



University of Stuttgart
Germany



Institute for Visualization
and Interactive Systems



Visualization Research Center
University of Stuttgart



European Union *Horizon 2020* project InTaVia



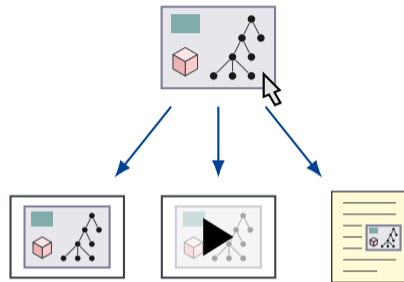
Deutsche
Forschungsgemeinschaft

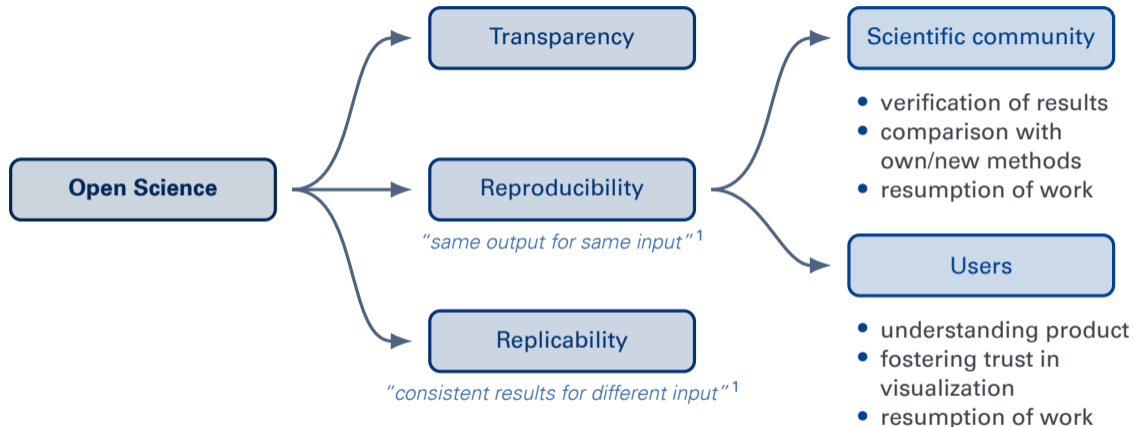
Toward Reproducible Visual Analysis Results

Max Franke, Guido Reina, Steffen Koch
University of Stuttgart, Germany



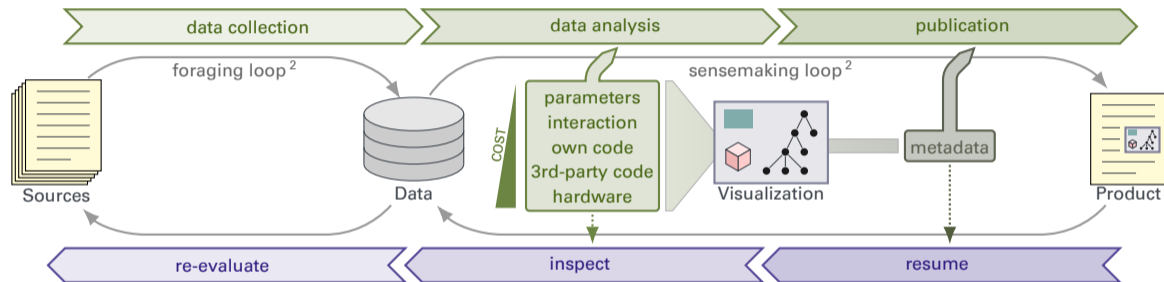
- typical visualization approaches output *products*
 - images
 - videos
 - documents
 - ...
- often all that is shared when communicating results





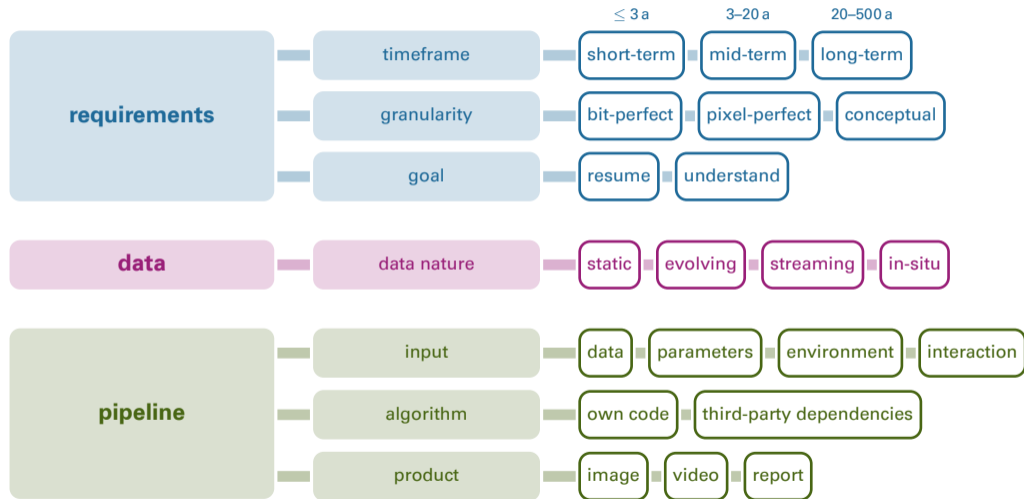
¹ Jean-Daniel Fekete and Juliana Freire. "Exploring reproducibility in visualization." In: *IEEE CG&A* 40.5 (2020), pp. 108–119. DOI: 10.1109/mcg.2020.3006412

Reproducibility Pipeline

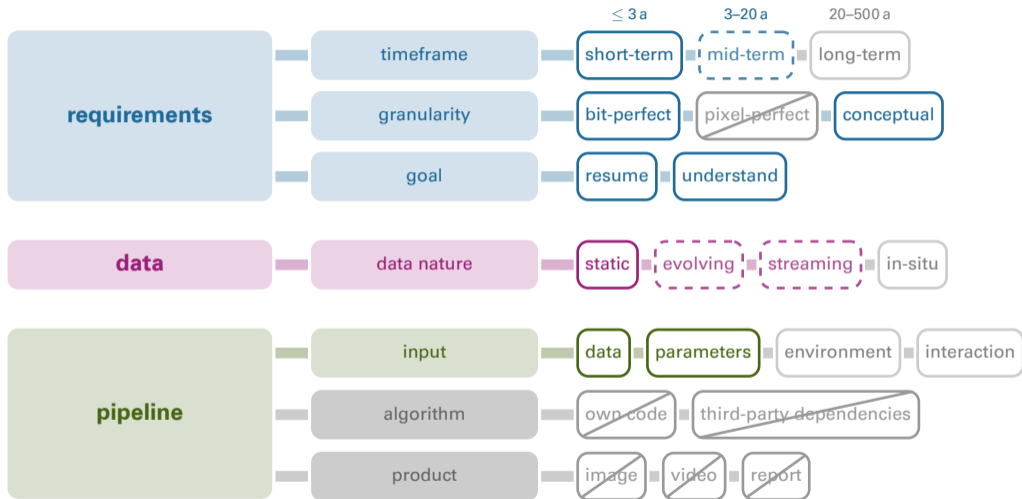


²Peter Pirolli and Stuart Card. "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis." In: *Proc. International Conference on Intelligence Analysis*. Vol. 5. 2005, pp. 2-7

Typology



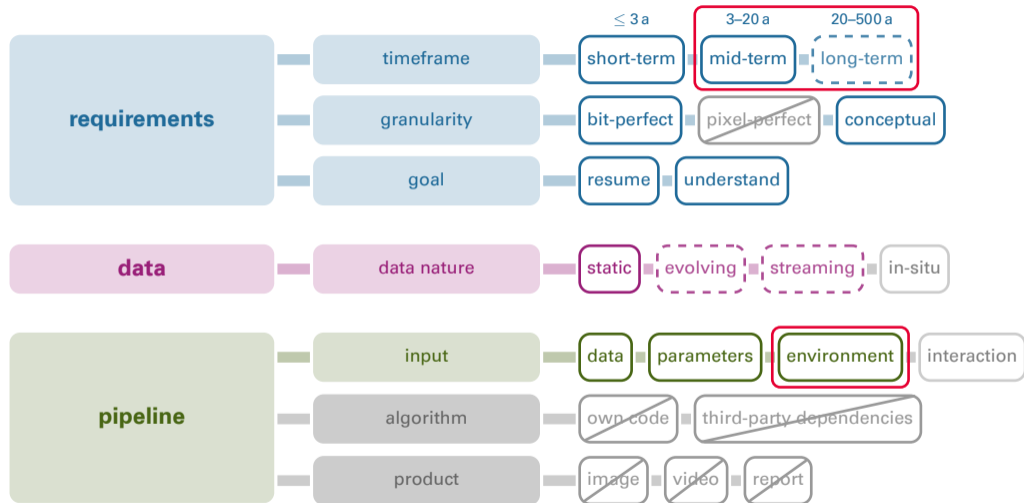
Stored datasets



Setup cost: ★☆☆

Maintenance cost: ★☆☆

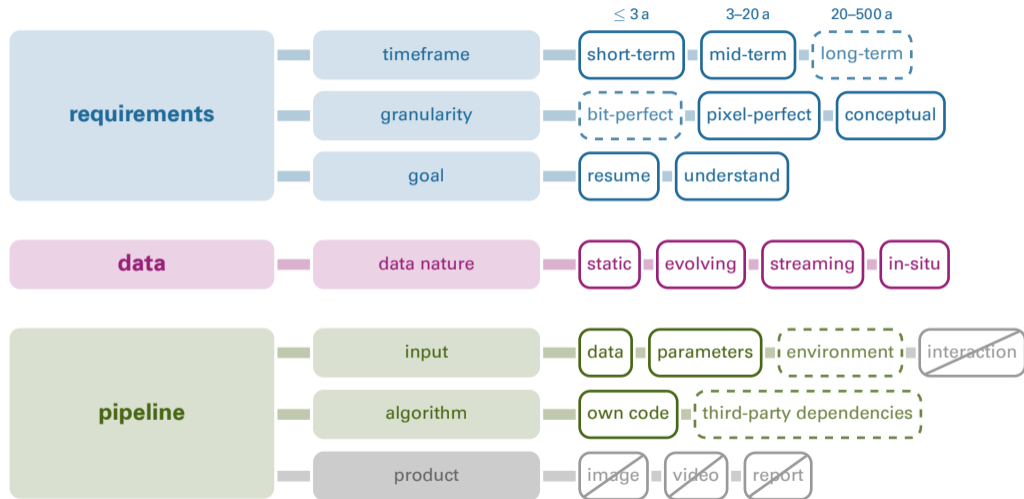
Stored datasets (long-term data repositories)



Setup cost: ★★☆☆ ←

Maintenance cost: ★★★☆☆ ←

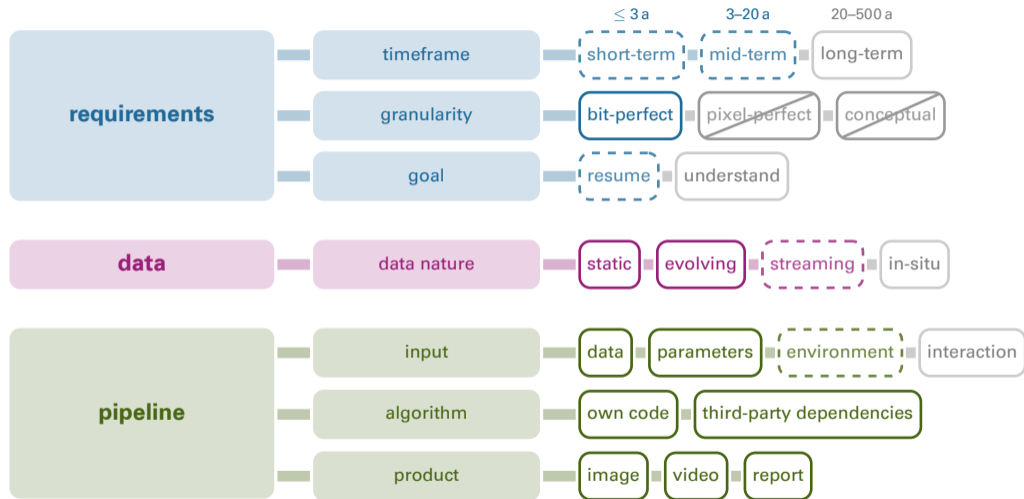
Deterministic datasets



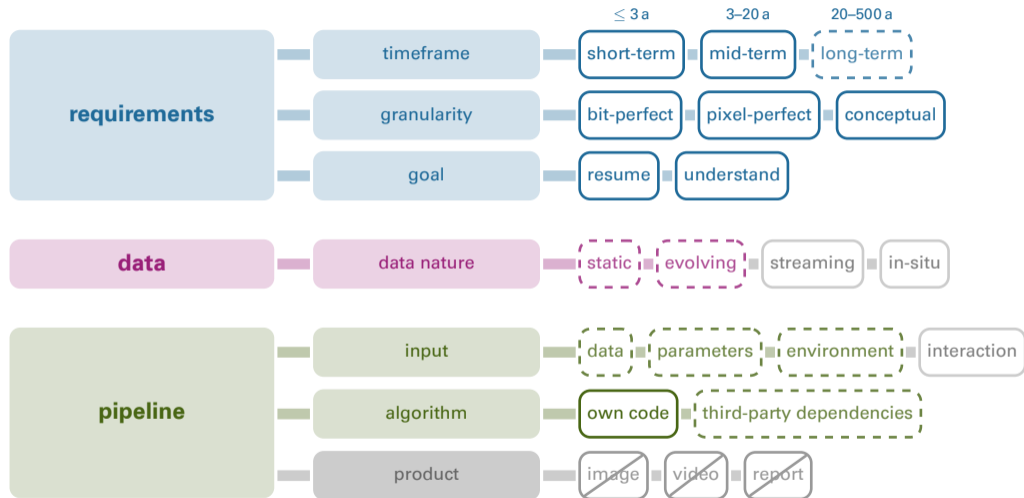
Setup cost: ★★☆☆

Maintenance cost: ☆☆☆

File hashes



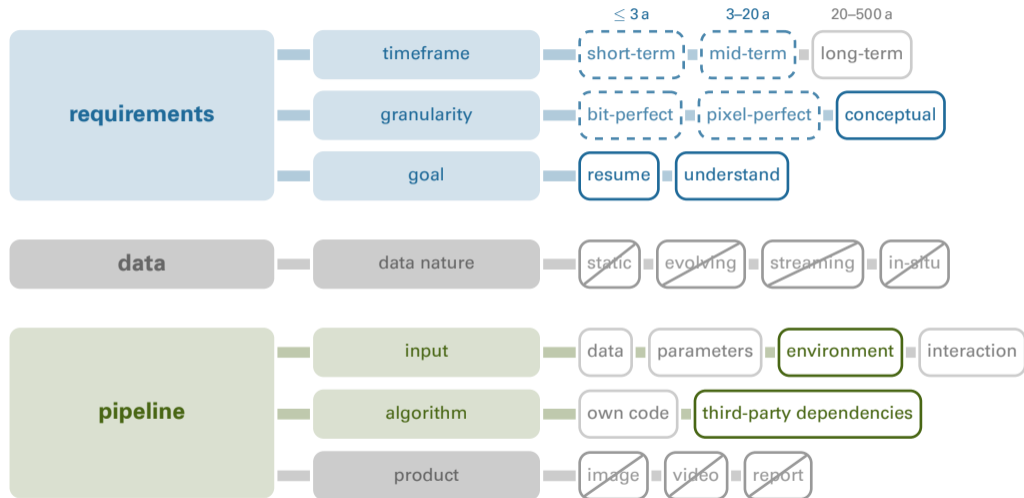
(Distributed) Version control



Setup cost: ★☆☆

Maintenance cost: ★☆☆

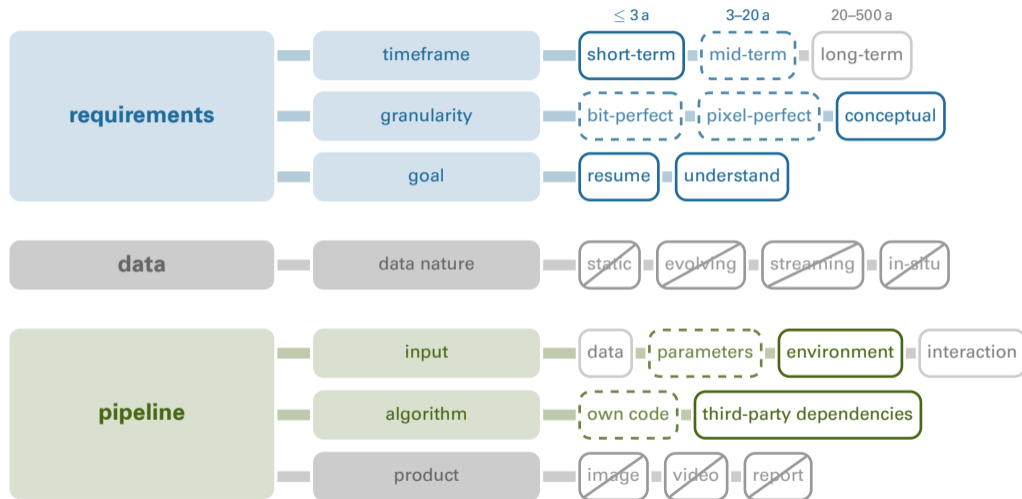
Pinning dependency versions (e.g., vcpkg.json, package-lock.json, requirements.txt)



Setup cost: ★☆☆

Maintenance cost: ★★☆☆

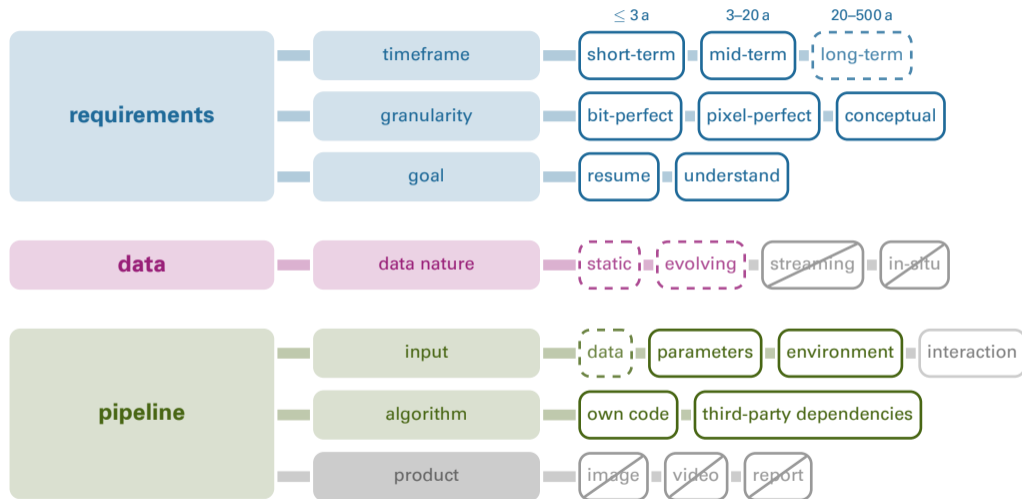
Storing a description of the system state (e.g., Dockerfile)



Setup cost: ★☆☆

Maintenance cost: ★★☆☆

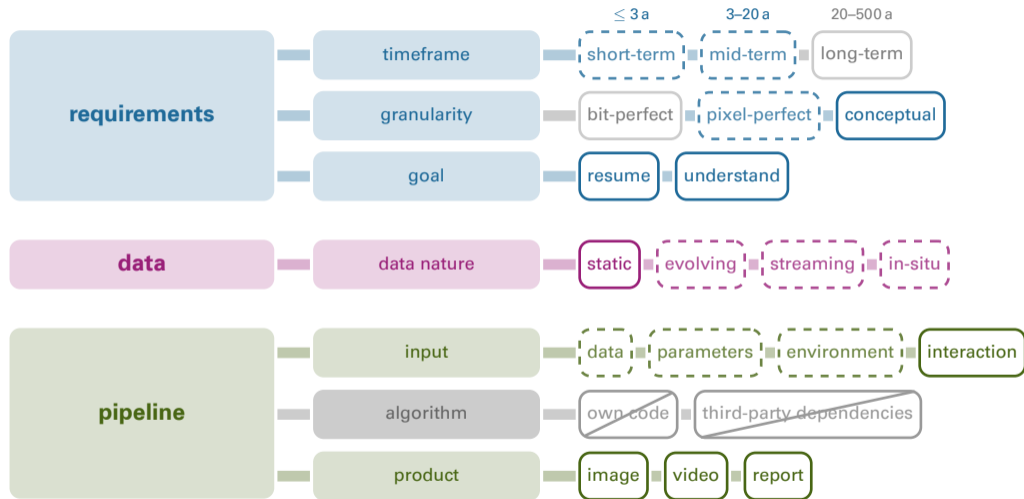
Storing the entire virtualized system (e.g., Docker image, VM)



Setup cost: ★★☆☆

Maintenance cost: ★☆☆☆

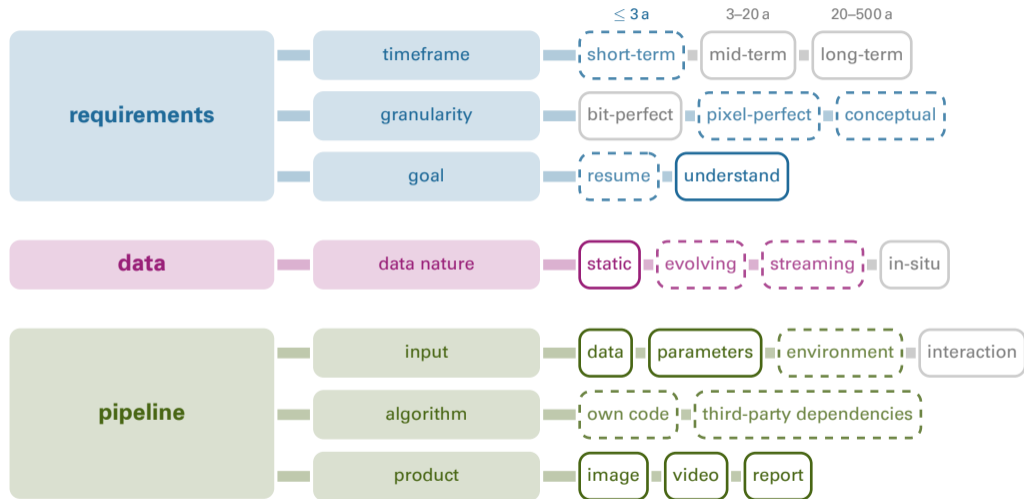
Interaction logging



Setup cost: ★★☆☆

Maintenance cost: ★☆☆☆

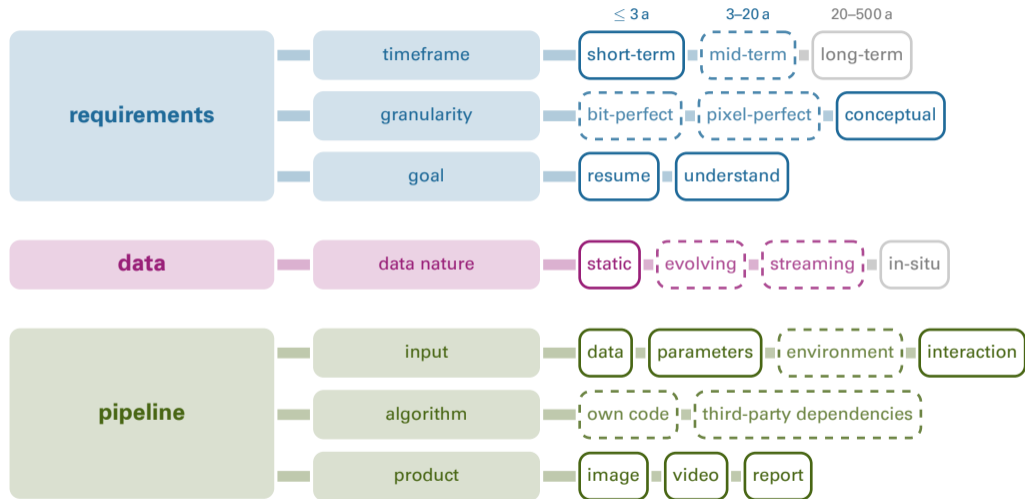
Storing pipeline metadata inside the product



Setup cost: ★★☆☆

Maintenance cost: ★☆☆☆

Linking to bundled metadata from the product

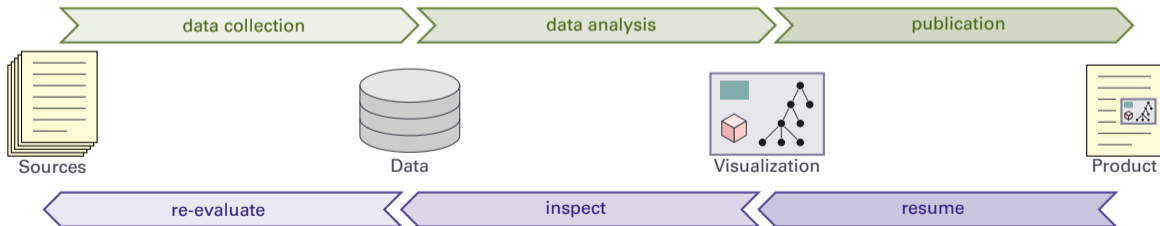
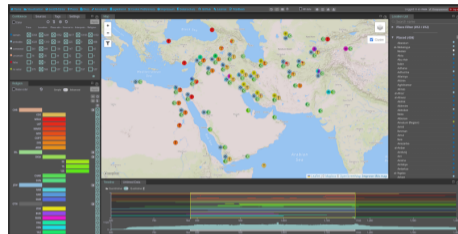


Setup cost: ★☆☆

Maintenance cost: ★★☆☆

Damast³

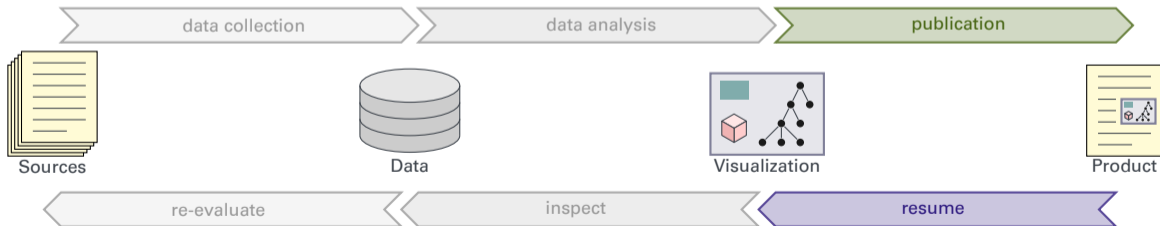
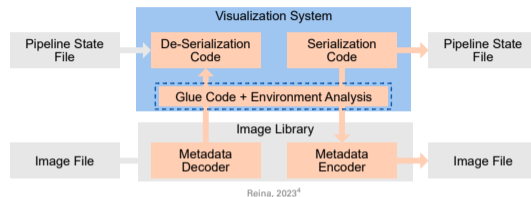
- co-existence of religions in the medieval Middle East
 - collection of data from many text sources
 - visual analysis
 - export and share findings
- backwards direction fully supported
- workflow in multiple, iterative cycles
- data in LTR, source code on GitHub
- server code in Docker image



³Max Franke and Steffen Koch. "Damast: A visual analysis approach for religious history research." In: *Proc. IVAPP*. SciTePress, Feb. 2023, pp. 40–52.
DOI: 10.5220/0011609700003417

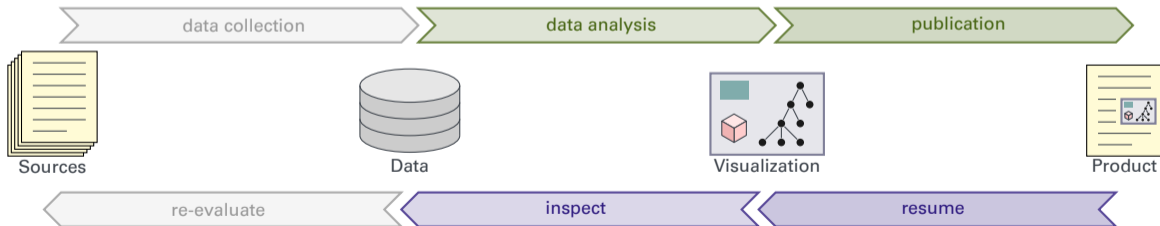
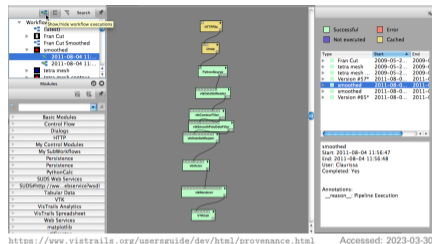
Image Metadata⁴

- focused on images as products of visualization
- generalized with regard to visualization software, prototyped in MegaMol
- includes entire visualization state in image metadata
- references to code, dependencies, data
→ additional external infrastructure required
- low overhead



⁴Guido Reina. “Can image data facilitate reproducibility of graphics and visualizations? Towards a trusted scientific practice.” In: *IEEE CG&A* 43.2 (Mar. 2023), pp. 89–100. DOI: 10.1109/MCG.2023.3241819. To be presented at EuroVis 2023 in Leipzig, Germany (June 12–16, 2023).

- formal specification format for the visualization pipeline
- long-term reproducibility was not an explicit goal
 - data referenced via file paths or URLs
 - little control over used hardware and third-party software
- powerful for
 - resumability
 - parameter and interaction input
 - own code
- no longer maintained



⁵L. Bavoil et al. "VisTrails: Enabling interactive multiple-view visualizations." In: *Proc. VIS. IEEE*, 2005, pp. 135–142. DOI: 10.1109/visual.2005.1532788

Conclusion

- reproducibility is essential for open science
 - strengthens scientific rigor
 - but also benefits users: *resumability*
- reproducibility can be ensured on a wide spectrum, depending on multiple factors
 - for how long?
 - how granular?
 - for what purpose?
 - with what data?
- **Reproducibility should always be a consideration!**
- pragmatic measures already constitute a great improvement

Slides: <https://dx.doi.org/10.5281/zenodo.7846110>



European Union *Horizon 2020* project
grant agreement no. 101004825

