JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Parallel Publishing from Converis to DSpace Using Custom Export and REST API

Miika Nurminen, JYU (University Services / IT Services)

Ari Häyrinen, JYU (Open Science Centre / Library)
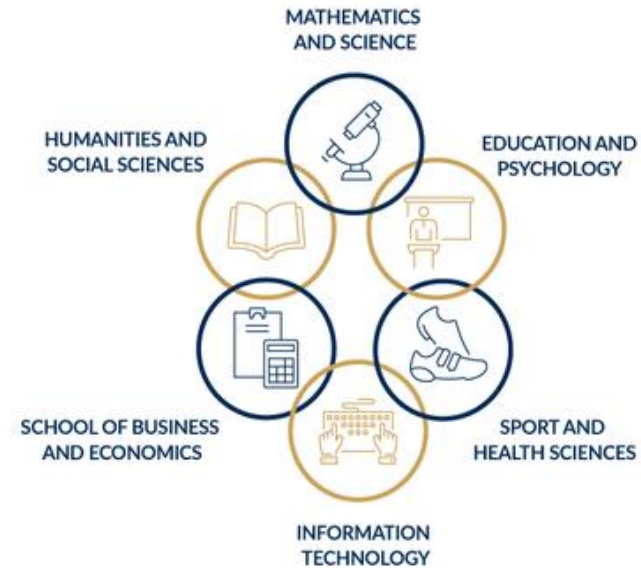
University of Jyväskylä

JYU – significant multidisciplinary research university and expert in the fields of learning and teaching. One of the most popular HEIs in Finland.

Background (CRIS)

- Legacy publication registry (+projects, travels, activities) TUTKA has been in use since 2003
  - / In-house development (started as student project, maintained by IT Services) based on Zope server
  - / Provides data for internal and external reports incl. performance assessments

- Reporting to the Ministry of Education and Culture shifted from statistics to metadata in 2011, publications became significant funding model component
  - / Recording publication data was originally left to researchers (department-level acceptance process) but this resulted to issues with data quality and motivation

- *The need for change* was identified – a working group (eventually *steering group for research information systems)* was established at the end of 2012
  - / An important process change with the legacy system was realized: in April 2014, publication recording was centralized to the Library (=now Open Science Centre)
    - → Implementing the process required management support and intense collaboration between the Division of Strategic Planning and Development (=Research management), University Library (setting up the team and publication alerts), and IT Services (adapting the systems)
  - / The group outlined a roadmap and evaluated implementation options, eventually recommending procurement of a new Current Research Information System

Background (IR)

- Institutional repository JYX launched in 2008

  / DSpace + local customizations, various types of content and interfaces (e.g. harvested to Finna search service and OpenAIRE portal)

  / One of the top-100 IRs (webometrics)

  / Open Science Centre manages the repository and provides support; development shared by OSC and IT Services

- Parallel publishing TUTKA → JYX introduced in 2011

  / Transfer based on XML-RPC calls, SWORD protocol and a custom file ingester to return handle to TUTKA on success

  / Centralized publication recording (2014) is a win-win situation to all stakeholders

    → With a small number of expert users, metadata quality of publications improved drastically; researchers are satisfied because the recording process has been simplified

    → **With the new process** backed by a university-level mandate*, **OA deposit activity has multiplied** (=c. 250 publications in 2013)

- Major DSpace update (1.7.3 → 6.2) completed in 05/2018

  / New theme and services implemented by OSC (GLAMpipe-based workflows utilizing REST interfaces, Omeka-based exhibitions)

Jyväskylä University Digital Repository

JYX contains digital materials of University of Jyväskylä from master's and licentiate theses to dissertations and from digitized journals to originally electronic materials. The majority of the collection is openly available. There may, however, be also some restricted items as well.

Search

Browse materials
> Articles
> Conferences and seminars
> Electronic books
> Historical maps
> Journals
> Tunes and musical notes
> Photographs
> Presentations and posters
> Publication series
> Research reports
> Research data
> Study materials
> Theses

Open Access Publishing in Jyväskylä University

Jyväskylä University promotes open publishing by offering an open publishing archive JYX for the use of the staff members. It is a place where you can self-archive your papers, share research data or store your digitised image or multimedia collections.

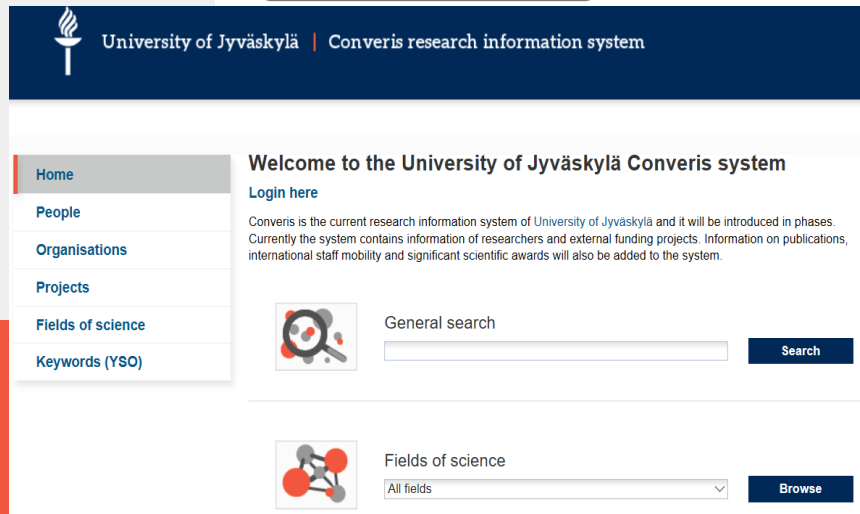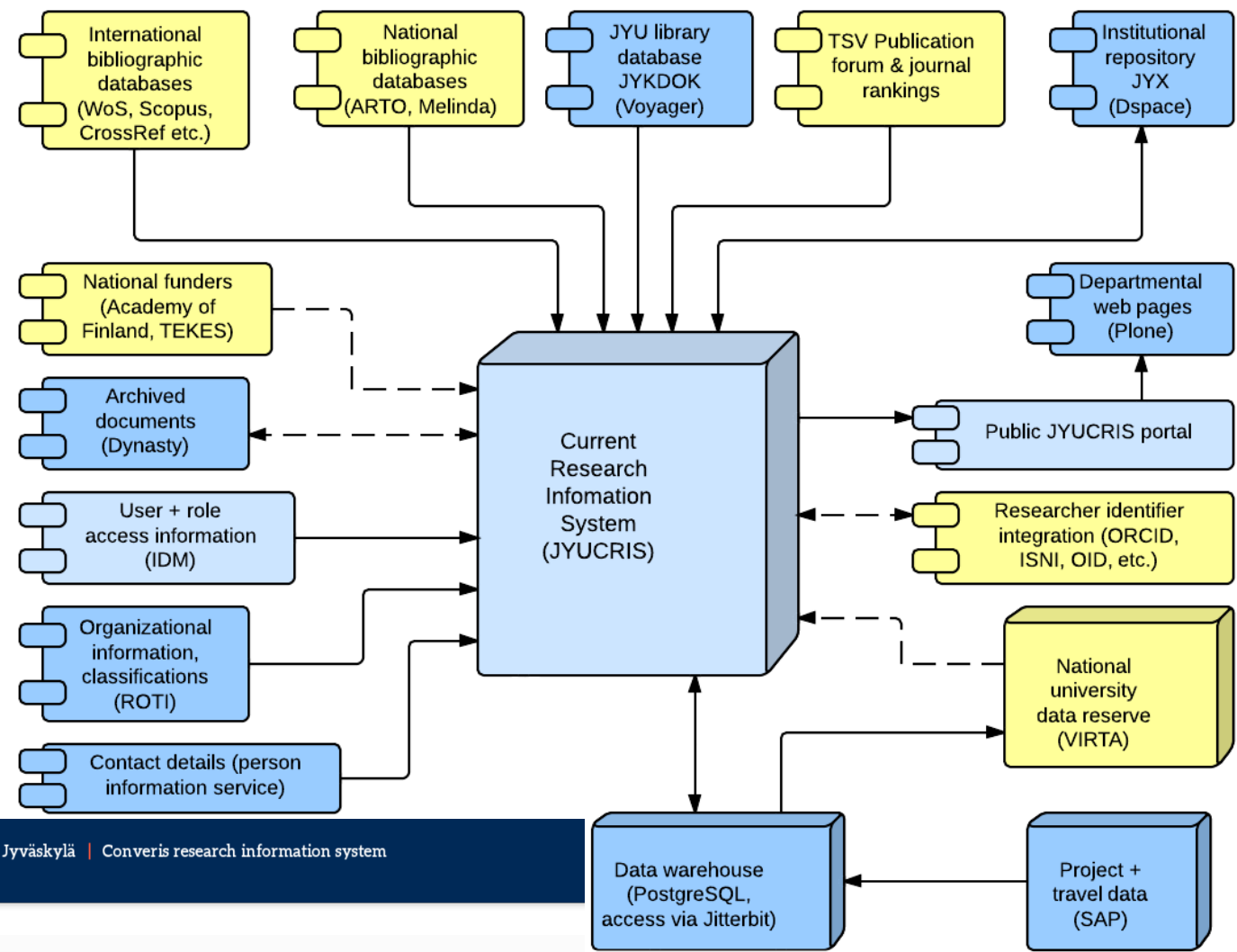| Date Issued |
| --- |
| 2017 (1542) |
| 2016 (1193) |
| 2015 (935) |
| 2014 (676) |

Implementation project started 12/2015

Launch (HR data, project applications) 09/2016

Public portal with projects opened 05/2017

Current focus: publications

- Data model configuration completed

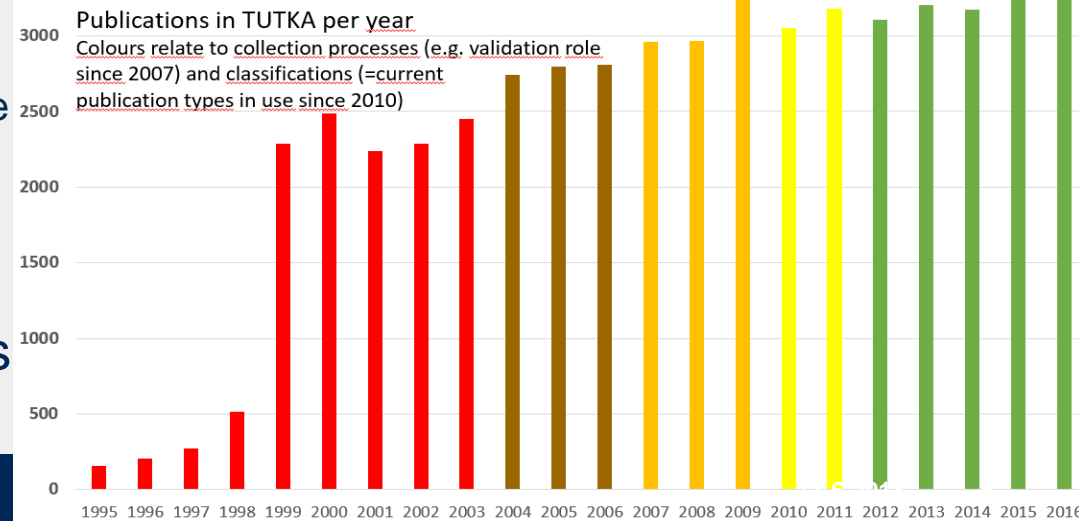- Interfaces to external systems in progress

Converis Transition

Implementation of the publication recording module in Converis

- Converis allows various means of customization within the system (e.g. data model, input forms, scripting) and integration options from external applications (REST, Pentaho Kettle, database connections)
  - / These have been used extensively at JYU
  - / Publication-related customizations include copying of metadata from the parent publication or linked publication channel, and generation of the citation field
- Interfaces to external systems require substantial amount of work
  - / WoS and Scopus integrations need configuration with field matching
  - / Additional integrations implemented in-house
    - → Publication channel + rankings import from the national service Publication Forum*
    - → Library system MARCXML import (=rework of equivalent functionality in the legacy system)
    - → Publications (c. 57000) from the legacy system will be imported with data cleaning (e.g. duplicate removal, mapping organization fields to entities in Converis)
- The original intention was that for parallel publishing, standard functionality offered by Converis would be used. However…

Publications in TUTKA per year
Colours relate to collection processes (e.g. validation role since 2007) and classifications (=current publication types in use since 2010)

Challenges with Converis/DSpace publication export

- When IR export functionality (based on SWORDv2) of Converis 5.10 was tested with DSpace 6, there were numerous issues
  - / Initially, the transfer did not work at all
  - / After additional configuration (*swordv2-server.keep-original-package=false*, *org.dspace.sword2.SimpleZipContentIngester = application/zip*) a publication was eventually transferred, but not visible in any collection
    - → Not even in the acceptance workflow of the target collection
    - → Without customization of *SimpleZipContentIngester* code (=e.g. adding *finishCreateItem*) the item stayed in submission workflow of the **SWORD** user
  - / It is unclear whether this is a bug in Converis (e.g. missing *In-Progress*-header) or a problem (=assumed client functionality) with SWORDv2 server
    - → The transfer should work with earlier versions of DSpace
- The root cause is likely the change from numeric to UUIDs in DSpace 6
  - / Indicated by line (*Error creating IR export item. The handle is malformatted*) in Converis logs when attempted to update an existing item
  - / As a result, changes to Converis code are needed
    - → The problem was immediately communicated to CA
    - → Converis 6 (published 04/2018) should be compatible with DSpace 6
  - / However, we decided to proceed with a custom solution for parallel publishing

Why a custom solution?

- Even if fixed, standard IR export functionality in Converis is limited
  - / Only link entities that have the publication as left side (e.g. PUBL_has_AREA) are supported; if multiple objects are related only the first one will be exported
  - / Embargo dates or publication visibility types (e.g. open access, embargo, request copy, IP limitation) are not supported in IR export – even within Converis
  - / Deleted publication in Converis will not get deleted in DSpace
- Custom application provides additional flexibility
  - / Loosely connected systems, resilience to data model or system changes
  - / Custom interface enables authorized access to restricted content – standard REST retrieval interface is limited to public metadata
    - → Applicability to other integrations (e.g. data for custom reports) as well
- Organizational benefits
  - / Utilize the experience with the interface used in the legacy system
  - / Utilize modern REST interface offered by DSpace 6
  - / Distributed development and knowledge sharing: IT Services provides the Converis-side interface, OSC develops the client application

The application*

- Converis side – HTTPS endpoints (with authentication):
  / *getIrSendables* – returns identifiers of new and updated publications/files
  / *getMetadataForIr* – returns metadata of a single publication (as JSON) with field names mapped to DSpace schema
  / *getFileForIr* – returns the publication file to be dispatched to DSpace
  / *updateIrSentFile* – updates a file-specific attribute in Converis side that the file has been submitted
  / *updateIrHandle* – updates handle to Converis after a new item has been accepted to the collection in DSpace
  / *updateIrMetadataFetchDate* – updates the latest metadata retrieval time
- DSpace side – JavaScript client application (jQuery)
  / List of new (to be accepted) or updated publications
  / Editing capability for selected fields with a lightweight acceptance process (e.g. setting embargo and resolving permits from publishers)
  / Dispatch to DSpace (bitstreams, metadata, policies) using REST API, no further action with DSpace workflow UI needed

Example publication page in Converis portal

- Home
- People
- Organisations
- Projects
- **Publications**
- Fields of science
- Keywords (YSO)

A4 Conference proceedings

A Simple Cluster Validation Index with Maximal Coverage (2017)

Jauhiainen, Susanne; Kärkkäinen, Tommi (2017) A Simple Cluster Validation Index with Maximal Coverage. In ESANN 2017 : Proceedings of the 25th European Symposium on Artificial Neural Networks. http://www.elen.ucl.ac.be /Proceedings/esann/esannpdf/es2017-24.pdf

JYU authors or editors

Jauhiainen, Susanne

Kärkkäinen, Tommi

Publication details
All authors or editors: Jauhiainen, Susanne; Kärkkäinen, Tommi
Parent publication: ESANN 2017 : Proceedings of the 25th European Symposium on Artificial Neural Networks
Conference:

European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning

Place and date of conference: Bruges, Belgium, 26.-28.4.2017
Publication year: 2017
Publication country: Belgium
Publication language: English
Website address: http://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-24.pdf
Open Access: Publication published in an open access channel

Abstract
Clustering is an unsupervised technique to detect general, distinct profiles from a given dataset. Similarly to the existence of various different clustering methods and algorithms, there exists many cluster validation methods and indices to suggest the number of clusters. The purpose of this paper is, firstly, to propose a new, simple internal cluster validation index. The index has a maximal coverage: also one cluster, i.e., lack of division of a dataset into disjoint subsets, can be detected. Secondly, the proposed index is compared to the available indices from five different packages implemented in R or Matlab to assess its utilizability. The comparison also suggests many interesting findings in the available implementations of the existing indices. The experiments and the comparison support the viability of the proposed cluster validation index.

Free keywords: cluster validation
Fields of science:

113 Computer and information sciences (Natural sciences)

Contributing organizations
JYU units:

Faculty of Information Technology

```
{
  'dc:rights': {'fi': [u'openAccess']},
  'dc:relation.ispartofseries': [],
  'dc:relation.ispartof': u'ESANN 2017 : Proceedings of the 25th European Symposium on Artificial
    Neural Networks, Computational Intelligence and Machine Learning',
  'dc:contributor.oppiaine': [u'tietotekniikka'],
  'dc:relation.uri': u'https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-24.pdf',
  'dc:identifier.other': 'TUTKAID_77636',
  'dc:relation.isbn': u'978-287587039-1',
  'dc:identifier.volume': u'',
  'dc:type.uri': u'http://purl.org/eprint/type/ConferencePaper',
  'dc:identifier.conference': u'European Symposium on Artificial Neural Networks, Computational
    Intelligence and Machine Learning',
  'dc:identifier.issn': u'',
  'dc:format.extent': u'',
  'dc:contributor.author': [u'Jauhiainen, Susanne', u'K\xe4rkk\xe4inen, Tommi'],
  'eprint:status': u'http://purl.org/eprint/status/PeerReviewed',
  'dc:format.content': u'fulltext',
  'dc:contributor.publisher': u'ESANN',
  'dc:contributor.editor': [],
  'dc:contributor.laitos': {'en': [u'Faculty of Information Technology'],
                            'fi': [u'Informaatioteknologian tiedekunta']},
  'dc:format.mimetype': u'application/pdf',
  'jyx:tutka.pagetopage': u'293-298',
  'dc:identifier.issue': u'',
  'dc:subject.other': [u'cluster validation'],
  'dc:relation.numberinseries': u'',
  'dc:title': u'A Simple Cluster Validation Index with Maximal Coverage',
  'dc:date.issued': 2017,
  'dc:description.version': u'Final Draft',
  'dc:type': u'conferenceObject',
  'dc:identifier.citation': u'Jauhiainen, S., & K\xe4rkk\xe4inen, T. (2017). A Simple Cluster
    Validation Index with Maximal Coverage. In ESANN 2017 : Proceedings of the 25th European Symposium on
    Artificial Neural Networks, Computational Intelligence and Machine Learning (pp. 293-298). ESANN.
    Retrieved from https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-24.pdf',
  'dc:language': [u'eng'],
  'dc:identifier.doi': u''
}
```

A Simple Cluster Validation Index with Maximal Coverage

Jauhiainen, Susanne
Kärkkäinen, Tommi

JavaScript client application (prototype with partial functionality, retrieves metadata from Converis side)

metadata

nimi	arvo	ohje	tiedosto
dc.type.version	acceptedVersion ▾	Tämä on pakollinen. Lisätietoa: tästä	
dc.date.embargo		Embargon päättymispäivä muodossa dd.mm.yyyy	
dc.description.abstract	Clustering is an unsupervised technique to detect general, distinct profiles from a given dataset. Similarly to the existence of various different clustering methods and algorithms, there exists many cluster	Englanninkielinen abstrakti	

Tallettaminen

JYX-linkki	Ei ole viety JYXiin vielä!

Parallel published to JYX

A Simple Cluster Validation Index with Maximal Coverage

Publisher's PDF

View/Open

1.7Mb

Jauhiainen, S., & Kärkkäinen, T. (2017). A Simple Cluster Validation Index with Maximal Coverage. In *ESANN 2017 : Proceedings of the 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 293-298). ESANN. Retrieved from https://www.elen.ucl.ac.be /Proceedings/esann/esannpdf/es2017-24.pdf

Authors
Jauhiainen, Susanne | Kärkkäinen, Tommi

Downloads: 2

Show download details

Date
2017

Discipline
tietotekniikka

Copyright
© the Authors, 2017.

Clustering is an unsupervised technique to detect general, distinct profiles from a given dataset. Similarly to the existence of various different clustering methods and algorithms, there exists many cluster validation methods and indices to suggest the number of clusters. The purpose of this paper is, firstly, to propose a new, simple internal cluster validation index. The index has a maximal coverage: also one cluster, i.e., lack of division of a dataset into disjoint subsets, can be detected. Secondly, the proposed index is compared to the available indices from five different packages implemented in R or Matlab to assess its utilizability. The comparison also suggests many interesting findings in the available implementations of the existing indices. The experiments and the comparison support the viability of the proposed cluster validation index.

Publisher
ESANN

Parent publication ISBN
978-287587039-1

Conference
European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning

OPEN ACCESS

Implementation status

- Work in progress – parallel development related to DSpace 6 update and Converis-related interfaces
- Metadata schemas and mappings have been specified and there is a working prototype for serving files and reading publication-specific metadata
  - As a related service, REST-based transmission of metadata and bitstreams is already in used with thesis submission forms
- Incremental file/metadata updates have not yet been implemented
  - It is still somewhat an open question what is considered as an significant update to metadata
  - All-encompassing (but only within the entity, not relations) *updated on* attribute in Converis is not sufficient, a hash based on transfer-related attributes+relations would be better
- Planned to go live in August with the launch of the publication module

Outlook

- New solution will streamline the parallel publishing process
  - / The client application simplifies OA status monitoring (i.e. pre-print, final draft or published version) and embargo management of individual files
  - / Incremental metadata update functionality reduces duplication of work and improves data quality
- Next steps in Converis implementation project
  - / Utilizing Converis to national publication reporting via data warehouse
  - / Update to Converis 6
- Collaboration with other Converis/DSpace users is essential
  - / Activity in Converis user groups and CA's Enhancement Request platform is a necessity to prioritize urgent development needs
  - / Running a DSpace 6 installation in production benefits from participation with the development community
- Publishing the custom solution -related code is still an open issue, university-level principles on open source are under construction
  - / Advancing Open Science and Research is listed as one of the strategic goals at the JYU

References

- *Converis 5.10 – Configuration Manual Supplement*. Clarivate Analytics 2017.

- *DSpace 6.x Documentation – REST API*. Wiki page, https://wiki.duraspace.org/display/DSDOC6x/REST+API, DuraSpace 2017.

- R. Jones, S. Lewis *et al*. *SWORD 2.0 Profile*. Technical report, http://swordapp.org/wp-content/uploads/2011/04/SWORDProfile.html, 2011.

- J. Kokko. *Implementation Project of the Research Information System*. Web site, https://www.jyu.fi/yliopistopalvelut/str/erityistoiminnot/tietohallinto/cris-kayttoonottoprojekti, University of Jyväskylä, 2015.

- M.-L. Harjuniemi. *Kirjasto yliopiston julkaisurekisterin ylläpitäjänä: kokemuksia keskitetystä kirjaamisesta* [Library as the Maintainer of University's Publication Registry: Experiences From Centralized Recording]. *Signum* 2/2015.

- A. Häyrinen. *Editing DSpace Metadata with GLAMpipe*. Screencast, https://www.youtube.com/watch?v=Ll_nKfB2otM, 2016.

- A. Häyrinen, M. Lassila, V.-M. Häkkinen & M. Nurminen. *Julkaisuarkisto JYX uudistuu* [Institutional repository JYX is being renewed]. Blog post, https://osc.jyu.fi/fi/blogi/julkaisuarkisto-jyx-uudistuu, University of Jyväskylä, 2018.

- K. Miettinen, A. Auer, J. Kokko, S.-L. Korppi-Tommola, S. Liimatainen, M. Nurminen, J. Ojala & K. Tuominen. *TUTKA-ryhmän loppuraportti* [TUTKA working group final report]. University of Jyväskylä, 2013.

- M. Nurminen. *Preparing for CRIS: Challenges and Opportunities for Systems Integration at Finnish Universities*. Poster at *Open Repositories 2014*.

- P. Olsbo. *Boosting for Green Open Access – Why, Who and How?* Presentation at *Oulu OpenCon* seminar, 2016.

- I. Ylönen. *Rinnakkaistallentamisen kriittiset menestystekijät* [Critical Success Factors of Parallel Publishing]. Presentation at *Parallel Publishing in Finland* seminar, 2017.

Thank you

minurmin@jyu.fi

http://users.jyu.fi/~minurmin/

https://converis.jyu.fi/

https://jyx.jyu.fi/