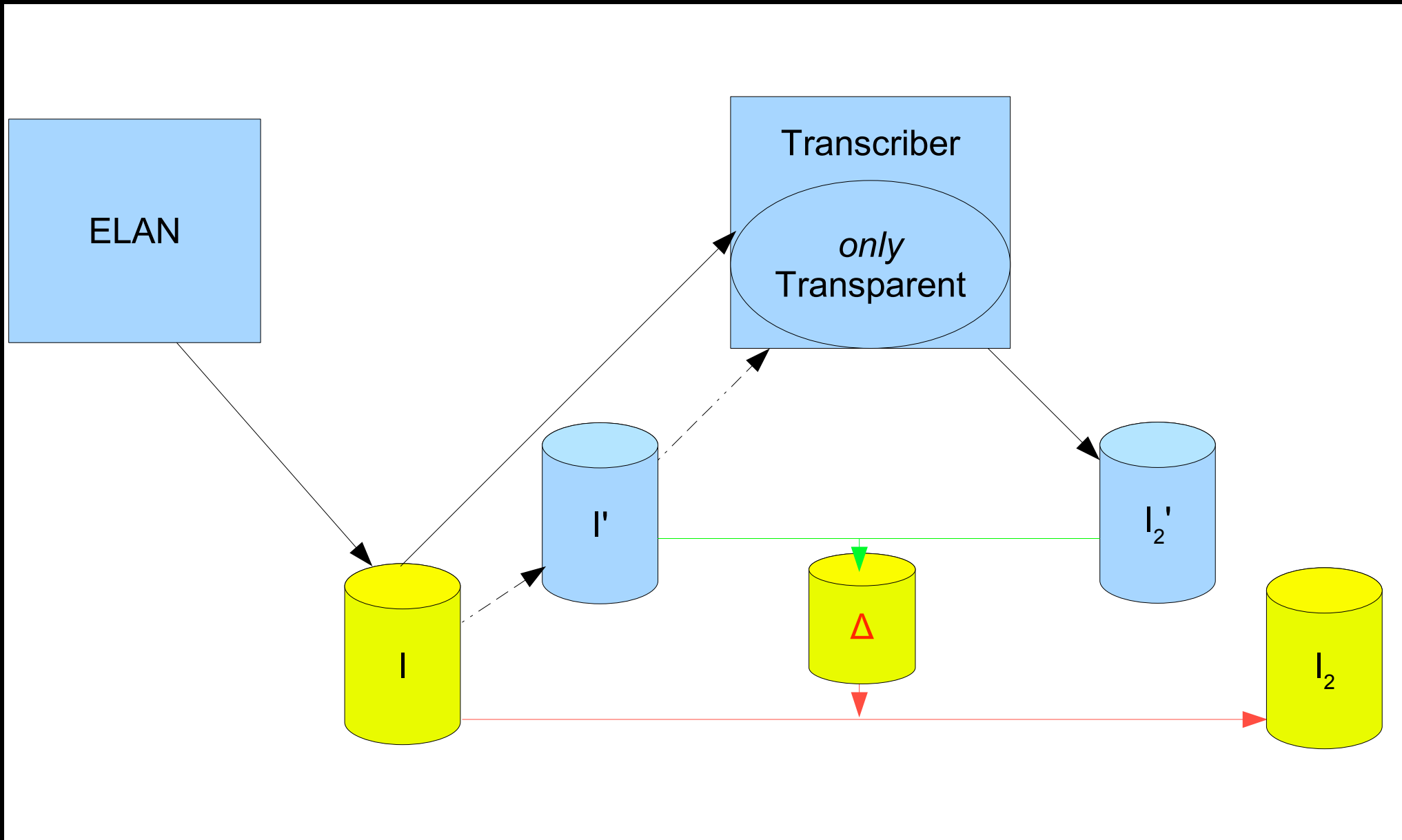
Conflicts

- A key desideratum is good conflict resolution
- A user should be warned of any problems introduced due to tool interoperation
- There's work to be done on how best present warnings

Conflict typology

- We know of several kinds of conflicts
 - Conflicts caused by tier types supported by one tool but not the other (e.g., time-aligned, symbolic-aligned)
 - Conflicts caused by annotation relationships (only one-to-many or many-to-many)
 - Cross-cutting tree structures
- And we know there must be more, and some will be very hard

Structure of problem



Conflicts

- Conflict resolution is hard
 - It requires detailed knowledge of the data constraints for each tool
 - It's too burdensome for each tool developer to be aware of every other tools' model
 - Rather, we would want central mechanism to handle data model interoperability

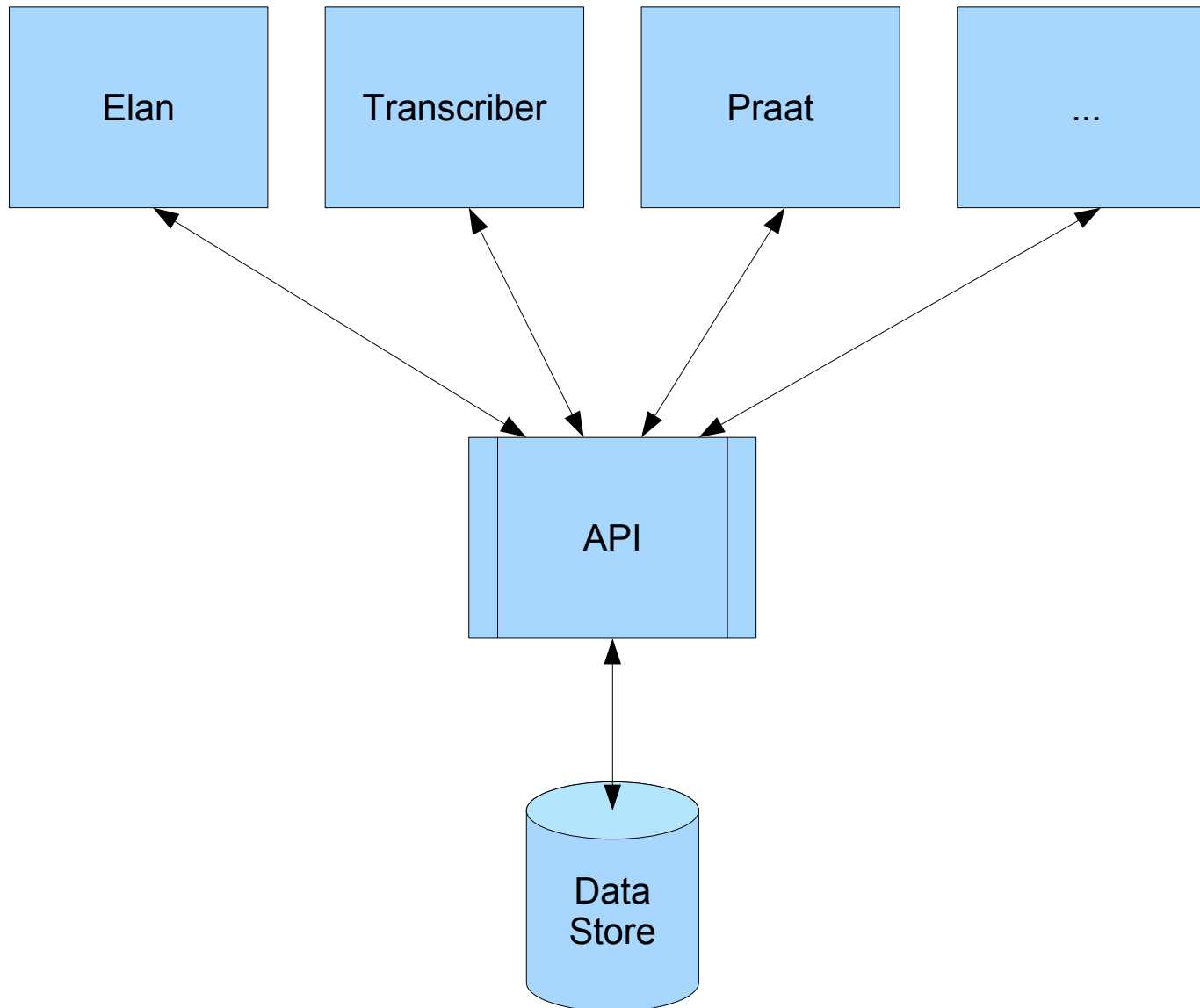
Proposed API functions

- Mediates between data and tools
- Has a well-defined interlingua (e.g., Annotation Graphs)
- Validates data to maintain data integrity
 - A tool facilitating conflict resolution could be built on top of such validation facilities
 - ...and is definitely needed

Proposed API functions

- Mediates between data and tools
- Has a well-defined interlingua (e.g., Annotation Graphs)
- Validates data to maintain data integrity
 - A tool facilitating conflict resolution could be built on top of such validation facilities
 - ...and is definitely needed

One model



What we have here

- A roadmap suitable for a grant proposal