

General Visualization Methods for Analysis and Exploration of Provenance Data

Goals

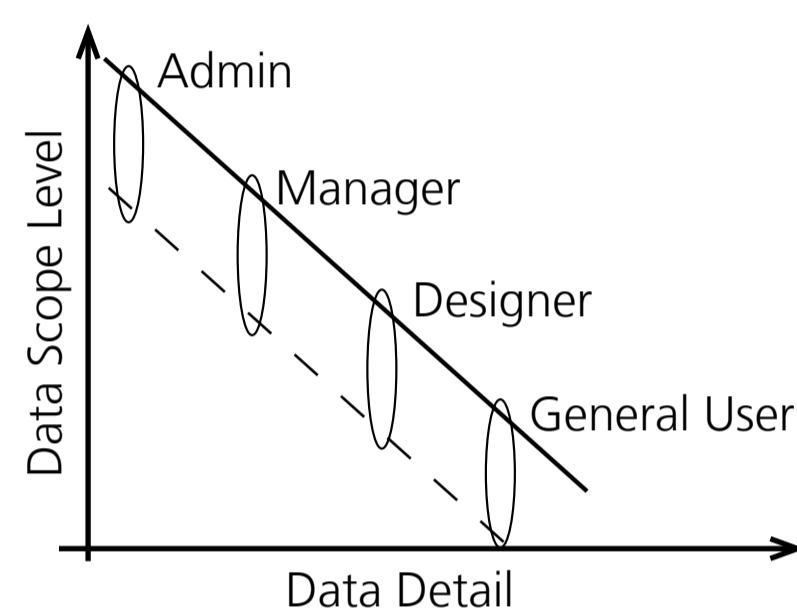
- Development of a general *visualization panel* for browsing the contents of provenance stores
- Applications in *scientific technical data management, Software Engineering* and analysis and management of *distributed systems* and services
- More comfortable verification of recorded Provenance data in comparison to XML, log file or spreadsheet browsing

User Groups

(derived from requirements of EU project "Grid Provenance")

Scope: High-level border, range of access.

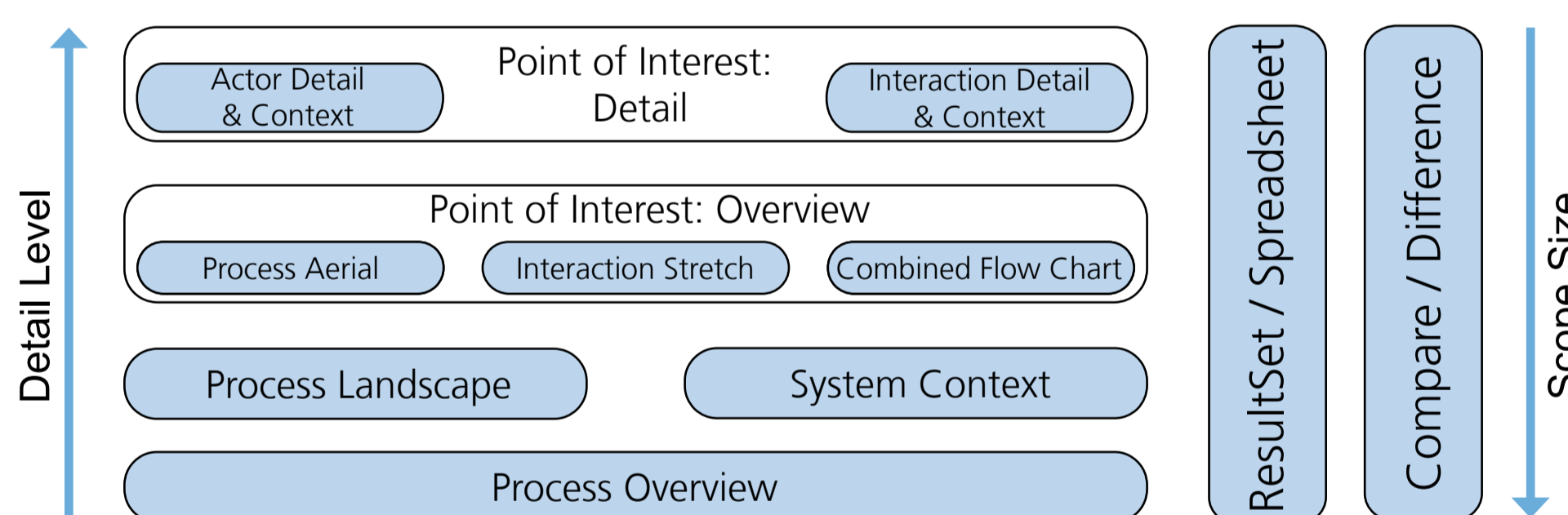
Intention: Low-level border, type and detail of interest in Provenance data



Classification	Examples	Scope	Intention
General User	Scientist Engineer Portal User	own work, own results, origin of used data	reliability and authenticity of results, reproducibility
Designer	Software Engineer Workflow Developer	Project related, all origins, monitored system, partner-made components	workflow behavior, service interaction, product evolution
Manager	Workflow Provider Provenance Analyst User Support	all assigned user and system Provenance	correctness of services, interpretation support, quality of the P-system
Administrator	Developer / Admin of Provenance System	all P-data available in connected P-stores	building the P-system and maintaining its function

Suitability of Visualization Methods

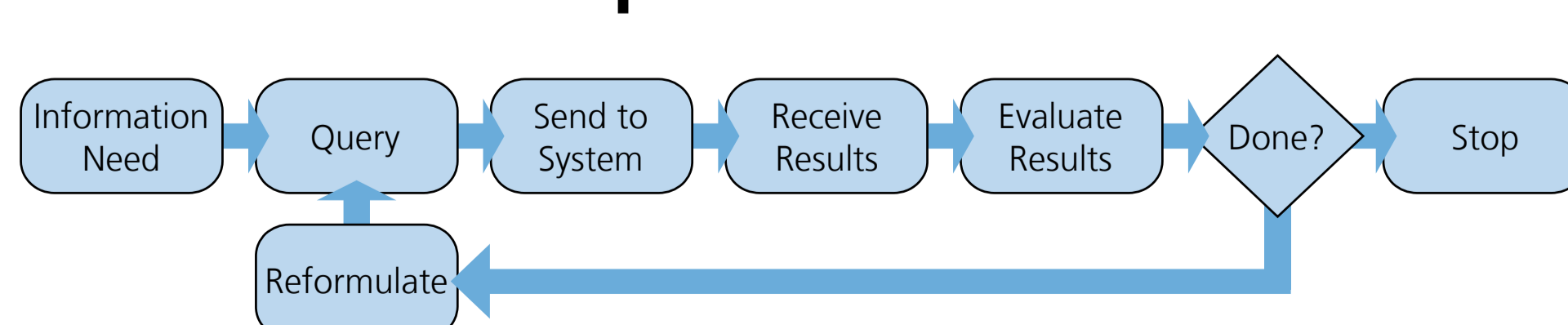
Visualization methods grouped by wanted level of detail and accessible scope of information:



Projection of user requirement types towards suitable visualization methods:

Type	Description	Provenance Data	Visualization
Process	Evaluation of the approach of a workflow	Actors, interactions, sequence of process steps	Process Landscape Process Aerial Combined Flow Chart
Results	Quality of intermediate and end results of processes	Dependencies of inputs and outcome	Interaction Stretch Process Landscape
Relationship	Analysis of the evolution of data	Relationships of interactions or actors	Interaction Stretch
Time Line	Finding performance bottlenecks, improving workflows	Evolution of results, comparison of actor behavior	Interaction Stretch Process Aerial Process Landscape
Participation	Trust to result	Participating actors	Process Landscape
Comparison	Validate correctness of processes and results, by comparing documented executions with reference structures	Processes, views on interactions, results	Compare Difference Process Aerial
Interpretation	Custom visualization requirements, deriving knowledge from Provenance data	Custom, probably all aspects	interactive, recursive exploration and combination of visualization methods.

General Data Exploration Procedure

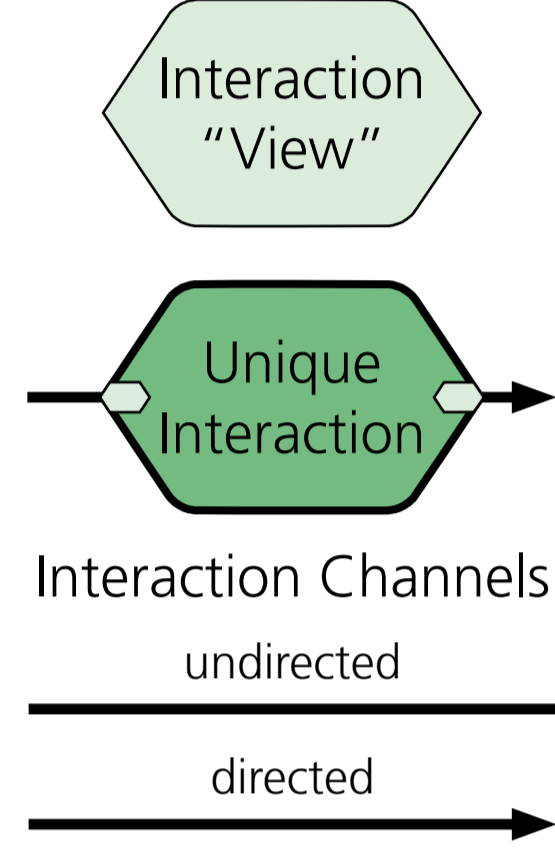


Provenance Concepts

An abstraction of the Provenance architecture developed at the University of Southampton, UK.

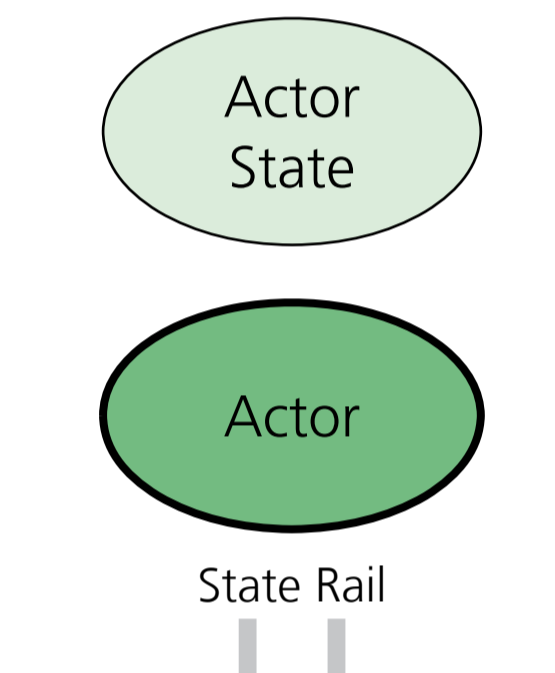
Interaction

- **globally unique** message between two actors
- directed from one **sender** to one **receiver**
- sender and receiver have different knowledge about the same interaction and thus document **different views** towards it



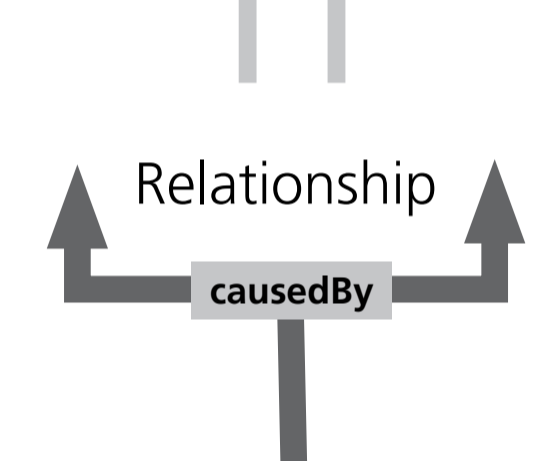
Actor and Actor State

- each actor is **unique**
- actors are located on **distributed** resources
- actors **interact** with each other
- **consume and produce data**
- reach different **internal states**, some of which may cause interactions



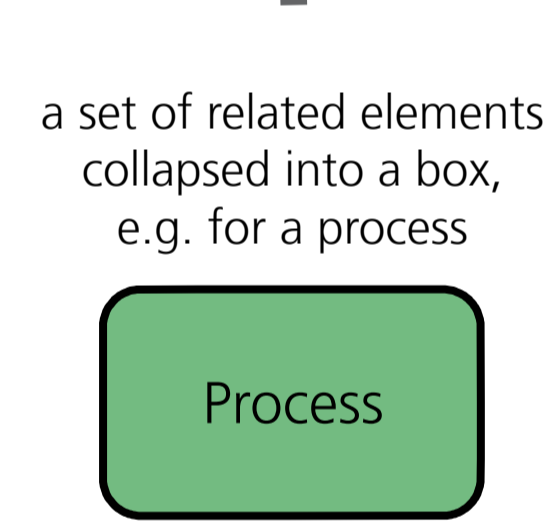
Relationship

- description of a **cause-and-effect relationship** between interactions
- one subject, one to many objects
- defined by a **relation type** (passive predicate, e.g. "caused by")



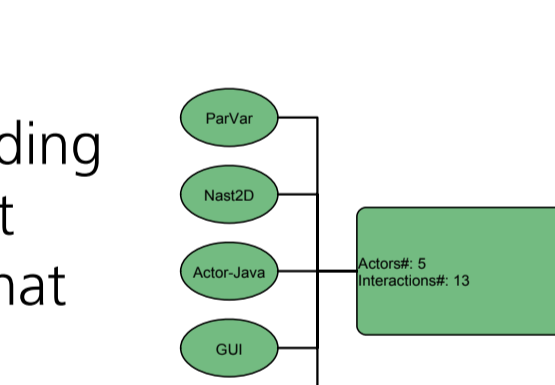
Grouping

- combined, structured documentation of incidents that **share an application context**, e.g. workflow executions
- recorded as specific data structures or marked with **Tracers**



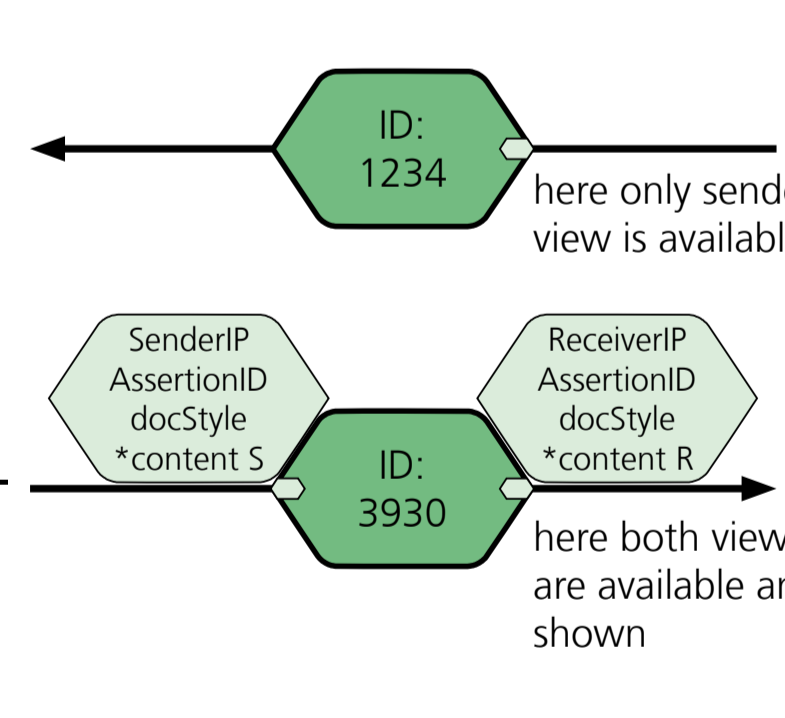
System Context

- shows the interface of a process regarding to input, output and execution context
- implicitly contained in all chart types that allow collapsing of process elements



Interaction Views

- little bright hexagons at the edge joints indicate available actor views that can be shown on request
- the order depends on the interaction direction
- the message "content" is application specific and may afford display in a dedicated panel



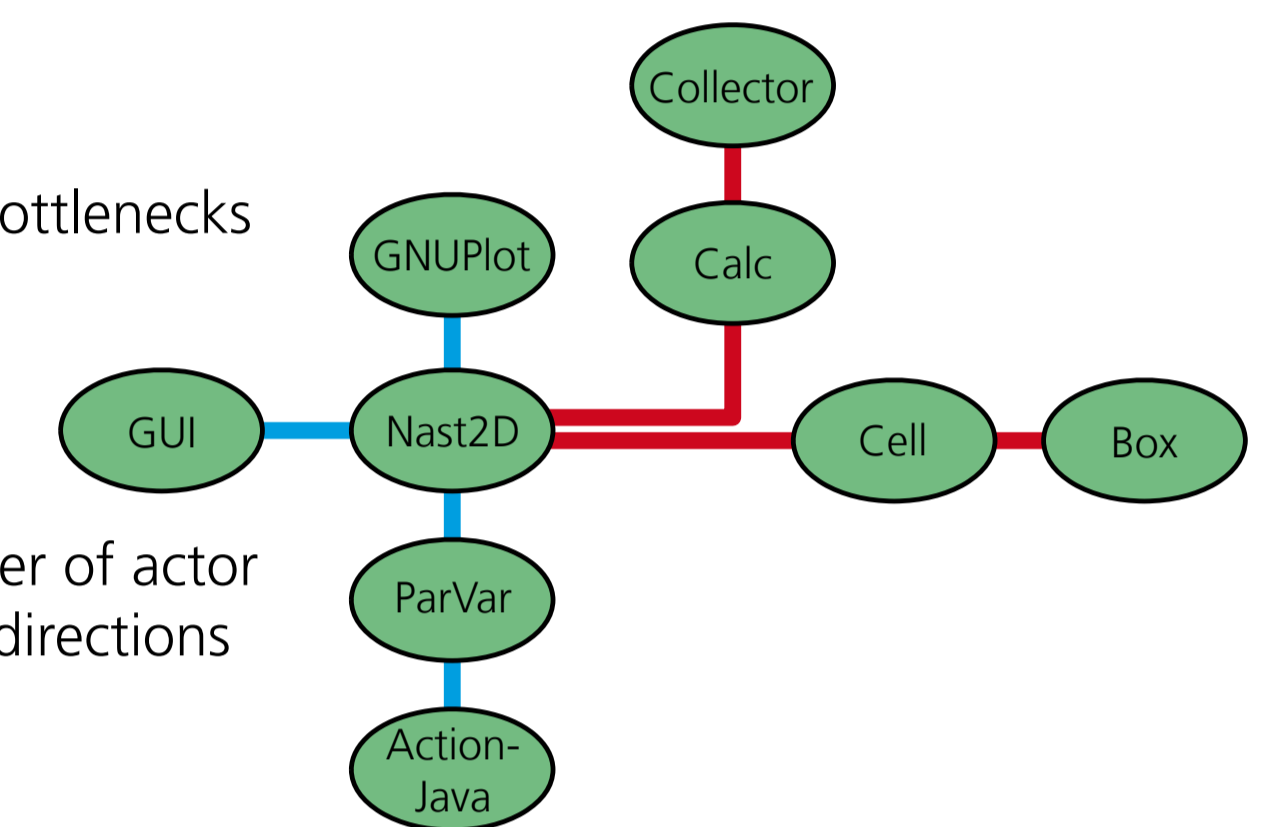
Visual Exploration Methods

- **Brushing & Linking:** Reflect the manipulation performed in one view simultaneously in other views that show the same content just projected differently.
- **Grouping / De-Grouping:** Logically connect selected elements to continuously manipulate them as one element, while maintaining the structure inside the group unchanged.
- **Collapsing:** Hide unwanted and show wanted information. Groups of connected elements can be shown as one process element for simplification and interaction views of can be shown on request (see "Interaction Views").
- **Switching Levels:** Change vis. method to display more or less detail.
- **Grasp & Release:** Move elements manually, if automatic layout is not sufficient.
- **Panning & Zoom:** Navigate the shown graph, by moving the view point in 2D/3D.
- **Search:** Find elements by text queries.
- **Selecting:** Select several elements to manipulate them all at once.
- **Undo:** Undo manipulation or navigation steps.

Visualization Chart Types

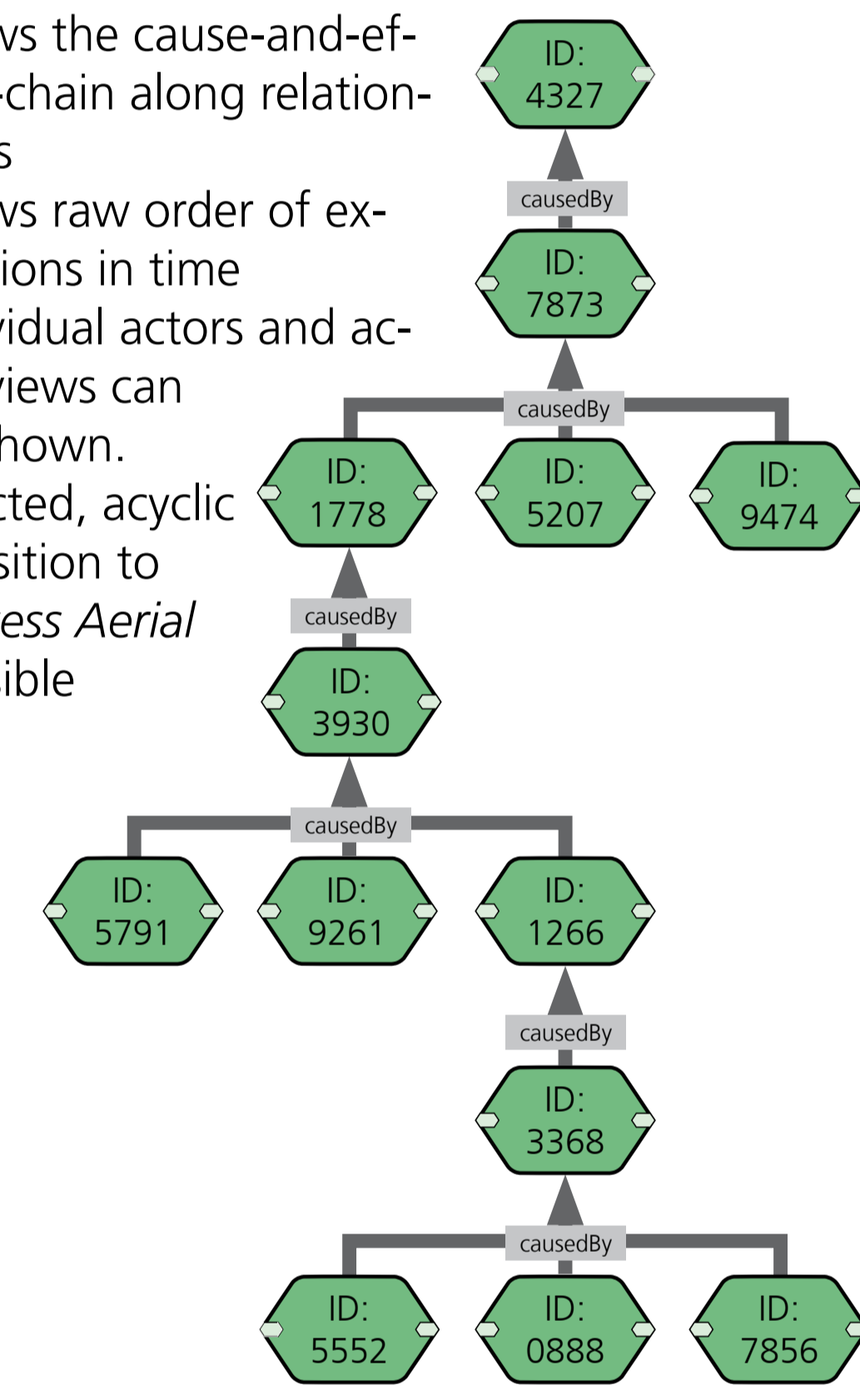
Process Landscape

- Inspired by "tube map"
- participation of actors in processes
- Supports locating possible system bottlenecks as actors shared by many processes
- Edges denote interaction channels
- Tracer-coloring edges serves to distinguish processes
- Extensible to reflect recorded number of actor states, interactions and interaction directions



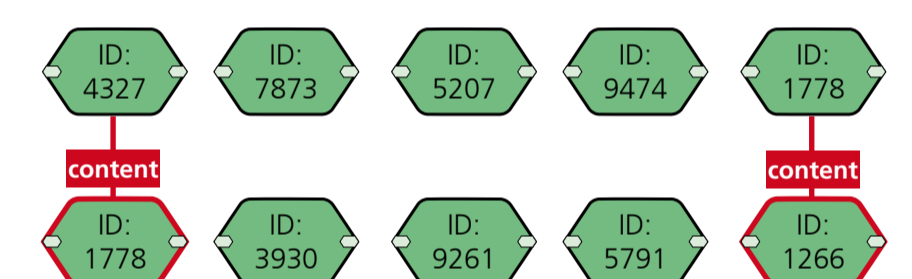
Interaction Stretch

- shows the cause-and-effect-chain along relationships
- shows raw order of executions in time
- Individual actors and actor views can be shown.
- directed, acyclic
- transition to *Process Aerial* possible



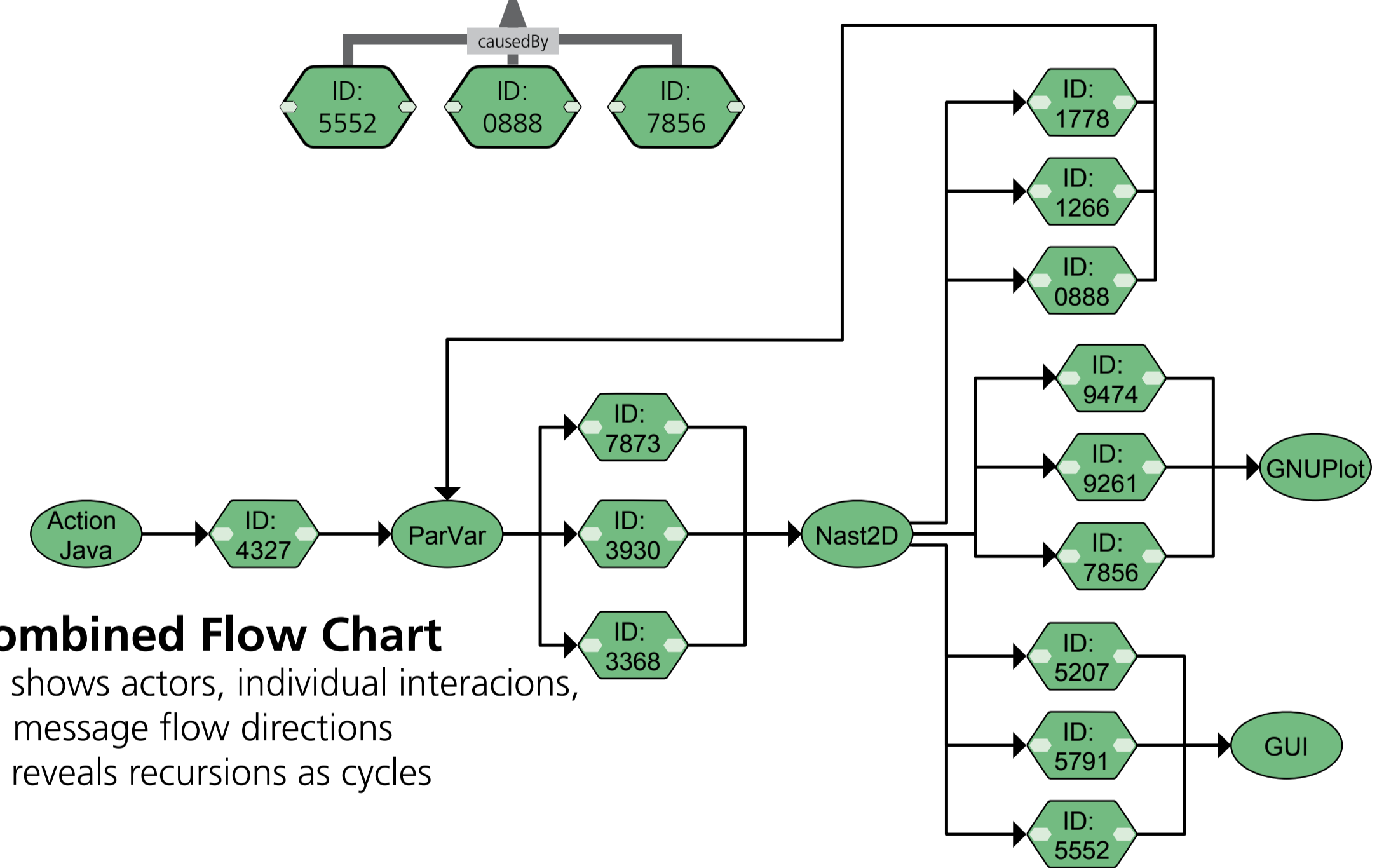
Difference Comparison

- identifies differences in specific attributes of a selection and reference elements
- recognizing differences in process executions requires definition of application specific roles



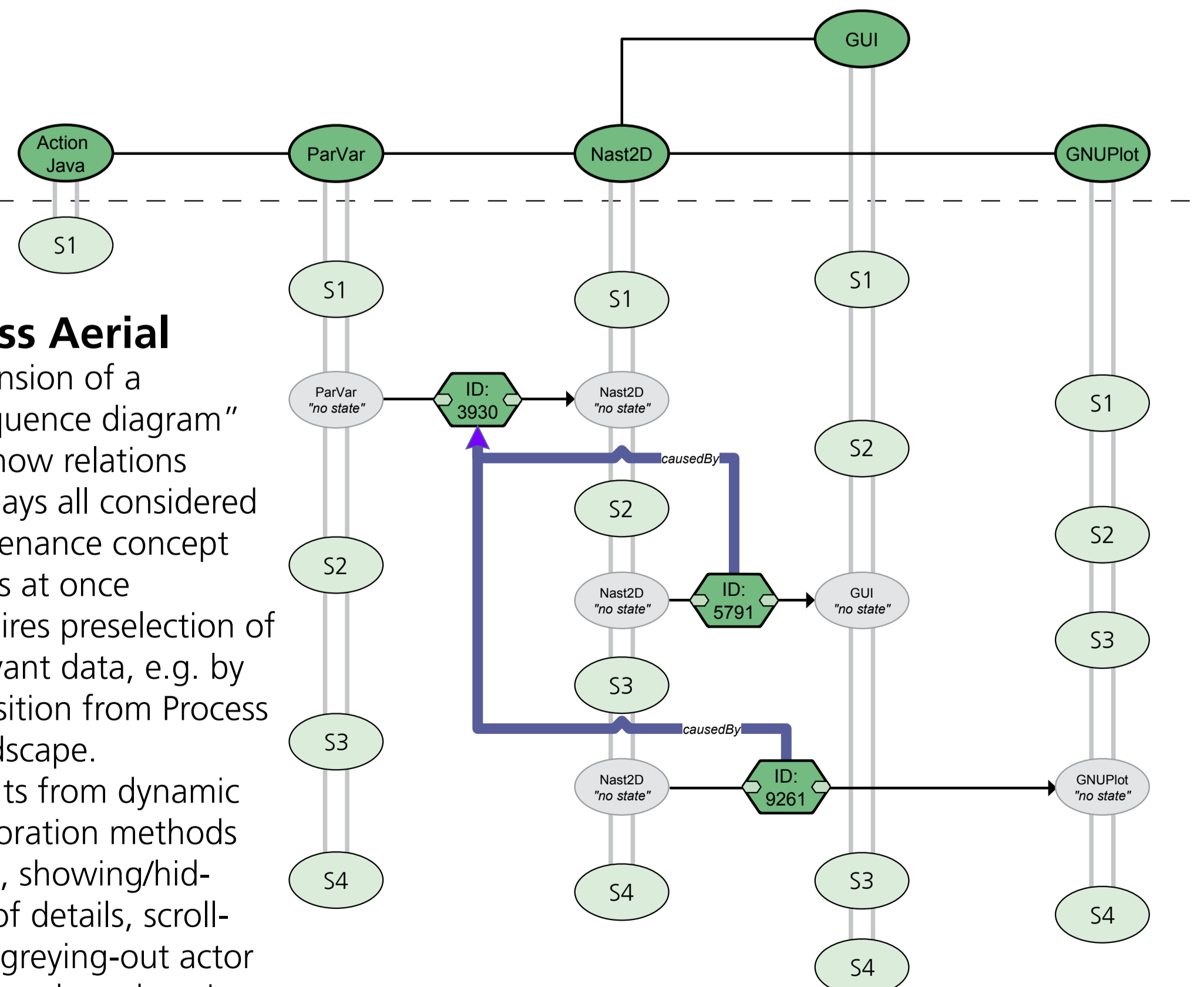
Combined Flow Chart

- shows actors, individual interactions, message flow directions
- reveals recursions as cycles



Process Aerial

- extension of a "sequence diagram" to show relations
- displays all considered Provenance concept types at once
- requires preselection of relevant data, e.g. by transition from Process Landscape.
- profits from dynamic exploration methods (e.g., showing/hiding of details, scrolling, greying-out actor states when choosing relations)



Markus Kunde
Research Associate
German Aerospace Center (DLR)
Simulation and Software Technology
Distributed Systems and Component Software
Linder Höhe, 51147 Cologne, Germany
Email: Markus.Kunde@dlr.de

Henning Bergmeyer
Research Associate
German Aerospace Center (DLR)
Simulation and Software Technology
Distributed Systems and Component Software
Linder Höhe, 51147 Cologne, Germany
Fon.: +49 2203 601-3581
Email: Henning.Bergmeyer@dlr.de

Andreas Schreiber
Head of Department
German Aerospace Center (DLR)
Simulation and Software Technology
Distributed Systems and Component Software
Linder Höhe, 51147 Cologne, Germany
Fon.: +49 2203 601-2485
Fax: +49 2203 601-3070
Email: Andreas.Schreiber@dlr.de



Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

Simulation and
Software Technology