



Research Data Day 2025

Experience report: Developing and implementing a research data management strategy and policy at the Institute for Combustion Technology

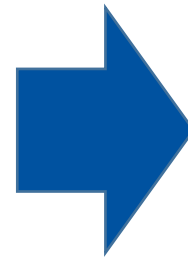
Raymond Langer, Fabian Fröde, Marianna Cafiero, Heinz Pitsch
Institute for Combustion Technology, RWTH Aachen University

11/18/2025

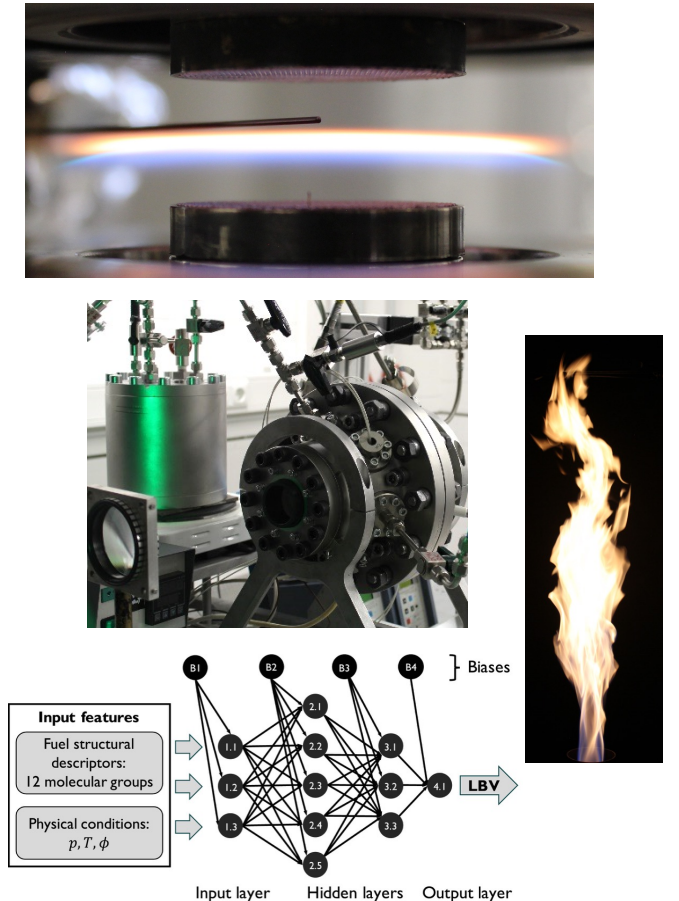
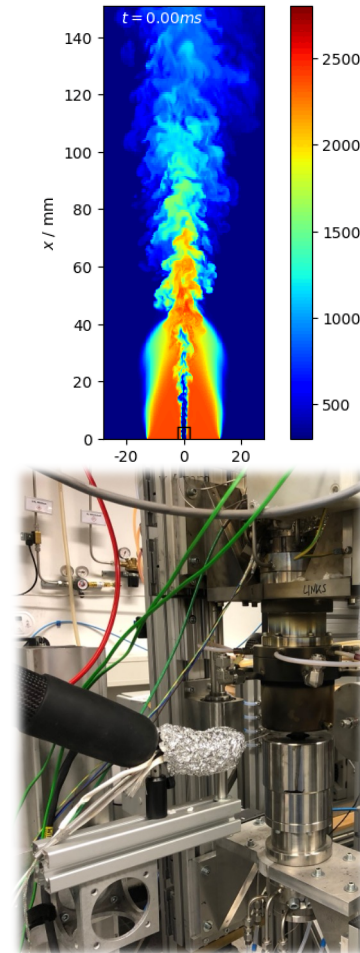
From tools to practice: making RDM work at ITV



Tools & Ideas



What does it take to move from tools and ideas to practice?

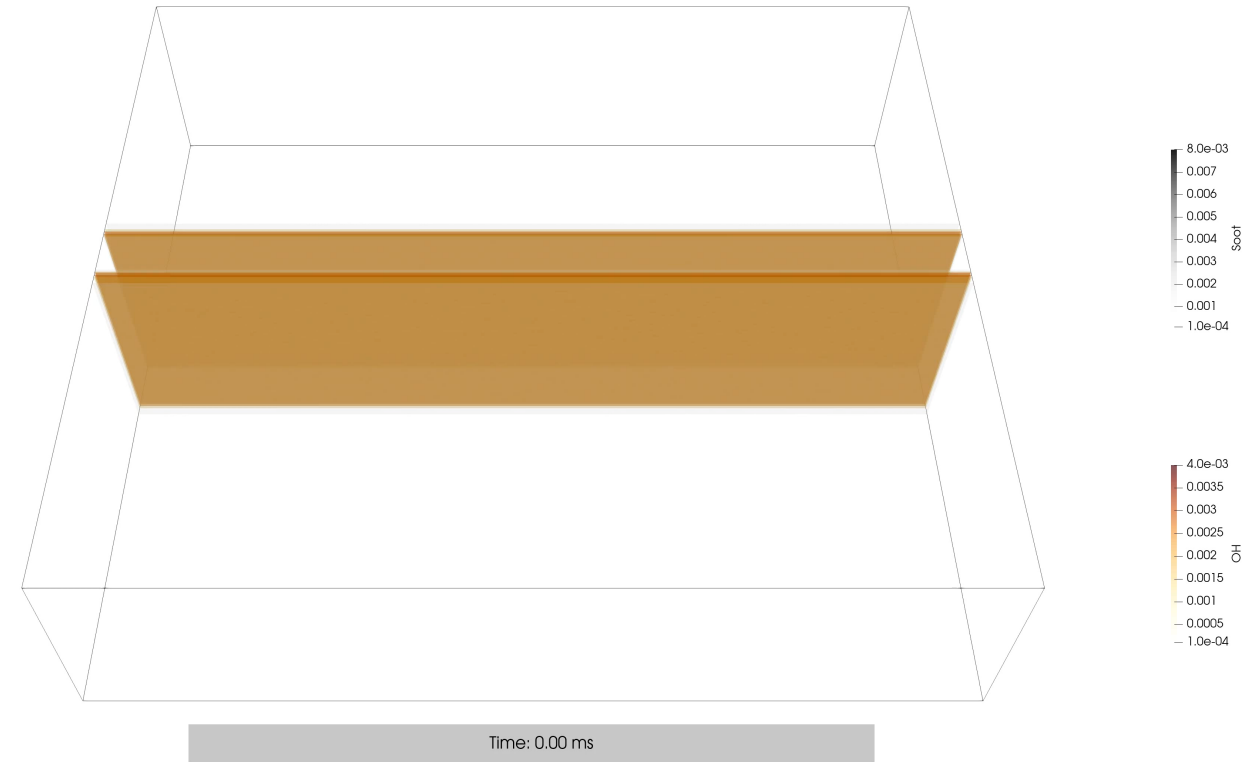


Practice

Concrete example of scientific work at the Institute for Combustion Technology



- 💧 Soot evolution in turbulent flows characterized by a large range of time scales
- 💧 Complex non-linearities due to turbulence-chemistry interaction
- 💧 Soot break-through (“smoking flame”) due to local extinction induced by turbulence
- 💧 High-fidelity 3D data for modeling better predictive soot emission models
- 💧 ~360 million grid points; more than 70 degrees of freedom
- 💧 11 million core-hours per full dataset



Soot mass generated inside the flame and the flame front location, using the hydroxyl radical (OH, yellow) as a marker.

Scialabba et al., Combust. Flame, (2025) 114093.

Concrete example of scientific work at the Institute for Combustion Technology



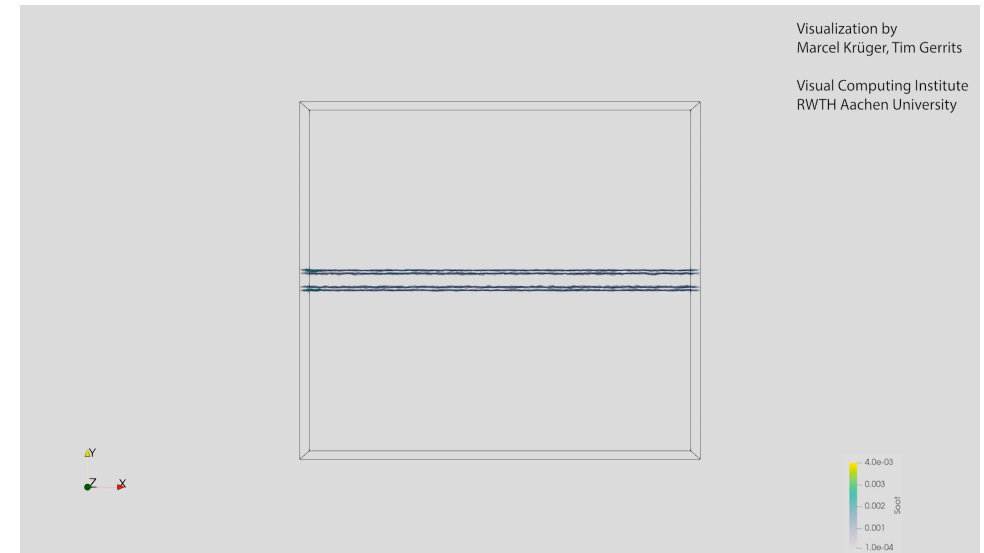
DATASET SIZE

Dataset size

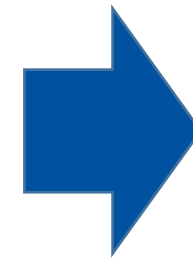
- Full dataset ~ 120 TB
- Two additional datasets with different soot breakthrough levels are available

Challenges

- Accessibility outside of HPC centers
- Storage and backup
- Efficient transfer between HPC systems or other systems to share the data for scientific purposes
- More efficient retrieval of a specific data subset without having to load the full dataset (“on-demand” extraction)



Scialabba et al., Combust. Flame, (2025) 114093.



Address challenges?

Cosine

Many approaches → one coordinated strategy



25+ PHD STUDENTS & POSTDOCS, DIVERSE WORKFLOWS

What we do

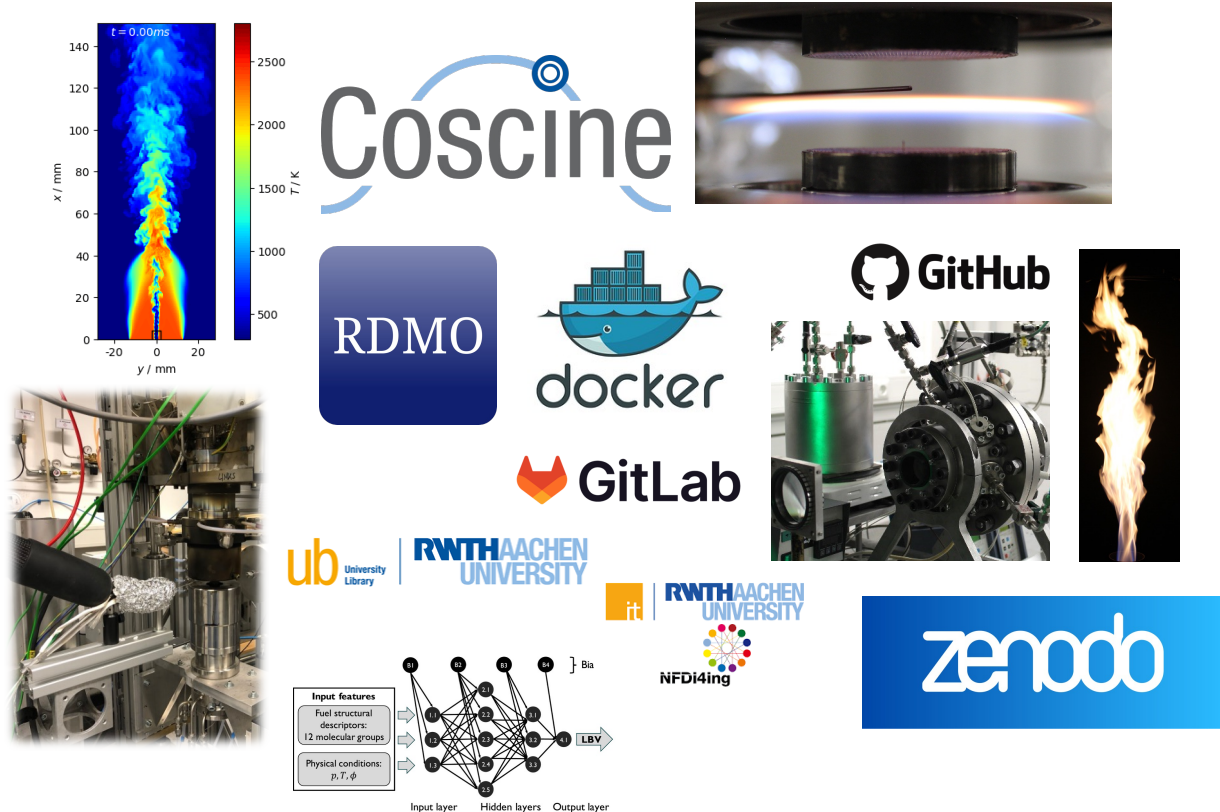
- Modeling
- Experiments
- Simulations

Tools and services we use

- Coscine, RWTH GitLab
- RDMO, RWTH publications, Zenodo, GitLab CI, Docker

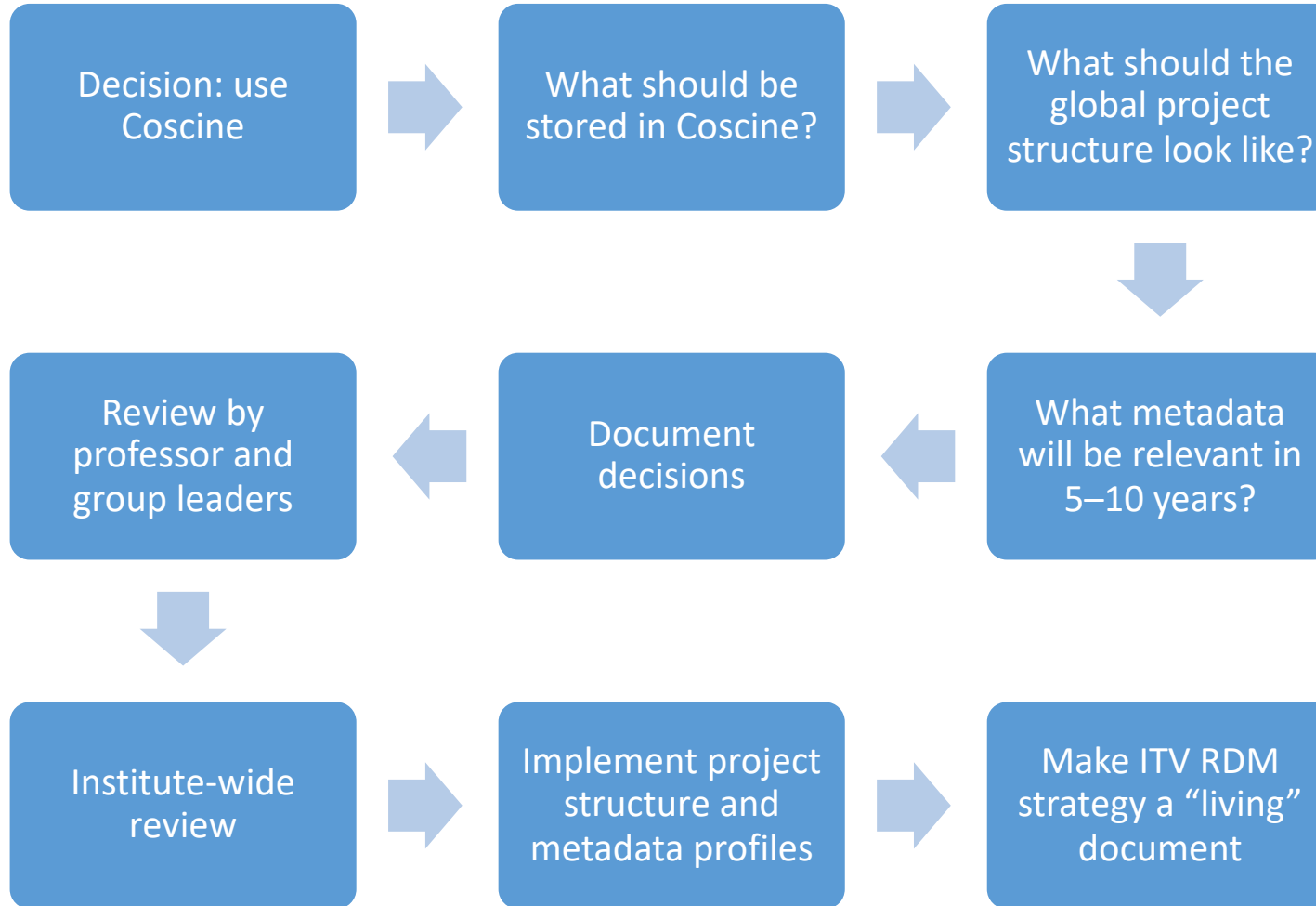
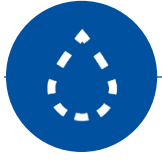
An institute-wide RDM strategy

- Share proven practices
- Reuse and align
- Continuous improvement
- Ensure long-term findability & accessibility



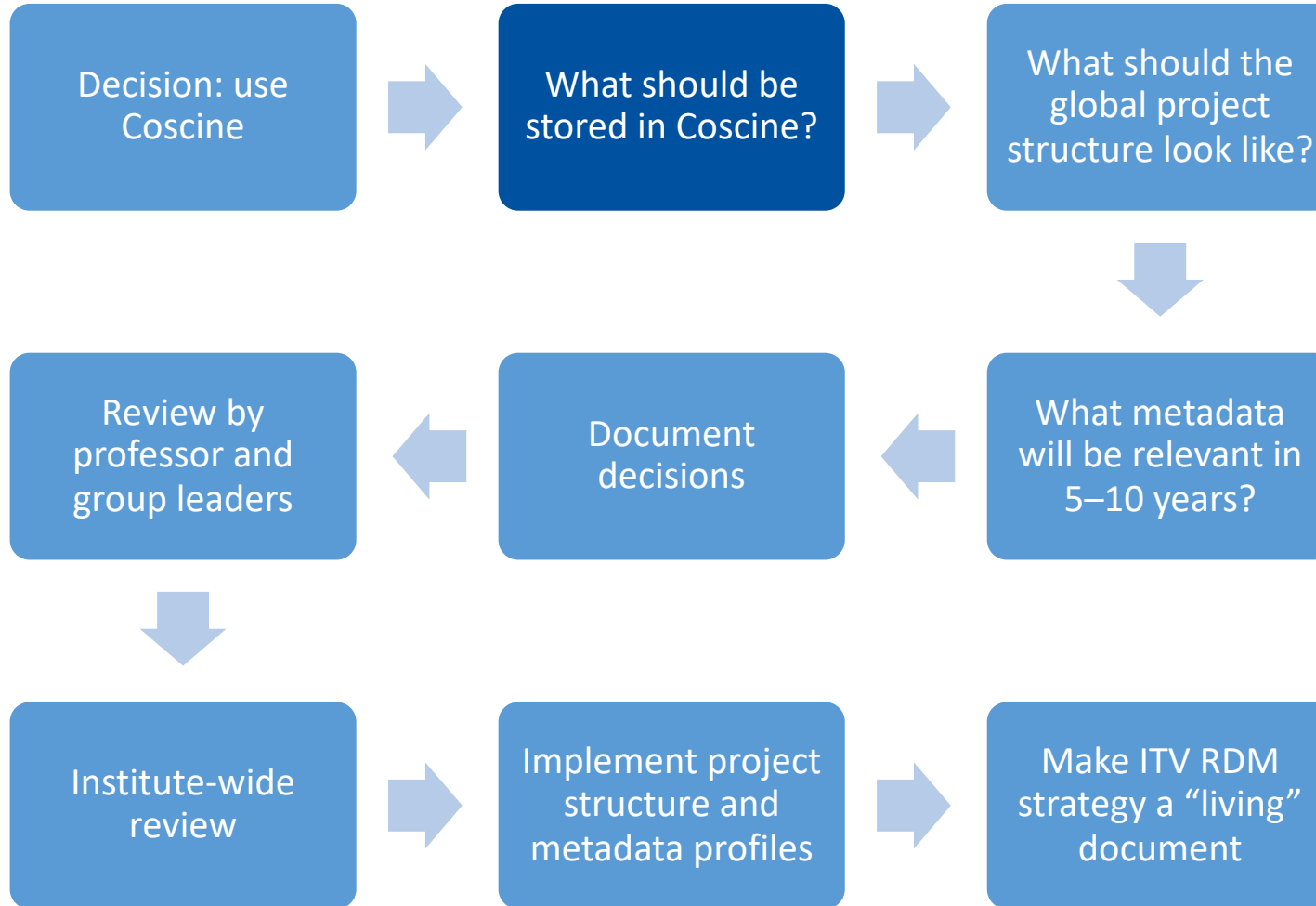
How can we coordinate these efforts?

Our approach: from “use Coscine” to a coordinated strategy



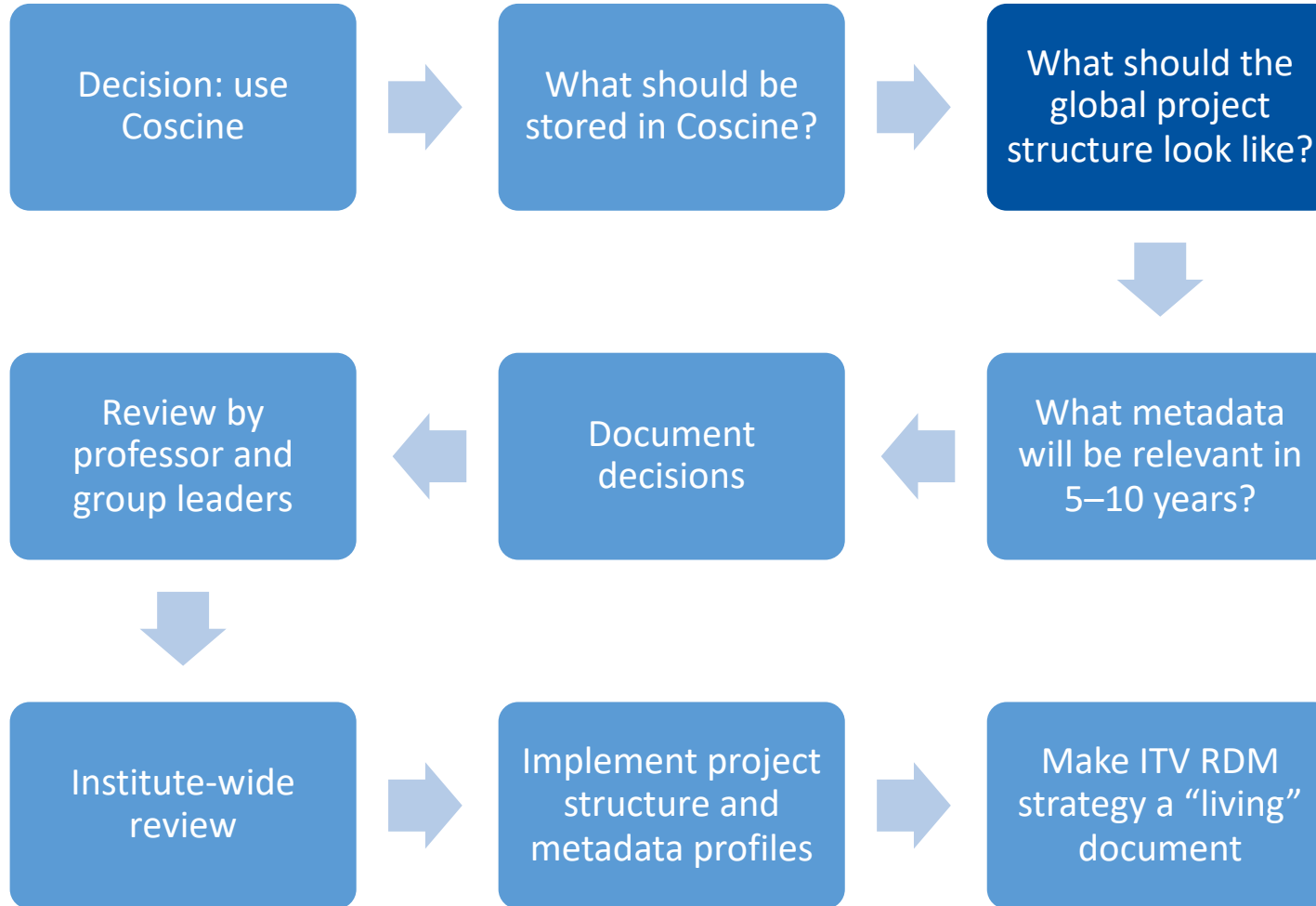
- 💧 No polished process initially
- 💧 Strategy emerged while learning Coscine

Our approach: from “use Coscine” to a coordinated strategy



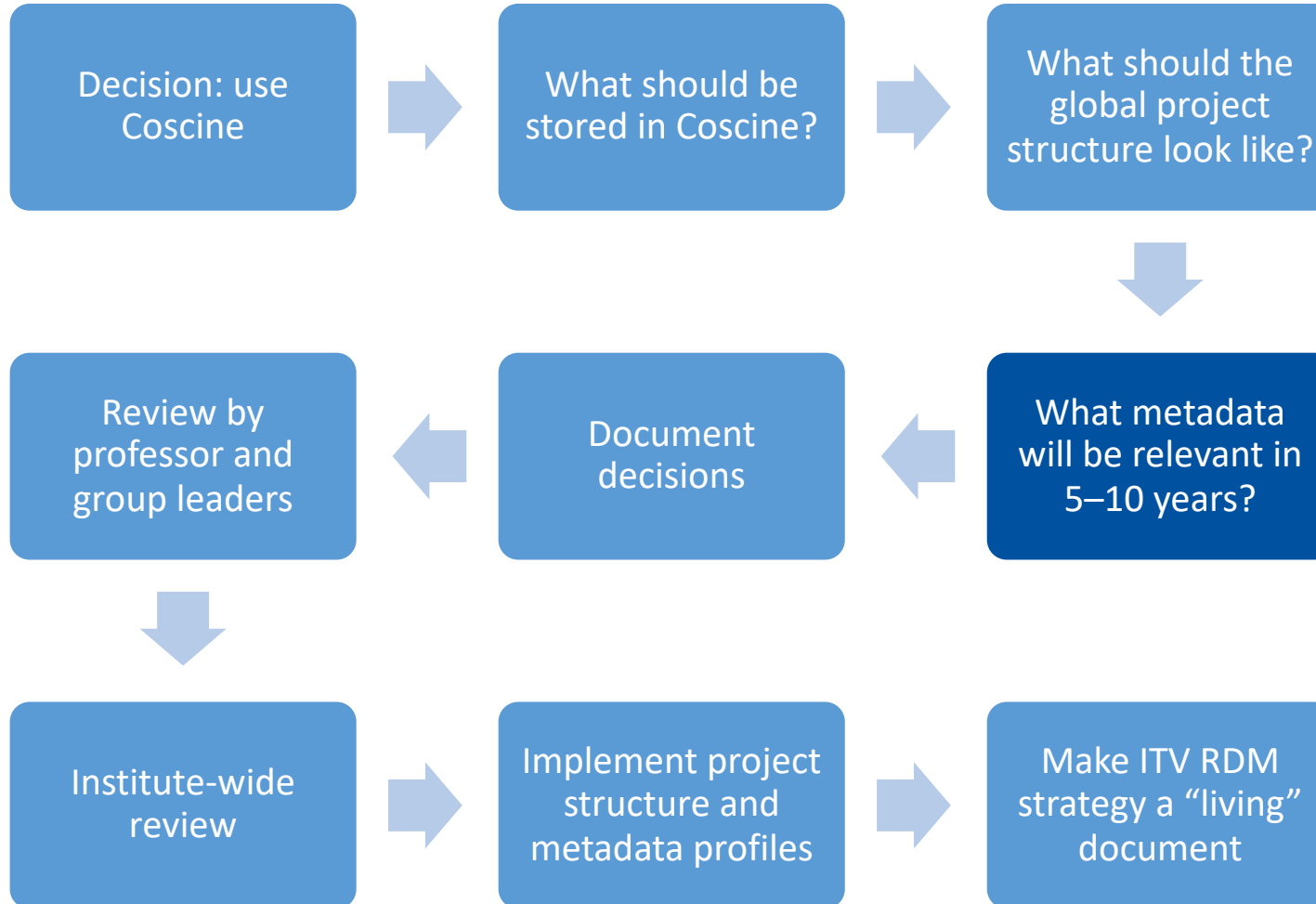
- 💧 No polished process initially
- 💧 Strategy emerged while learning Coscine
- 💧 Decision: keep ITV servers for operational/experimental workflows
- 💧 Use Coscine for curated, shareable datasets + metadata

Our approach: from “use Coscine” to a coordinated strategy



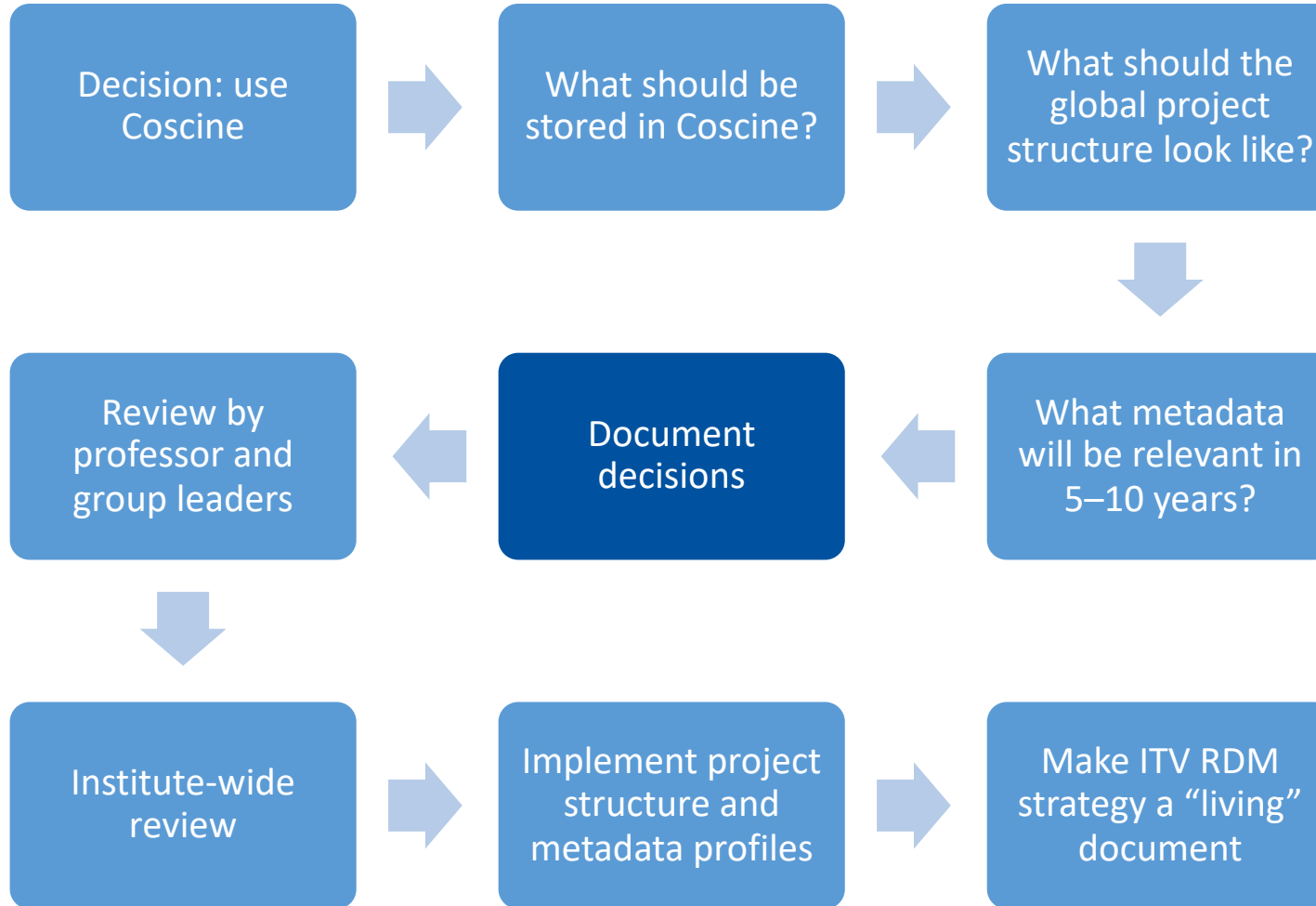
- 💧 No polished process initially
- 💧 Strategy emerged while learning Coscine
- 💧 Decision: keep ITV servers for operational/experimental workflows
- 💧 Use Coscine for curated, shareable datasets + metadata
- 💧 Coscine is hierarchical (tree)
- 💧 Top level: person-centric (employees/alumni) to aid long-term recall

Our approach: from “use Coscine” to a coordinated strategy



- ◊ No polished process initially
- ◊ Strategy emerged while learning Coscine
- ◊ Decision: keep ITV servers for operational/experimental workflows
- ◊ Use Coscine for curated, shareable datasets + metadata
- ◊ Coscine is hierarchical (tree)
- ◊ Top level: person-centric (employees/alumni) to aid long-term recall
- ◊ Domain experts decide what is most relevant
- ◊ Constraint: Metadata profile cannot be deleted

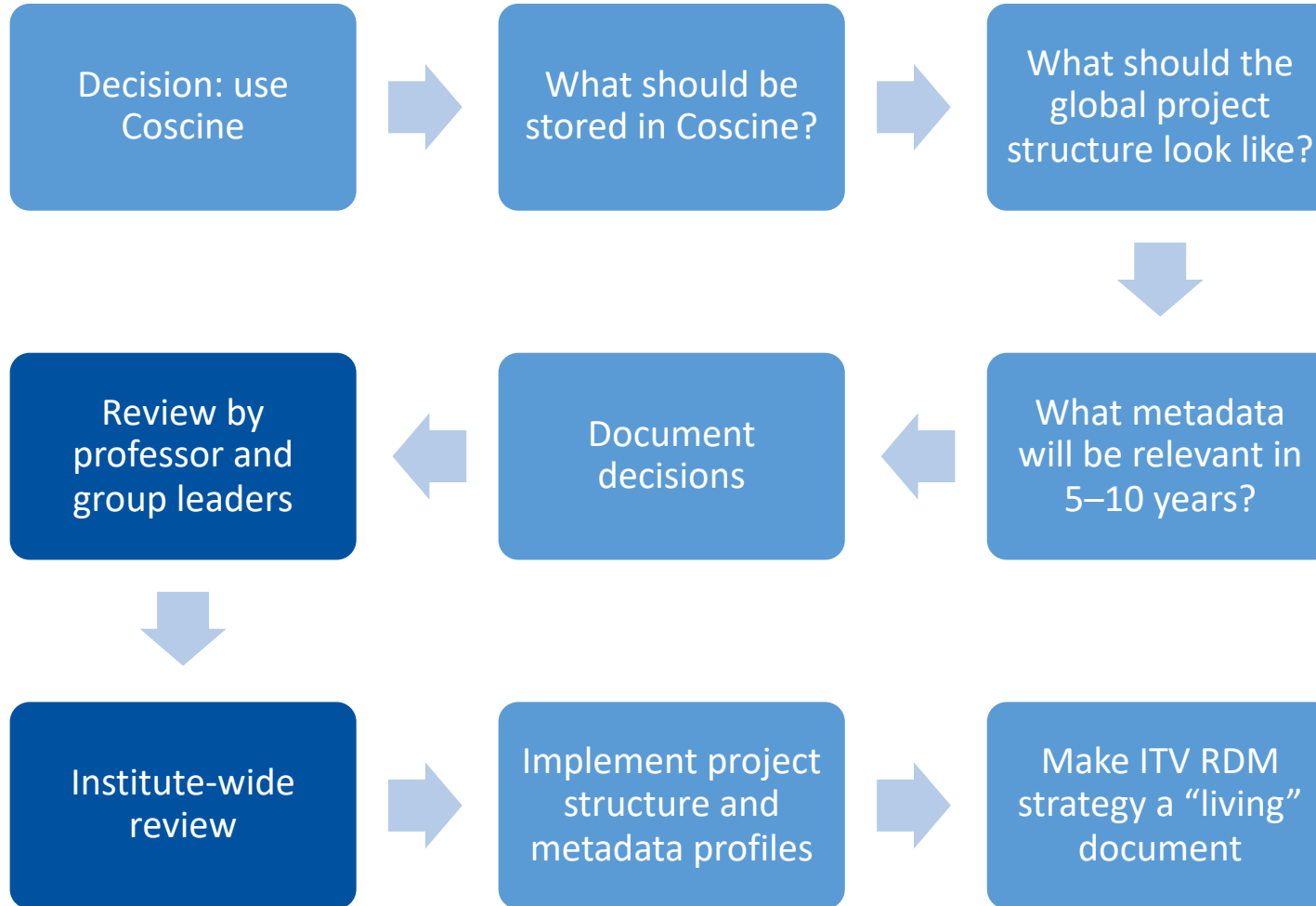
Our approach: from “use Coscine” to a coordinated strategy



💧 Idea of an institute-wide RDM strategy emerged

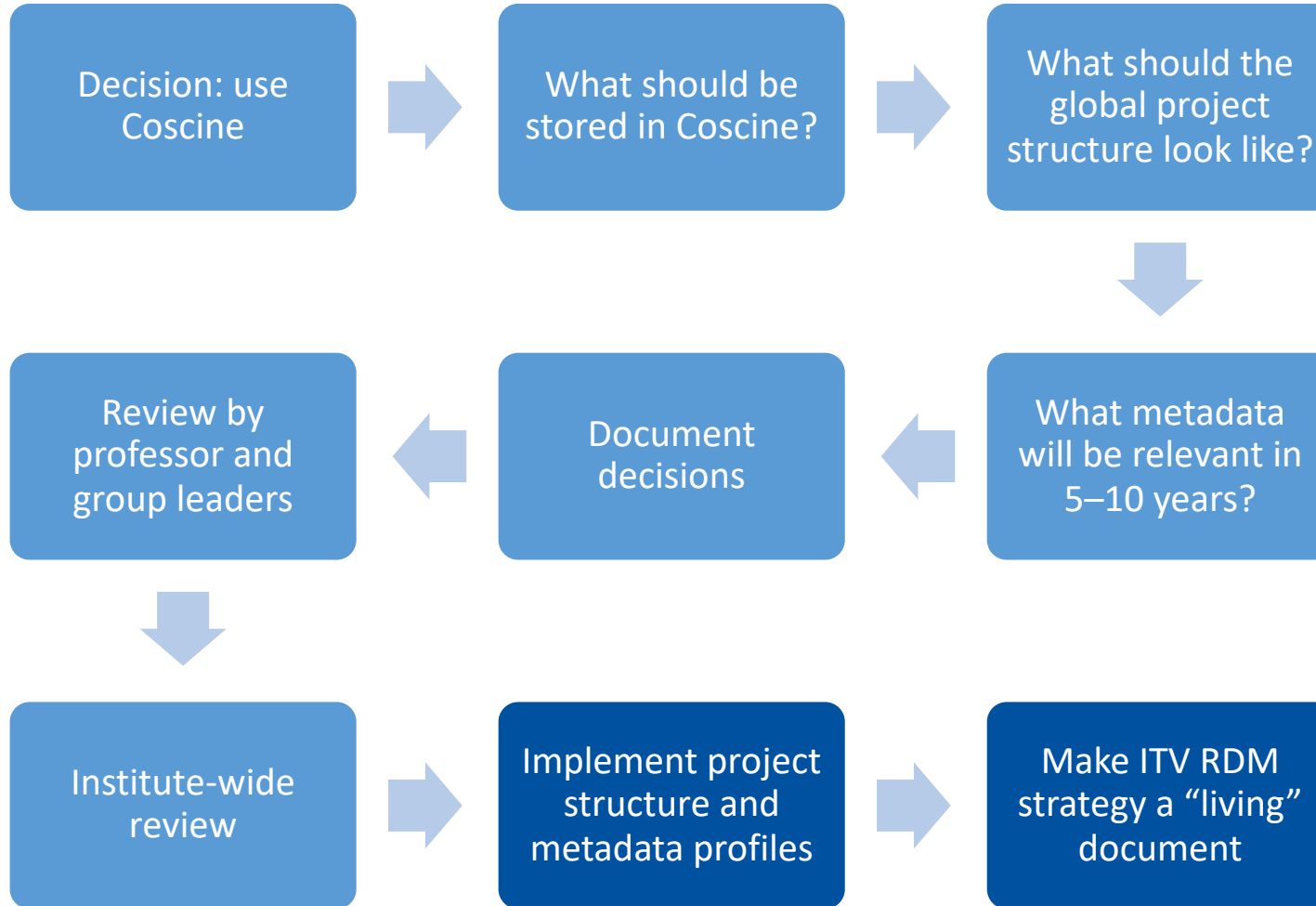
- 💧 Discussion already covered a broad scope (processes, IT-infrastructure, domain knowledge, ...)
- 💧 Extend scope of documentation to all RDM-related topics

Our approach: from “use Coscine” to a coordinated strategy



- 💧 Idea of an institute-wide RDM strategy emerged
 - 💧 Discussion already covered a broad scope (processes, IT-infrastructure, domain knowledge, ...)
 - 💧 Extend scope of documentation to all RDM-related topics
- 💧 First draft was prepared by a small team
 - 💧 Group leaders + data stewards
 - 💧 Experienced PhD students
- 💧 → Feedback from ~25 researchers

Our approach: from “use Coscine” to a coordinated strategy



- 💧 Idea of an institute-wide RDM strategy emerged
 - 💧 Discussion already covered a broad scope (processes, IT-infrastructure, domain knowledge, ...)
 - 💧 Extend scope of documentation to all RDM-related topics
- 💧 First draft was prepared by a small team
 - 💧 Group leaders + data stewards
 - 💧 Experienced PhD students
- 💧 → Feedback from ~25 researchers
- 💧 Ongoing: Implementation
- 💧 Iterate as we learn

Guidelines for Handling Research Data at the Institute for Combustion Technology



MAIN GUIDELINES

💧 RDM concerns everyone at ITV

💧 Contribute to the RDM policy

1. All projects require a data management plan (DMP)
2. All source code and other text-based data must be managed in GitLab
3. All research data must be in Coscine

→ Three sections

1. DMPs in RDMO
2. ITV Groups in GitLab and GitLab
3. Coscine

Core Guideline

Research data management (RDM) concerns everyone at ITV. All personnel are expected to familiarize themselves with this policy and to implement the procedures explained according to their role. We encourage active participation in the periodic review and enhancement of this policy to ensure its relevance and effectiveness.

Guideline 1

A data management plan maintained in the Research Data Management Organizer (RDMO) is required for each project.

Guideline 2

Software developed at ITV and other text-based data that meets the license conditions and project-dependent confidentiality requirements must be managed in a git repository in the RWTH GitLab instance.

Guideline 3

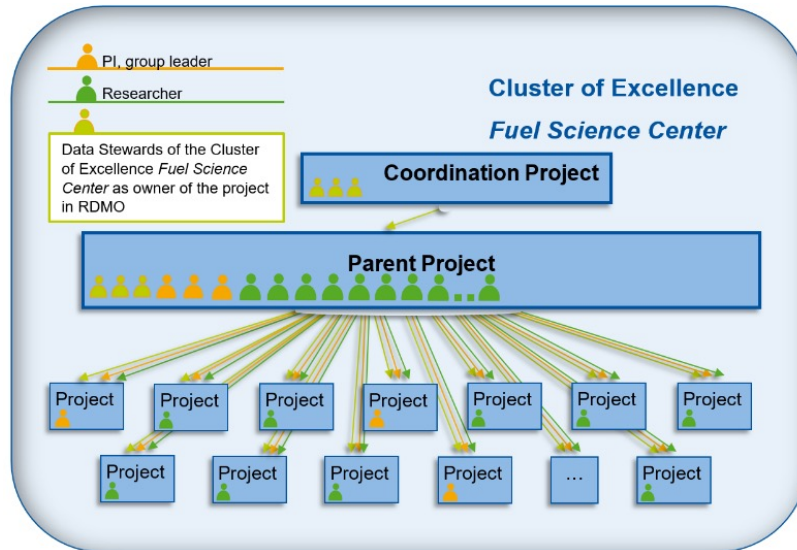
All research data, including L^AT_EX source code and figures for publications, must be stored or referenced and documented in the Collaborative Scientific Integration Environment (Coscine).

Data management plans in the research data management organiser



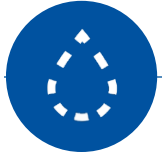
ADOPTED THE FSC APPROACH

- ITV catalogue and parent project
- Advantage: Centrally stored DMPs with minimal effort



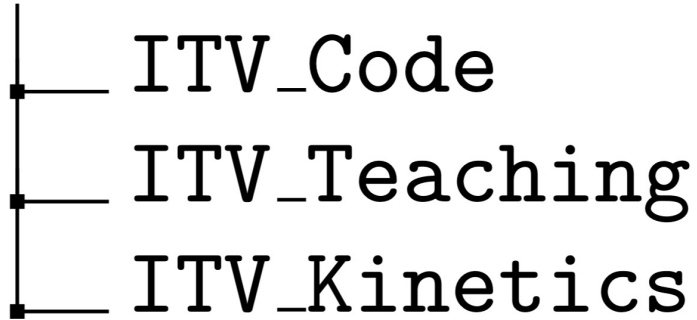
RDMO

Schönau, S., Cafiero, M., Kröckert K.W., Meißner, D., Langer, R., Gauding, M., Pitsch, H. Parent Projects in RDMO – How hierarchies in the Research Data Management Organiser facilitate data management in large-scale. Submitted to the Int. J. Digit. Curation.



ITV GitLab groups

RWTH GitLab/



- Improved findability, accessibility, and reuse
- Potentially extend for future work

ITV_Code / FlameMaster / Pipelines / #1833526

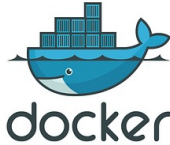
Running Raymond Langer created pipeline for commit 1245c582 6 days ago

For master

Scheduled latest branch 28 jobs In progress, queued for 13 seconds

Pipeline Jobs 28 Tests 0

prerequisites	build	unitfm	testsc	testfm	release
- CeOSPrerequisites - UbuPrerequisites	- CeOSBuild - UbuBuild	- CeOSTestFMUnit - UbuTestFMUnit	- CeOS_SM_FM01 - CeOS_SM_FM02 - Ubu_SM_FM01 - Ubu_SM_FM02	- CeOS_FM_0D - CeOS_FM_DiffPlugFlow - CeOS_FM_DiffSteady - CeOS_FM_DiffUnsteady - CeOS_FM_PSR - CeOS_FM_PremUnstretched - CeOS_FM_Spherical - Ubu_FM_0D - Ubu_FM_DiffPlugFlow - Ubu_FM_DiffSteady - Ubu_FM_DiffUnsteady - Ubu_FM_PSR - Ubu_FM_PremUnstretched - Ubu_FM_Spherical	- CeOSDockerRelease - SrcRelease - UbuDockerRelease

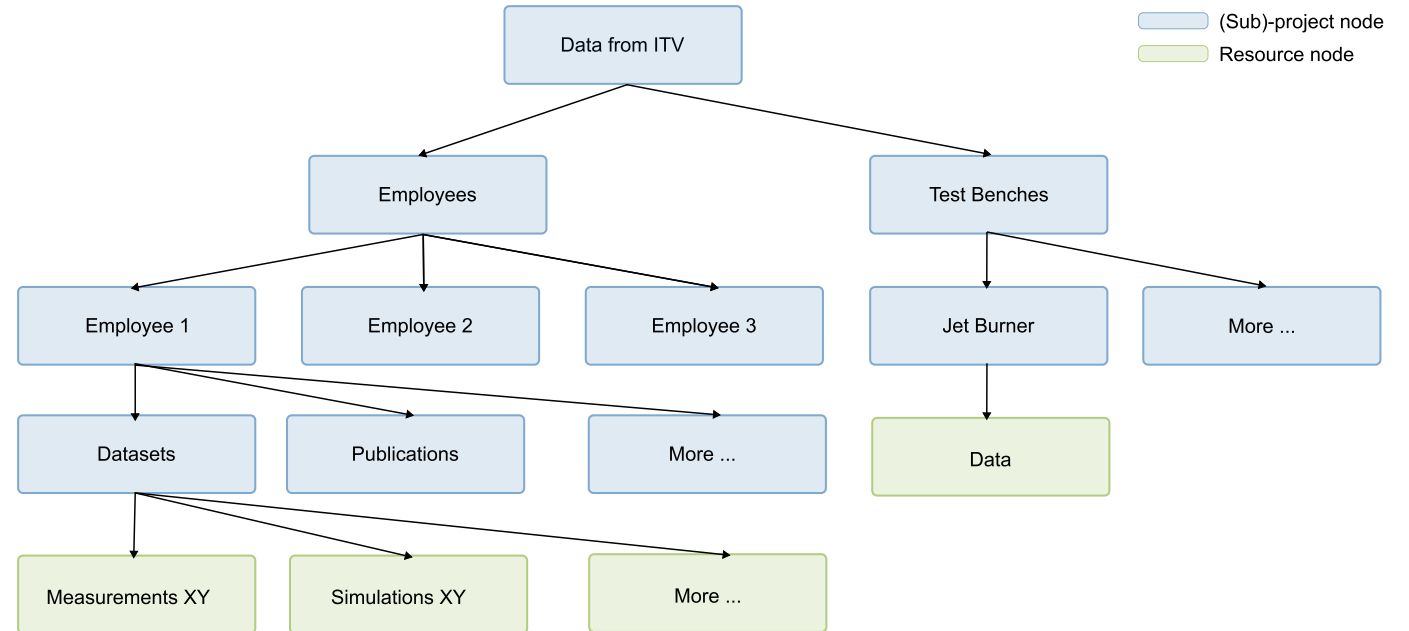
 **GitLab** +  + 

Reproducible, automated tests in containerized environments



Coscine section

- 💧 Brief introduction
- 💧 References to the (excellent) Coscine documentation
- 💧 JARDS platform proposals (> 100 GB)
- 💧 Integration with ITV processes
 - 💧 Internal peer-review
 - 💧 Group leader's responsibilities
- 💧 Metadata profiles

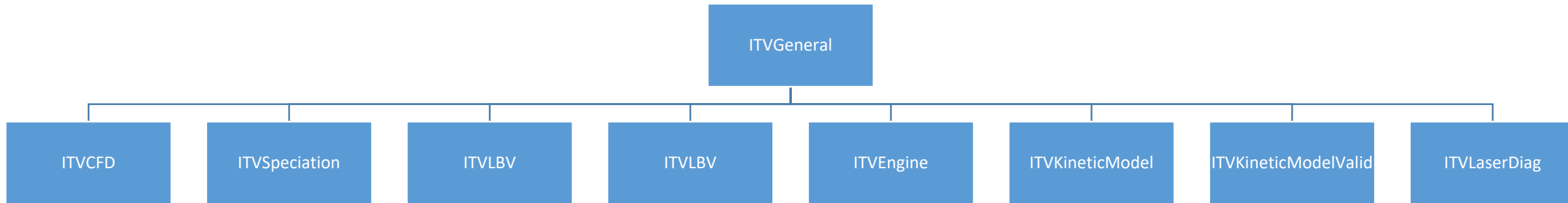


Global ITV project structure



- 💧 Coscine resources require a metadata profile
- 💧 Profiles are reviewed; they can't be deleted
 - design carefully
- 💧 All profiles start with "ITV"
 - Facilitates selection in Coscine

Our approach: Keep required fields minimal



ITV Metadata profiles: ITVGeneral and ITVKineticModel



Table 2: Metadata profile: ITVGeneral.

Field Name	Description
Title	Provide a descriptive title for the data
Contact person	Provide the name (Firstname Surname, e.g., Max Mustermann) of the contact person
Producer / Author	List of all names of persons involved (Firstname Surname, e.g., Max Mustermann)
Keywords / Tags	Provide keywords or tags that describe the data
Legal information	
Date of creation	Provide the date of creation (Format: mm-yyyy, e.g., 01-2024) Provide start and end month if creation took longer than one month (Format: mm-yyyy - mm-yyyy)
Status publication	Select one of the options: not relevant for publication, in preparation, submitted, published
Funding	Provide the acronym of the funding project(s). A list of all available acronyms can be found on the ITV wiki.
Reference to paper	If the data is published, provide authors, title, journal, year, and doi. e.g., F. Fröde, T. Grenga, S. Dupont, R. Kneer, R. Tischendorf, O. Massopo, H-J. Schmid, H. Pitsch, Large eddy simulation of iron oxide formation in a laboratory spray flame, Applications in Energy and Combustion Science, 2023, 10.1016/j.jaecs.2023.100191
Comments / Explanation	Provide comments and explanations related to the data.
Goal	Provide the scientific goal of the data.

Screenshot ITV RDM policy

Table 7: Metadata profile: ITVKineticModel.

Field Name	Description
Repository kinetic model	Provide the repository for the kinetic model. For example, git@git.rwth-aachen.de:ITV/mech.git
Subdirectory	Provide the subdirectory in the git repository that contains the chemkin files and species dictionary, e.g., SootMechanisms/ITV_2015/Langer2021/kritika.liming/soot.langer
Git hash kinetic model	Provide the git hash of the kinetic (tags are preferred).
Repository validation database	Provide the link to the validation database in Cosine
Git hash validation database	Provide the git hash of the commit used for the validation of the kinetic model.
Fuels	State their common name(s) and the name used in the model
Validation targets	This field provides a summary of the validation targets considered. Possible values in the list provided in the field are ignition delay time data from shock tubes (IDT(ST)), IDTs from rapid compression machines (IDT(RCM)), laminar burning velocities (LBV), extinction strain rates (ESR), species data from counterflow flames (CF), species data from jet-stirred reactors (JSR), or species data from flow reactors (FR)

Might be too detailed

ITVGeneral

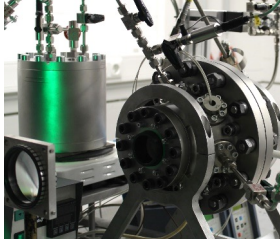
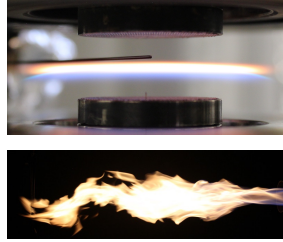
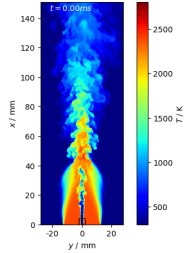
Publication status
Date of creation
Legal information
Keywords
Title
Publication title
Contact person ORCID
Publication journal / conference
Contact person email
Producer(s) / Author(s)
Contact person
Publication DOI
Comment / Explanation
Funding

Screenshot AIMS

ITVKineticModel

Repository kinetic model
Subdirectory
Git hash kinetic model
Repository validation database
Fuels
Validation targets
Git hash validation database

Do a peer-review in AIMS



Institute-wide RDM Policy

- Share proven practices
- Establish processes
- Ensure long-term findability & accessibility
→ Coscine + GitLab

- High acceptance
 - Improved processes
 - Institute-wide standard
- Time-consuming
 - Long policy





Thank you for your attention!

Raymond Langer

Institute for Combustion Technology
RWTH Aachen University
Templergraben 64
52056 Aachen
Tel.: +49 241 80 94607

Internet: www.itv.rwth-aachen.de

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence 2186 “The Fuel Science Center” – ID: 390919832.