NEA
NUCLEAR ENERGY AGENCY

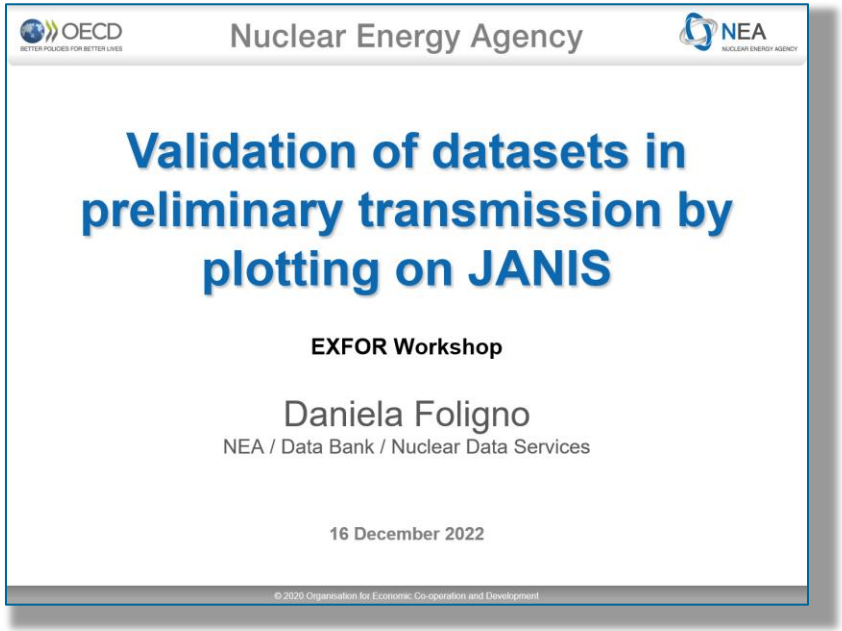
Exfor updates

New developments at the NEA

Julia Sprenger
Nuclear Energy Agency Databank

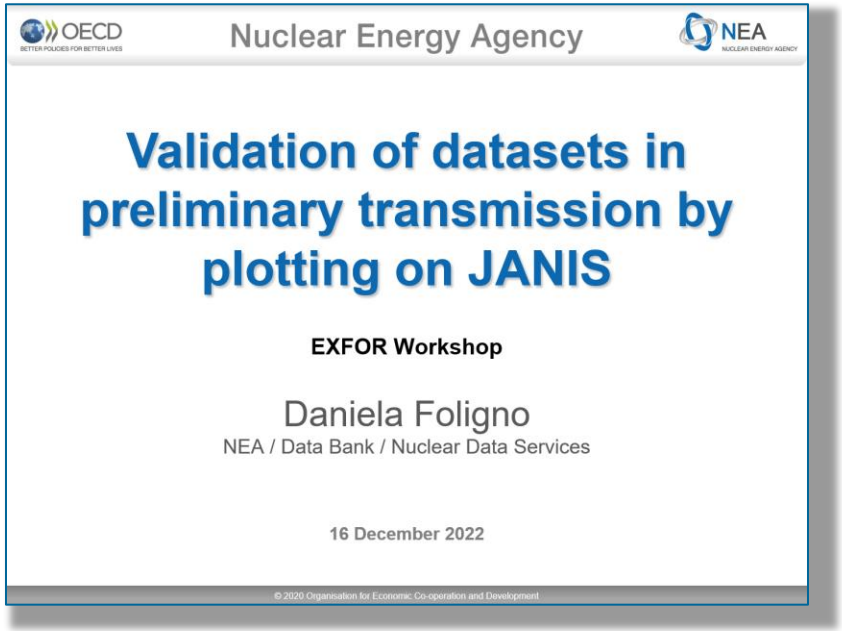
Motivation

- FAIR and Open data
  - Reproducibility
  - Transparency
  - Required by many funding organizations
- Project Management & Workflows
  - Currently requires a lot of manual steps
  - Manual tracking of versions



Motivation

- FAIR and Open data
  - Reproducibility
  - Transparency
  - Required by many funding organizations
- Project Management & Workflows
  - Currently requires a lot of manual steps
  - Manual tracking of versions

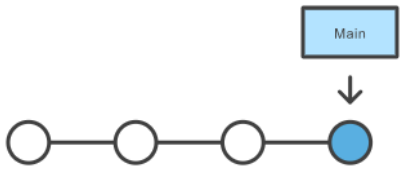


How to get Exfor ready for the next 60 years?

Modern technologies to the rescue!

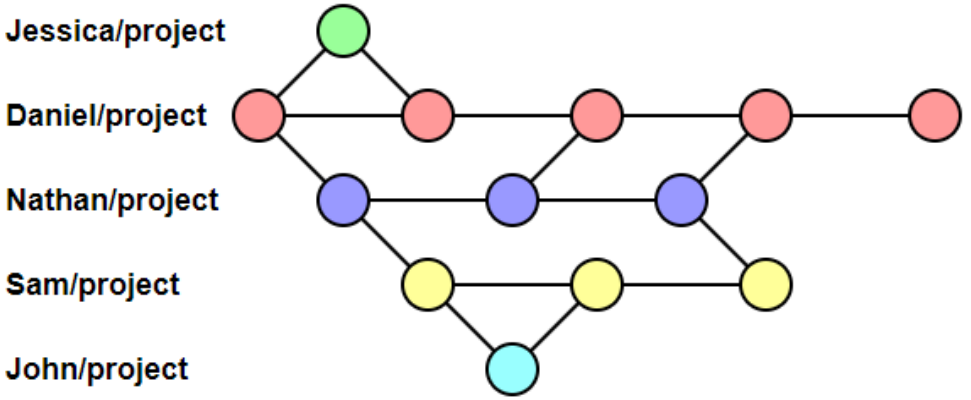
Version Control System (VCS)

- Leading VCS since 2014
- Originated from Software Development Community
- Central to Reproducibility efforts in research
- Concept
  - **Distributed** system
  - Linking of versions via `commits`



Commit information

- Who
- When
- Which changes
- Why
- Unique identifier



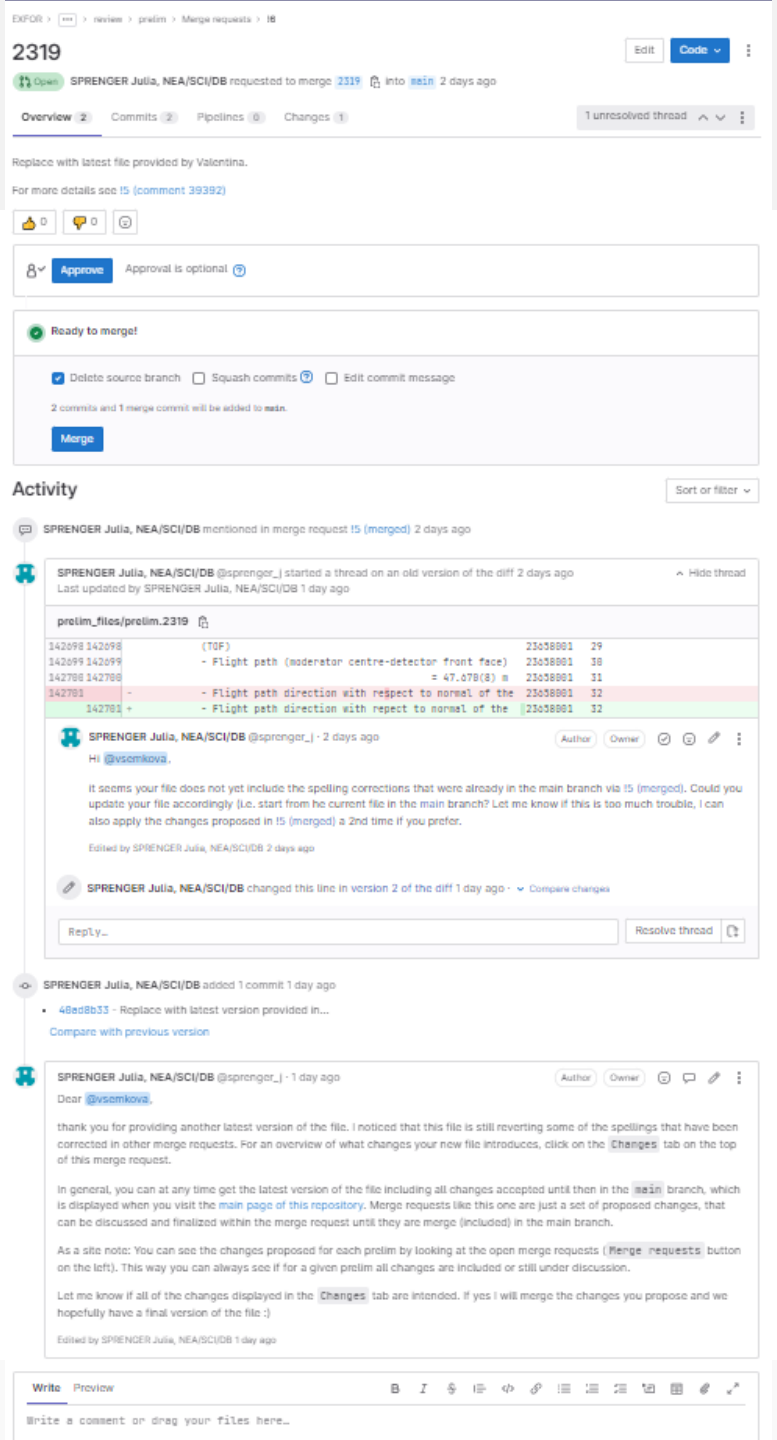
Git Logo: CC BY 3.0 Jason Long | Git commit sketch: https://www.atlassian.com/git/tutorials/rewriting-history | Git Tree: Daniel Imms, https://www.growingwiththeweb.com

Modern technologies to the rescue!

GitLab/ GitHub

Collaboration Platforms

- > 100 million / > 30 million users
- > 420 million repositories (>28 million public)
- GitLab is open source, available at gitlab.com or self-hosted
- Numerous collaboration / coordination features on top of git
  - Issue tracker
  - Discussion platform for changes (MRs)
  - Review mechanics
  - Automation of pipelines
  - Visualization of
    - Changes
    - Contributions
    - Project statistics
    - …
  - User management
  - Wiki
  - Integration with other platforms
  - Email notifications



Additional considerations

Using modern standard practices

- Makes onboarding from related communities easier
- Teaches core transferrable skills required in any scientific community
- Makes project workflow more transparent
- Reduces the Bus Factor by sharing & documenting knowledge

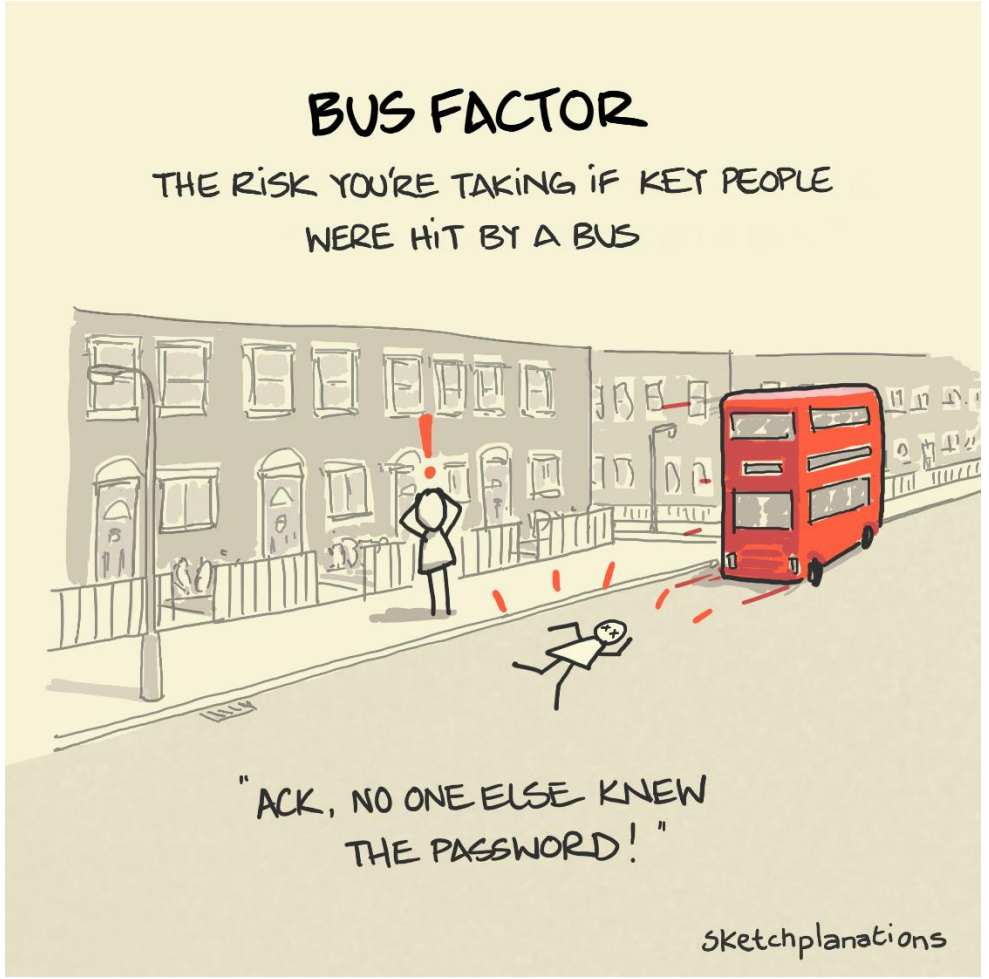


Image: Sketchplanations

A glimpse of the GitLab Webinterface

Repository view

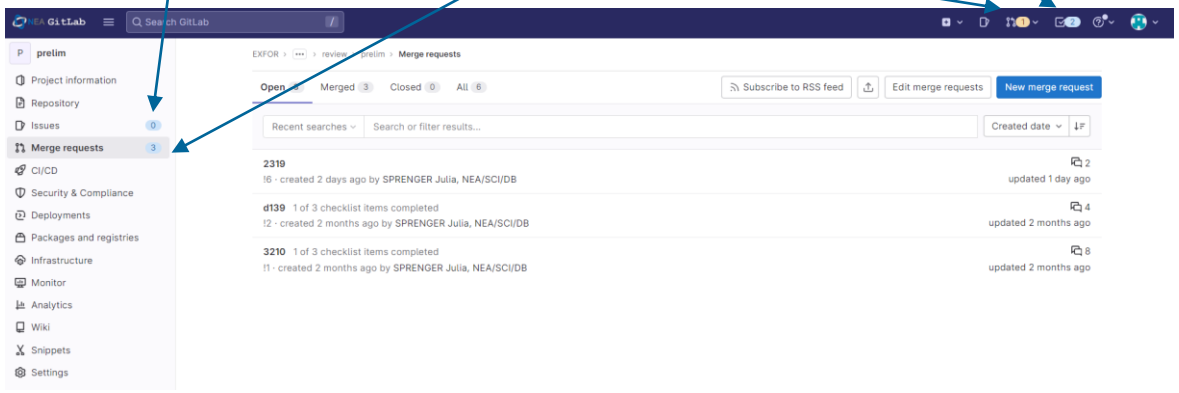
Open merge requests (MR)

Open Issues

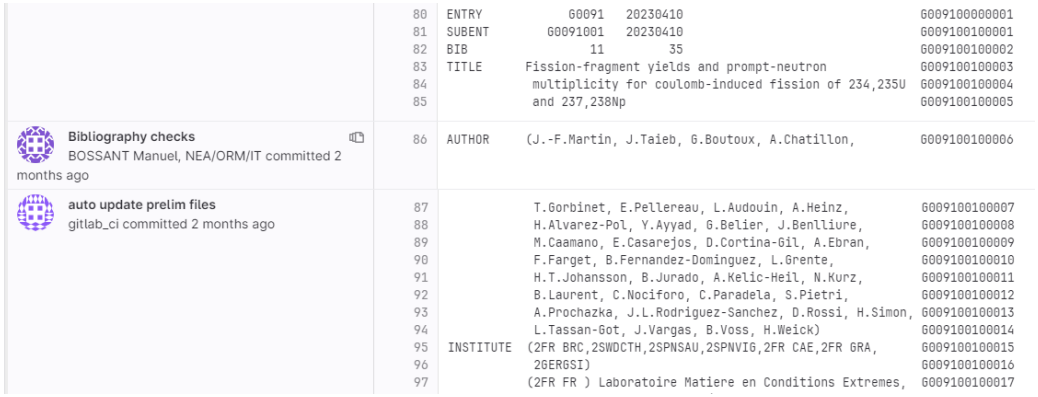
For a repository

For a user

Tasks



Line change history (git blame)



	80	ENTRY 60091 20230410 6009100000001
	81	SUBENT 60091001 20230410 6009100100001
	82	BIB 11 35 6009100100002
	83	TITLE Fission-fragment yields and prompt-neutron 6009100100003
	84	multiplicity for coulomb-induced fission of 234,235U 6009100100004
	85	and 237,238Np 6009100100005
Bibliography checks BOSSANT Manuel, NEA/ORM/IT committed 2 months ago	86	AUTHOR (J.-F.Martin, J.Taieb, G.Boutoux, A.Chatillon, 6009100100006
auto update prelim files gitlab_ci committed 2 months ago	87	T.Gorbinet, E.Pellereau, L.Audouin, A.Heinz, 6009100100007
	88	H.Alvarez-Pol, Y.Ayyad, G.Belier, J.Benlliure, 6009100100008
	89	M.Caamano, E.Casarejos, D.Cortina-Gil, A.Ebran, 6009100100009
	90	F.Farget, B.Fernandez-Dominguez, L.Grente, 6009100100010
	91	H.T.Johansson, B.Jurado, A.Kelic-Heil, N.Kurz, 6009100100011
	92	B.Laurent, C.Nociforo, C.Paradela, S.Pietri, 6009100100012
	93	A.Prochazka, J.L.Rodriguez-Sanchez, D.Rossi, H.Simon, 6009100100013
	94	L.Tassan-Got, J.Vargas, B.Voss, H.Weick) 6009100100014
	95	INSTITUTE (2FR BRC,2SWDCTH,2SPNSAU,2SPNVIG,2FR CAE,2FR GRA, 6009100100015
	96	2GERGSI) 6009100100016
	97	(2FR FR) Laboratoire Matiere en Conditions Extremes, 6009100100017

(MR) Change view

Overview of proposed changed lines

prelim_files/prelim.2319

1		-	TRANS	2319	20240226	20000000	0
	1	+	TRANS	2319	20240429	20000000	0
2	2		ENTRY C	22877	20231026	22877000	1
3	3		SUBENT C	22877001	20231026	22877001	1
4	4		BIB	14	57	22877001	2
			@@ -142691,9 +142691,9 @@ INC-SOURCE (PHOTO) 23658001 19				
142691	142691		- Two water filled Be-containers around U target			23658001	22
142692	142692		2x(14.6 cm x 21 cm x 3.9 cm) at the room temperature			23658001	23
142693	142693		as moderators placed above and below U target.			23658001	24
142694		-	- Density of moderator material = 1 g/cm3			23658001	25
142695		-	- Temperature of moderator : Room temperature			23658001	26
142696		-	- Moderator-room decoupler : None			23658001	27
	142694	+	- Density of moderator material = 1 g/cm3			23658001	25
	142695	+	- Temperature of moderator : Room temperatu			23658001	26
	142696	+	- Moderator-room decoupler : None			23658001	27
142697	142697		METHOD (TRN) Transmission in good transmission geometry			23658001	28
142698	142698		(TOF)			23658001	29
142699	142699		- Flight path (moderator centre-detector front face)			23658001	30
			@@ -142703,8 +142703,8 @@ METHOD (TRN) Transmission in good transmission geometry 2365800				
142703	142703		- Diameter of neutron beam at sample position = 45 mm			23658001	34
142704	142704		- Overlap suppression filter : 10B overlap filter			23658001	35
142705	142705		- Other neutron beam filters : Co, Pb			23658001	36
142706		-	DETECTOR (GLASD) Li-glass scintillator (152.4 mm diameter,			23658001	37
142707		-	6.35 mm thick) in the beam			23658001	38
	142706	+	DETECTOR (GLASD) Li-glass scintillator (152.4 mm diam, 6.35 mm			23658001	37
	142707	+	thick) in the beam			23658001	38
142708	142708		ANALYSIS Transmission Texp was derived by			23658001	39
142709	142709		Texp=N*([Cin-K*Bin]/[Cout-K*Bout]),			23658001	40
142710	142710		where			23658001	41

Version Controlled Exfor Files

- Automatically updated repositories via GitLab Actions
- [In development] use GitLab Actions for automatic review and quality control
- Can serve as versioned input for other tools and workflows
- Trigger pipeline to
  - Generate latest Exfor file once trans file is available
  - Registers a DOI for each Exfor file generated
    (currently on Sandbox Zenodo)

Repository references

NEA compilations

- https://git.oecd-nea.org/exfor/compilation/prelim

NRDC *Open Area* files

- https://git.oecd-nea.org/exfor/nrdc

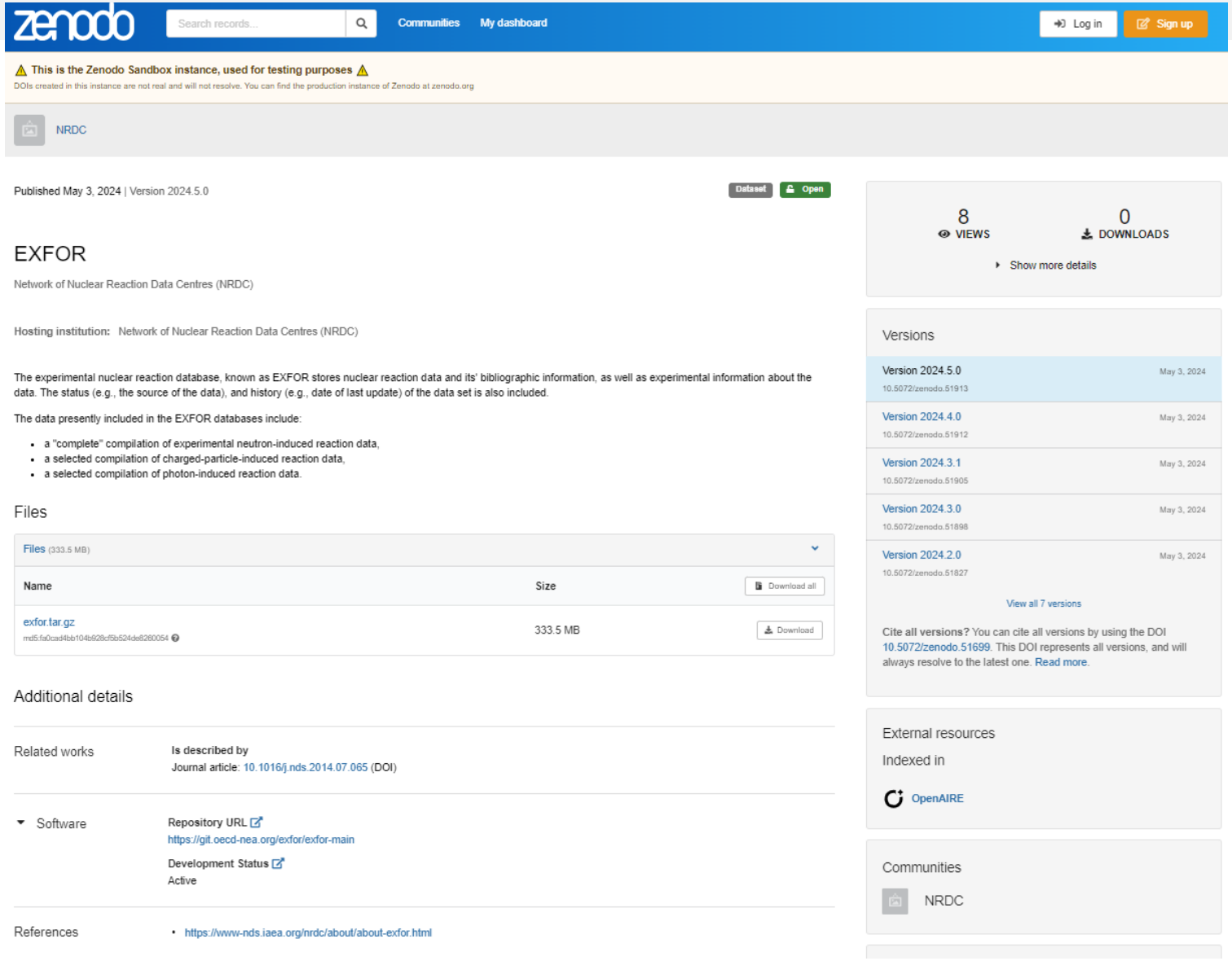
NEA reviews of prelim files

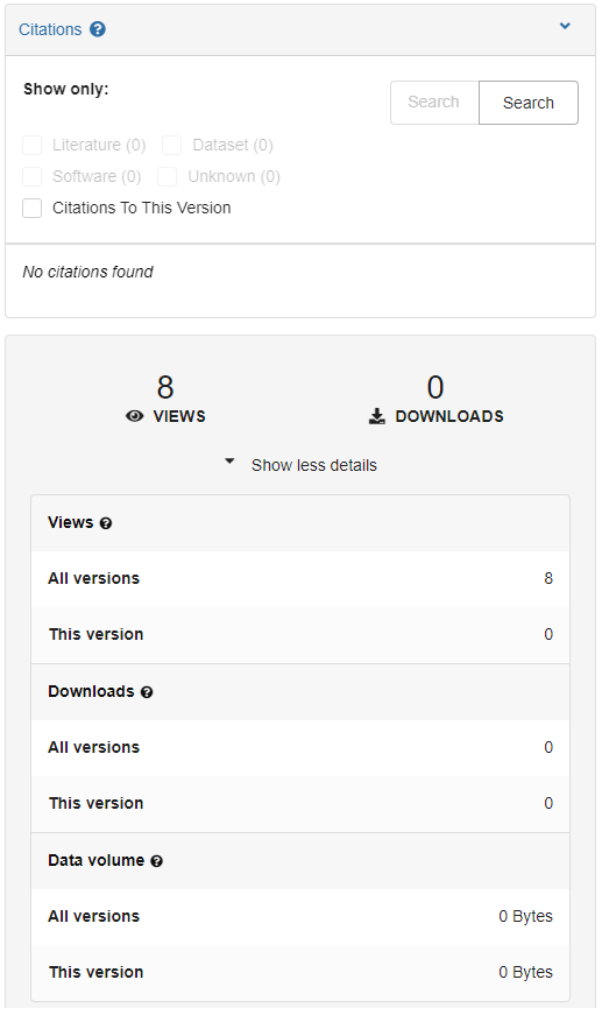
- https://git.oecd-nea.org/exfor/nrdc/review/prelim

Latest Exfor Files

- https://git.oecd-nea.org/exfor/exfor-main
  (exploded file version, compare also to EXFOR-Archive)

Digital Object Identifiers @ Zenodo

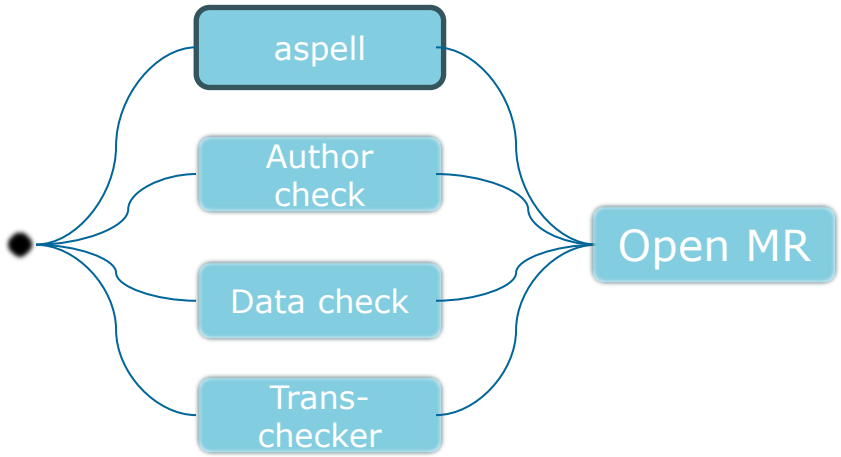




NEA reviews of prelim files

Review includes

- Spell checking
  - Automatic spell correction proposal coming soon!
- Author checking
  - In planning
- Data consistency check
  - In planning
- Run Janis trans-checker
  - Automatically include report [coming soon!]



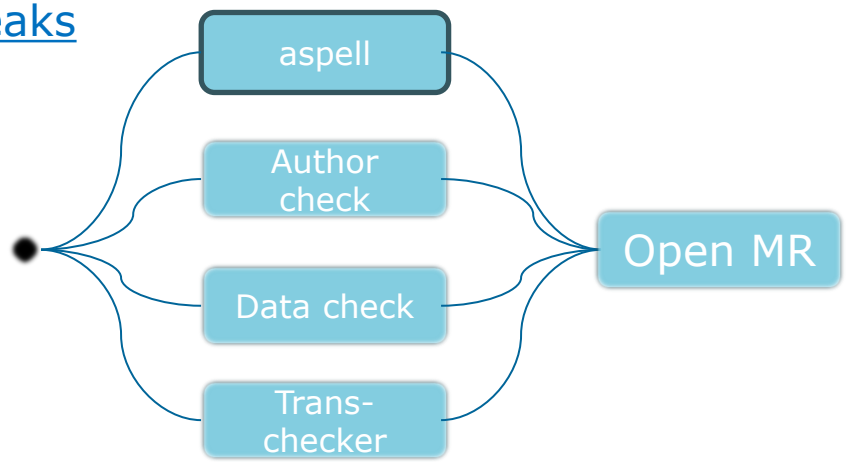
Exfor tools language extension for aspell

Aspell_speaks

- Exfor language extension for aspell
- Adds Exfor vocabulary as English language variant
- Additionally uses topic specific custom dictionaries
- Python utility function to correct spelling based on best-guess
- Development repo: https://git.oecd-nea.org/sprenger_j/aspell_speaks

X4_utils

- Version controlled code of x4_utils
- Wrapped as installable python package
- Repo: https://git.oecd-nea.org/exfor/tools/x4util



What's next?

For the NEA

- Continue automation of manual tasks
  - Finalize aspell workflow to automatically spell correct prelim files
  - Integrate with Janis trans-checker to automatically add feedback in GitLab MR
  - Proposal: Connect automatic email notification of new files in `Open Area` to Exfor technical mailing list
- Collect requirements for a non-expert entry editor

General Ideas

- Use journal feeds / notifications to populate list of articles to compile
- Register *our* NRDC as research organization at ror.org to get a unique identifier
- Create a NRDC community on Zenodo.org

For everyone

- Check out our repos
  - Create your own fork / clone
  - Set up your own workflows
  - Feel free to open issues / merge requests
    - Let us know what doesn't work
    - Let us know what would be useful for you!
  - Let me know if your git.oecd-nea.org account is not activated yet
- Join the NRDC Zenodo Community at zenodo.org
- Help uniquely identifying you by getting an ORCID (orcid.org)

ORCID stands for Open Researcher and Contributor ID

Thank you for your attention