Safety demonstration and structured argumentation

IFE Digital Systems
Peter Karpati
Svein Tore Edvardsen
André A. Hauge
Bjørn Axel Gran
Silje Olsen
Fabien Sechi
Vikash Katta

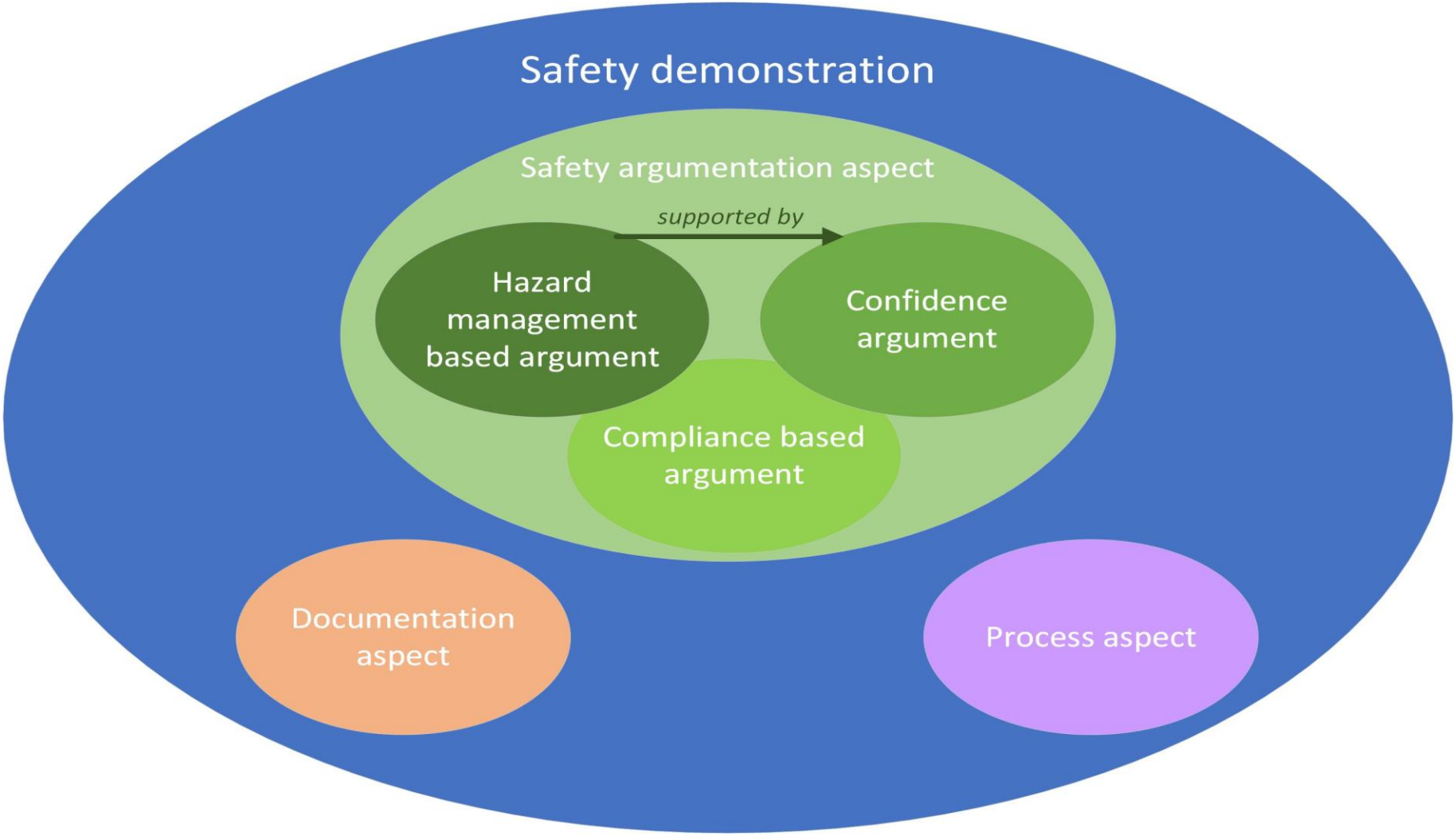
VTT
Markus Airila

6.12.2018, Lillehammer

Outline

1. Concepts of safety demonstration and structured argumentation
2. Tool support for information structuring: InStrucT
3. Application for decommissioning

Safety demonstration: *documents*, *tasks*, and ***argumentation*** intended to demonstrate that the safety of a system and/or related activities are sufficiently taken care of.



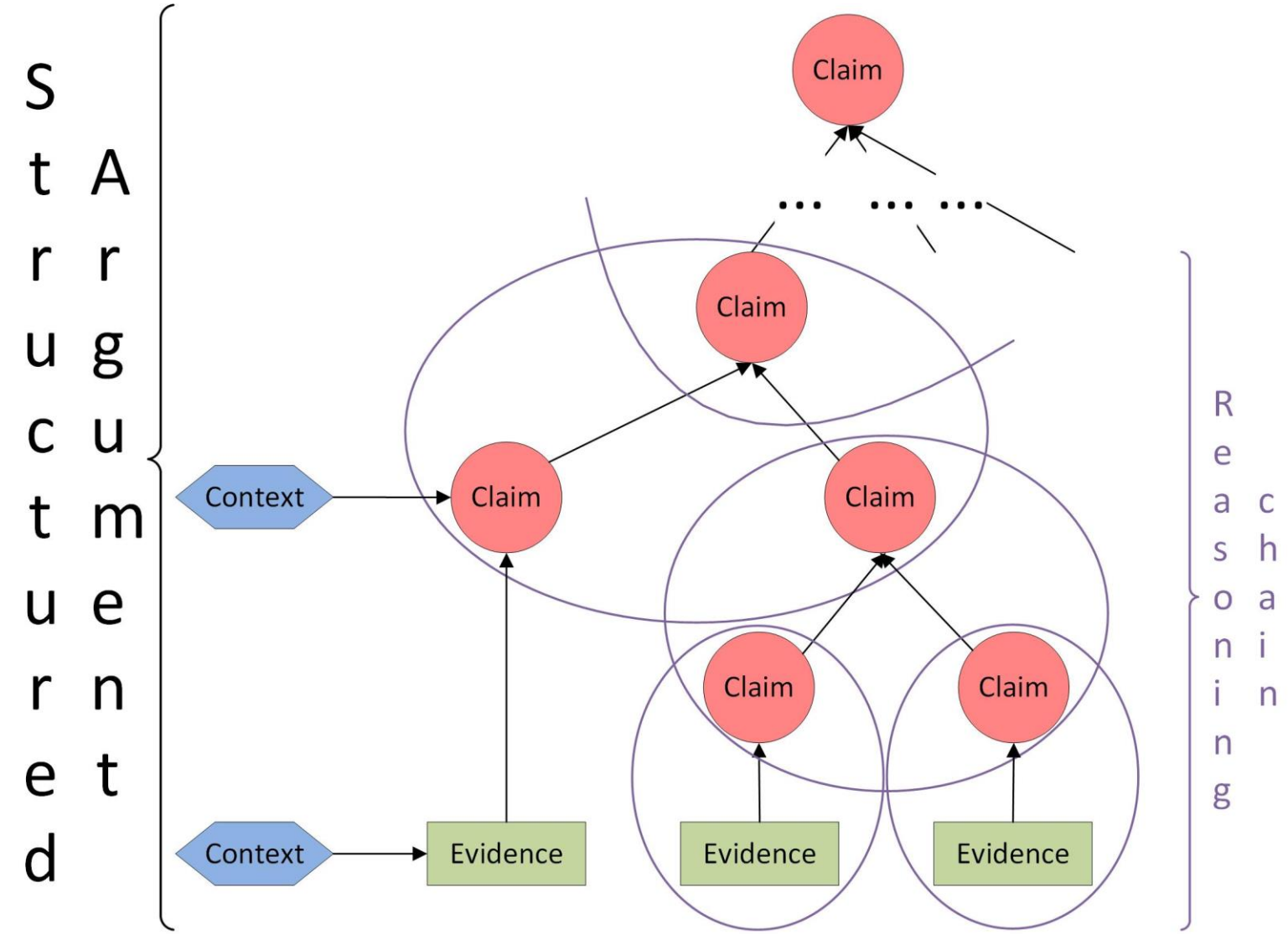
Structured safety arguments

Safety demonstration is usually presented as a set of linear, natural language documents in pdf.

Structured safety arguments can be used to present the relevant information and its logical structure explicitly.

- Better assessable
- Supports communication between parties
- Improves safety
- Reduces regulatory uncertainty
- Saves costs

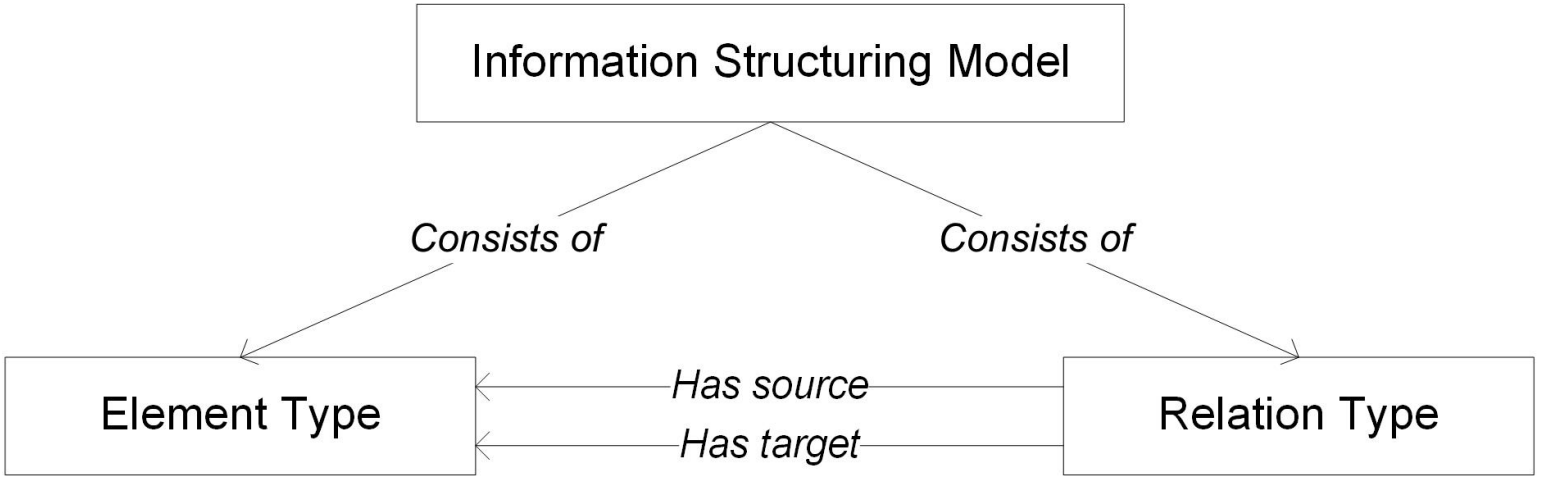
Basic model of structured argument



2. Tool support for information structuring: InStrucT

- Information structuring – usual generic case
  - *Input*: mixed information, linearly presented (e.g. in pdf)
  - *Process*: extracting the important pieces, categorizing and organizing them according to a goal
  - *Output*: categorized and interrelated information pieces
- Motivation
  - Helps pinpointing unclear parts and missing information
  - Helps avoiding/reducing misunderstandings
  - Helps communication and discussion
  - Reduces related risks and thus costs

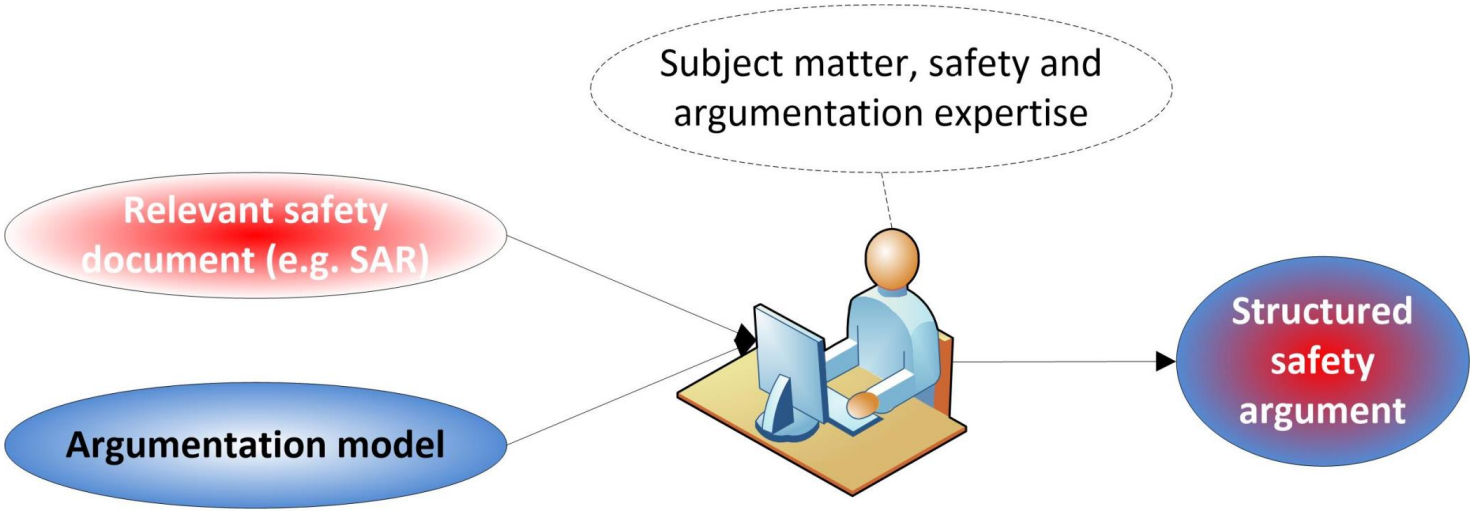
Information structuring model



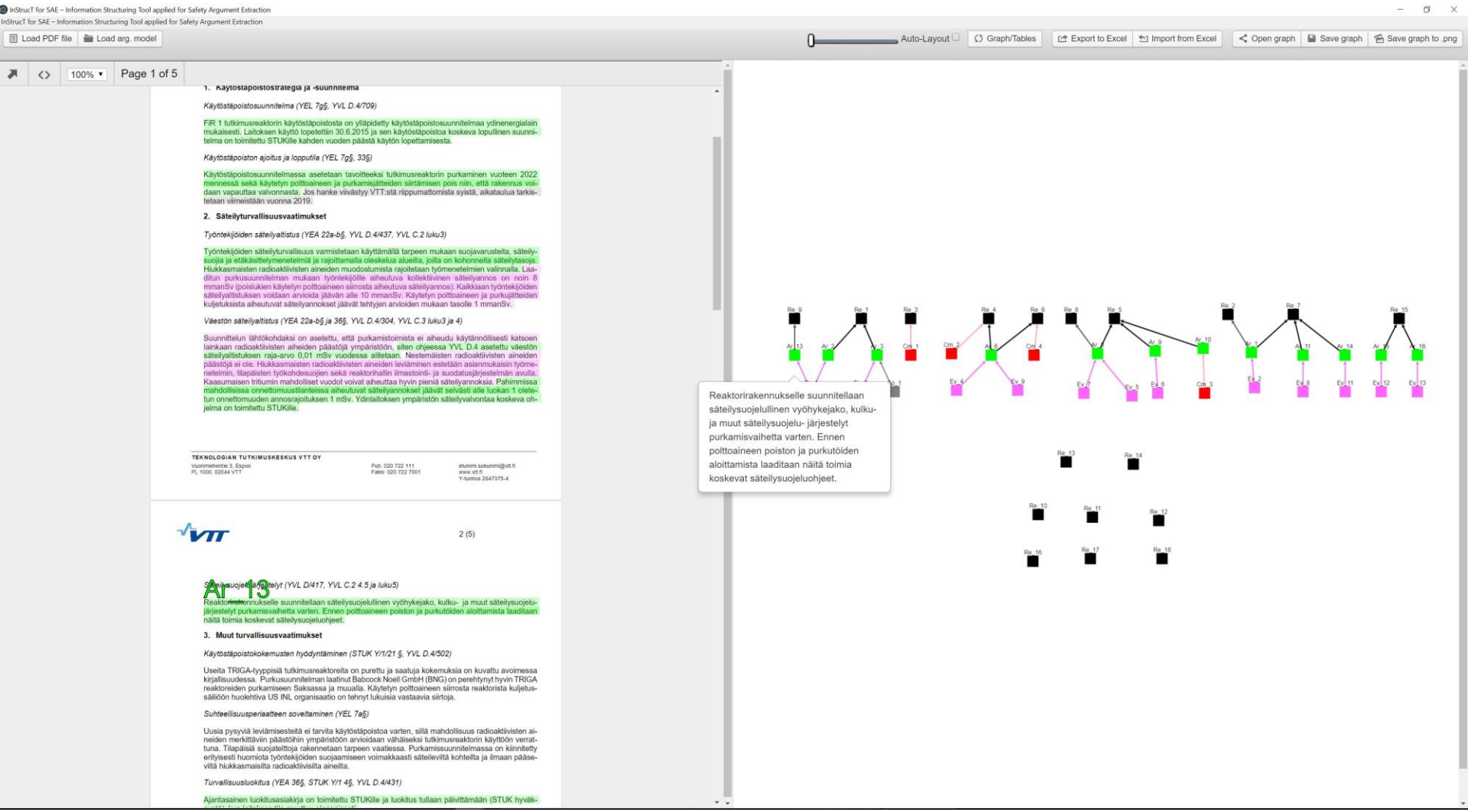
- *Element Type*: categories to group and tag the same kind of information pieces
- *Relation Type*: links between categories representing the nature of their relations

InStrucT: Information Structuring Tool

- Prototype
- Used in 2 case studies to create safety arguments (reasoning structures)
- Functionality
  - Organising and structuring information according to *pre-defined categories and relations* between them



InStrucT in use (decom. case)



Main functionalities of InStrucT

- *Reading* one *pdf* document and an information structuring model description
- *Tagging* continuous text parts in pdf-s
- Presenting the structured information graphically as a *directed graph,* or as a *table*
- Creating *freely definable nodes and rel*ations
- Saving and loading extracted information structure as a *graph* (*keeping links* to the pdf)
- Saving and loading extracted information structure as a *table* (*loosing links* to the pdf)

InStrucT Viewer

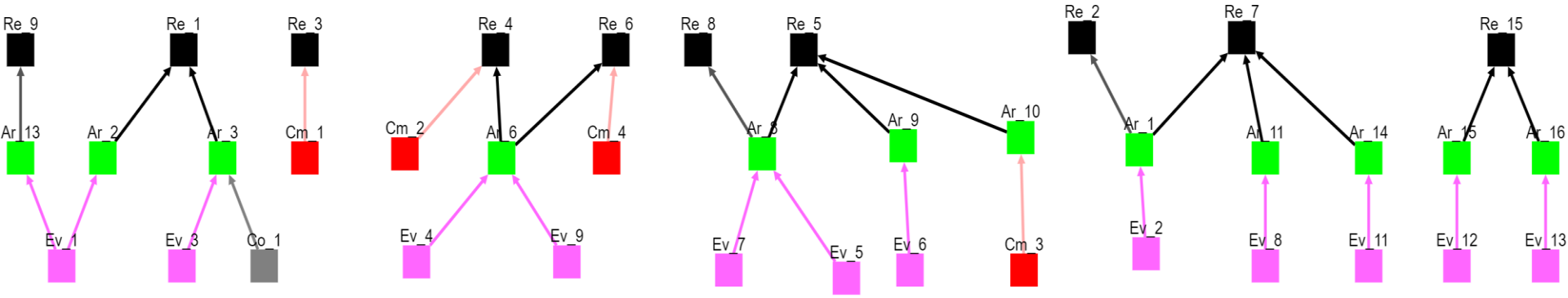
- Goal: to be able to share an information structure with another party for viewing without the need to install InStrucT
- The graph created in InStrucT can be viewed through this online viewer
- The owner of the graph has to share the graph and the related pdf document with the targeted person
- Ready but needs testing yet
- Web address

  http://instruct-viewer.hrp.no/

3. Application for decommissioning

- In cooperation with VTT in Finland
- Case
  - Research reactor FiR 1 in Espoo, Finland
  - The reactor is currently in permanent shutdown state, and VTT's license application for decommissioning is under review by Finnish authorities
- InStrucT was used for
  - Stage 1 – regulatory documents
    - Extracting and analyzing regulatory requirements
  - Stage 2 – applicant documents
    - Analyzing a part of the decommissioning license application
    - Defining the reasoning structure how parts of the license application fulfill the regulatory requirements from Stage 1

Illustration of the case study in 7.1



- Back rectangles: requirements from the regulations
- Purple rectangles: pieces of evidence from the license applications
- Green rectangle: arguments from the license applications, how the evidence show that the requirements are fulfilled
- Red rectangles: comments, remarks from the user
- Grey rectangles: context from the license application

Future sights

- Handling of multiple documents
- Tagging of documents (not just their context)
- Communication support for stakeholders
- Integration/extension into an information management system
  - Interrelated, queryable information
  - Change management
  - Traceability
  - Filtering of information
  - Pre-defined views/perspectives (e.g. safety argument, decommissioning plan, cost estimate, etc.)
  - Multi-media capable (e.g. safety argument integrated in a 3D scenario)

Thank you for your attention!

Questions?

Peter.Karpati@ife.no

Reserve slides

+1. Barriers to assuring of autonomous systems

- Based on the Assuring Autonomy International Programme at University of York
  - https://www.york.ac.uk/assuring-autonomy/
- Scope: assurance of Robotics and Autonomous Systems (RAS)
- Critical Barrier to Assurance and Regulation (C-BAR) is a problem that must be solved for a particular system or domain, in order to avoid one or more of the risks presented next.

Risks (to be avoided by coping with C-BARs)

- a safe system cannot be deployed (losing the benefit of the technology)
- an unsafe system is deployed (lack of clear evidence to assure operation)
- the adoption of safe technology is slow
- there is a lack of progress in adoption in a particular domain
- the level of accidents and incidents leads to a backlash

C-BARs

- Adaptation – of behaviour in operation
- Bounding Behaviour – safe operation within known bounds
- Cross-Domain Usage – known to be effective in one domain, how can it be assessed for adequacy in another environment
- Explanations – of decisions made by a RAS
- Handover – handing (back) control to a human
- Human-Robot Interaction – in sight of potential for physical harm to humans
- Incident and Accident Investigation – information needed to be provided to support incident/accident investigations

C-BARs – cont.

- Monitoring – retain sufficient levels of attention and concentration of operators
- Risk Acceptance – how can risk be estimated, communicated and accepted?
- Role of Simulation – how can it enable assurance and regulation, and when does it provide sufficient evidence to allow controlled use of the RAS?
- Systems of Systems –when given SoSs which are ‘individually safe’ how can safe interaction be assured, in their intended operational environment?
- Training and Testing AI – how can it be shown that the training sets (and test sets) give enough coverage of the environment to provide sufficient evidence (in itself or in combination with other means of V&V) to allow controlled use of the RAS?
- Validation & Verification – effective means of RAS/AI V&V