

Software-Publikationen: Überblick, Herausforderungen, Bedarfe

Fachgespräch “Neue digitale Formate für die Forschung”,
Allianz-AG 8 “Wissenschaftspraxis”,
Bonn, Hochschulrektorenkonferenz, 2019-09-30

Stephan Druskat (Deutsches Zentrum für Luft- und Raumfahrt (DLR),
Friedrich-Schiller-Universität Jena, Humboldt-Universität zu Berlin)



seit 1558



Fachspezifika

Linguistik - Formate

- Papers, Monographien (inkl. z.B. Grammatiken & Wörterbücher), Konferenzbeiträge

- Papers, Monographien (inkl. z.B. Grammatiken & Wörterbücher), Konferenzbeiträge
- Modelle [*d*]
- Datensätze (v.a. Korpora) aus analogen oder digitalen Quellen [*a/d*]
- Datenbanken (z.B. Treebanks, typologische Datenbanken) [*d*]
- Standards (z.B. Tagsets) [*a/d*]
- Software [*d*]
- technische Berichte (Annotationshandbücher, etc.) [*a/d*]
- Digitalisate (z.B. von Handschriften) [*d*]

allgemein nativ [*a*]nalog / nativ [*d*]igital

Linguistik - Wahrnehmung und Nutzung digitaler Formate

- datengetriebene Teilbereiche (z.B. Korpuslinguistik, Computerlinguistik, Phonetik, moderne hist. Linguistik)
- Forschungsdaten sind zentral und werden auch so wahrgenommen
- Forschungssoftware ist zentral, wird aber nur teilweise so wahrgenommen (Computerlinguistik)

Forschungssoftware - Qualitätssicherung

- analog zu Lebenswissenschaften
- softwaretechnische Qualitätssicherung: vs. Praxis (Fähigkeiten, Ressourcen, Anreize)
- wissenschaftliche Qualitätssicherung: nur über Metaformate (Papers)
- Validität sekundärer Qualitätssicherung?

Nein

Nein

(bzw. kaum, v.a. sekundär via Paper)

Forschungssoftware - infrastrukturelle Unterstützung - Infrastrukturen

z.B. CLARIN (ERIC)

- webbasierte Analysewerkzeuge: Nutzung?
- Virtual Language Observatory (Suchmaschine)

Forschungssoftware - infrastrukturelle Unterstützung - Infrastrukturen

z.B. CLARIN (ERIC)

- webbasierte Analysewerkzeuge: Nutzung?
- Virtual Language Observatory (Suchmaschine)

Distribution (Publikation)? Vernetzung? Auswertung?



Record 4 of 14

< previous

next >



ANNIS Search and Visualization System for Corpora

Record details

Links (1)



Availability

All metadata

Technical details

Name ANNIS Search and Visualization System for Corpora

Description ANNIS is an open source, cross platform (Linux, Mac, Windows), web browser-based search and visualization architecture for complex multilayer linguistic corpora with diverse types of annotation. ANNIS, which stands for ANnotation of Information Structure, was originally designed to provide access to the data of the SFB 632 - "Information Structure: The Linguistic Means for Structuring Utterances, Sentences and Texts". It has since then been extended to a large number of projects annotating a variety of phenomena. Since complex linguistic phenomena such as information structure interact on many levels, ANNIS addresses the need to concurrently annotate, query and visualize data from such varied areas as syntax, semantics, morphology, prosody, referentiality, lexis and more. For projects working with spoken language, support for audio / video annotations is also required.

Collection IMS, CLARIN-D Centre, University of Stuttgart **Modality** multimodal 

Landing page

HDL 11022/1007-0000-0000-8...

Forschungssoftware - infrastrukturelle Unterstützung - Infrastruktur

- Ausbildung?
 - Karrierepfade?
 - grundständige Finanzierung?
 - softwarespezifische Infrastruktur (VCS, CI, etc.)?
 - dedizierte Softwarerepositorien/-archive?
 - (disziplinspezifische) Policies / Guidelines?
 - zentrale Anlaufstellen?
 - nicht-projektgebundene Softwaregruppen?
- informell
 - Nein
 - kaum
 - teilweise
 - Nein/Zenodo
 - Nein
 - Nein
 - Nein

Entwicklungs-, Steuerungs-, Reformbedarfe bzgl. Software- Publikationen

State of the art



Andreas Zeller

@AndreasZeller



You have built a tool X that finds security bugs. You also submit a paper about X to double-blind review. You want to release X as open source ASAP, but reviewers would challenge the paper's novelty. Wait until paper acceptance: programs will stay insecure for months. Thoughts?

[Tweet übersetzen](#)

1:18 vorm. · 27. Sep. 2019 · [Twitter Web App](#)

7 Retweets **34** „Gefällt mir“-Angaben

“Will an accepted paper increase the odds that people will use it?”



Andreas Zeller

@AndreasZeller



Antwort an [@hillelogram](#)

The accepted paper would be there to (a) document the technical details, (b) ensure that your name is linked to the contribution, and (c) help you advance your scientific career, e.g. getting tenure.

[Tweet übersetzen](#)

1:40 vorm. · 27. Sep. 2019 · [Twitter Web App](#)

2 „Gefällt mir“-Angaben

Publikationen

- “Document the details”
- “Link name to contribution”
- “Help advance careers”

Was ist der Forschungsbeitrag?

- Paper?
- Service?
- Datenoutput?

Was ist der Forschungsbeitrag?

- Paper?
- Service?
- Datenoutput?
- **Software!**
- “Document the details”: Software ist “the details”
 - Code ist Dokumentation
 - (Softwaredokumentation ist Dokumentation)
- **Bedarf₁ : Kulturwandel**

Attribution

- “Link name to contribution”
- **Bedarf₂: Metadaten**
 - **Citation File Format (CFF)** (Druskat et al. 2018) – Autorenformat
 - **CodeMeta** (Jones et al. 2017) – Austauschformat
 - **Steuerungsbedarf:** Metadaten fordern
 - **Entwicklungsbedarf:** Metadaten in Publikationsplattformen integrieren

- “Link name to contribution”
- **Bedarf₃: Zitierpraxis**
 - neben allgemeinen - z.B. DFG Kodex 2019 ((DFG) 2019) - spezifische Richtlinien und Policies für Herausgeber*innen, Gutachter*innen, Autor*innen (derzeit entwickelt von **FORCE11** **SCIWG** Guidance Task Force)
- **Steuerungsbedarf:** Evaluation inkl. Softwareveröffentlichungen
- Softwarezitierung hat zentrale Rolle für Reproduzierbarkeit
- **Entwicklungsbedarf:** Software muss eigene Referenzen zitieren (eigene Dissertation)

Publikationsplattformen

- Journal of Open Source Software (JOSS) / Software Journals?
- Zenodo
 - DOI für Versionen
 - DOI für “Konzept”
 - Metadaten?
 - Peer Review?



Catherine Jones

@CM_J0nes

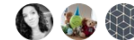
Follow



Panel [#ukrse19](#) now discussing the future of research software in five years.....
[@stdruskat](#) suggests that software publishing/journals should have gone away

9:56 AM - 17 Sep 2019

1 Retweet 3 Likes



1



1



3



Will Furnass @WillFurnass · Sep 17



Replying to [@CM_J0nes](#) [@stdruskat](#)

Yes, it's interesting to hear software journals being described as temporary things that will go away when recognition of software as a valid research output reaches a critical level. Makes sense! [#ukrse19](#)



1



1



3

Publikationsplattformen

- **Bedarf₄: geeignete Publikationsplattformen**
- **Entwicklungsbedarf:** wiss. Softwarearchiv mit PID, Versionen, (Zitations-)Metadaten; (Bibliotheken? NFDI?)
- **Entwicklungsbedarf:** Peer Review für Forschungssoftware; neue Rolle für Software Journals ohne Fließtextveröffentlichungen?

Karrierpfade

- “Help advance careers” – **Bedarf₅: Karrierepfade**
- Research Software Engineers analog zu wiss. Stellenbeschreibungen
- **Reformbedarf:** Grundlagen der Softwareentwicklung in Curricula
- **Reformbedarf:** Karrierepfade entwickeln und implementieren
- **Steuerungsbedarf:** (grundständige) Finanzierung softwarerelevanter Stellen, z.B. Grundsicherung durch Institutionen + Finanzierung durch Drittmittel
- **Steuerungsbedarf:** Einrichtung institutioneller RSE-Gruppen nach britischem Vorbild (?)

Danke!

Was sind Ihre Fragen?

stephan.druskat@dlr.de

Twitter: [@stdruskat](https://twitter.com/stdruskat)

ORCiD [0000-0003-4925-7248](https://orcid.org/0000-0003-4925-7248)

Appendix

Referenzen

(DFG), Deutsche Forschungsgemeinschaft. 2019. “Leitlinien zur Sicherung guter wissenschaftlicher Praxis (Kodex) [Guidelines for Safeguarding Good Scientific Practice (Code)],” August, 32 pages. <http://bit.ly/dfg-kodex-2019>.

Druskat, Stephan, Neil Chue Hong, Robert Haines, and James Baker. 2018. “Citation File Format (CFF) - Specifications,” August. <https://doi.org/10.5281/zenodo.1003149>.

Jones, Matthew B., Carl Boettiger, Abby Cabunoc Mayes, Arfon Smith, Peter Slaughter, Kyle Niemeyer, Yolanda Gil, et al. 2017. *CodeMeta: An Exchange Schema for Software Metadata. Version 2.0*. <https://doi.org/10.5063/schema/codemeta-2.0>.