

ÉCOLE D'AUTOMNE 2016

DONNÉES DE LA RECHERCHE

MODULE 8: Outils de gestion des données de recherche dans les humanités numériques

Niklaus Stettler

Schweizerisches Institut für Informationswissenschaft (SII)

Zurich, 25.10.2016

Licence CC BY 2.5

All cartoons courtesy of Jørgen Stamp,
digitalbevaring.dk
CC BY 2.5



Sommaire

Retour d'expérience

Prestations possibles en bibliothèque

Infrastructure pour la documentation de la recherche

Infrastructure pour la publication de données





Retour d'expérience de TextGrid et DARIAH.de

- Les prestations de Dariah ont reçu un accueil mitigé.
- Aucun centre de services mutualisés n'a été mis en place.
- Les chercheurs en humanités numériques travaillent souvent seuls ou en petits groupes.
- Les chercheurs souhaitent du conseil/aide concernant l'utilisation de logiciels.
- Lorsqu'ils débutent avec de nouveaux outils, ils ne rencontrent pas de problèmes d'utilisabilité.
- Il n'est pas facile de surmonter les problèmes de communication avec les équipes informatiques.
- Les bibliothèques devraient assumer un rôle plus actif dans la GDR.



Retour d'expérience de la gestion des données de recherche

- Les chercheurs utilisent très peu les services de gestion des données
 - Ils n'expriment presque aucun besoin par rapport à une gestion des données structurée.
- Malgré cela, les chercheurs ne sont pas satisfaits de la situation actuelle.
 - Ils souhaiterait une gestion des données pour leur propre activité de recherche.
 - Il existe un besoin de connexion entre les données et les publications (throughput).
 - Les chercheurs des sciences humaines et sociales attachent une grande importance à l'archivage à long terme.

Ce qu'on peut en conclure

- La gestion de données de la recherche doit cibler les chercheurs.
 - Ceux-ci ont besoin d'assistance concernant les outils disponibles.
 - Les possibilité de gestion des données doivent être améliorées.
- => Les bibliothèques doivent garantir l'accès aux données de recherche.

Sommaire

Retour d'expérience

Prestations possibles en bibliothèque

Infrastructure pour la documentation de la recherche

Infrastructure pour la publication de données





Les chercheurs souhaitent être conseillés par rapport aux outils

- De très nombreux outils sont à disposition – mais il en manque une vue d’ensemble.
- Les bibliothèques devraient rendre les chercheurs attentifs à ces outils.
- Elles peuvent garantir un accès à certains logiciels.
- Elles devraient également mettre à disposition des formations (ex.: MOOCs, vidéos).
- Important: il ne suffit pas de créer une simple liste d’outils. Les chercheurs ont aussi besoin de conseils.

- Exemples:
 - Jan Krause: A Selection of Research data Management Tools Throughout the Data Lifecycle EPFL, 2015: <https://infoscience.epfl.ch/record/211157>
 - Explications d’outils de TextGrid: <https://wiki.de.dariah.eu/display/TextGrid/Subject-specific+Tools+and+Services>
 - <https://www.dataone.org/all-software-tools>



Outils génériques pour tâches récurrentes

- Diverses méthodes peuvent s'appliquer dans différents projets.
- Les bibliothèques peuvent mettre à disposition des outils génériques.
- Exemples: outils génériques pour les tâches suivantes:
 - Création d'un plan de gestion de données (<https://dmptool.org/plans/19207.pdf>)
 - Annotations
 - Visualisation
 - Géoréférencement
 - Prosopographie
 - Analyse de discours
 - Ecriture collaborative

Exemple pour le géoréférencement

- La représentation spatiale peut être intéressante dans différents projets.
- Si les chercheurs ont un moyen simple de géoréférencer leurs données, de nouvelles questions ou domaines de recherche apparaissent.
- DARIAH met à disposition un navigateur géographique (geobrowser) – une prestation possible pour les bibliothèques.
- <https://geobrowser.de.dariah.eu/edit/index.html#id=404651>



.oad Data

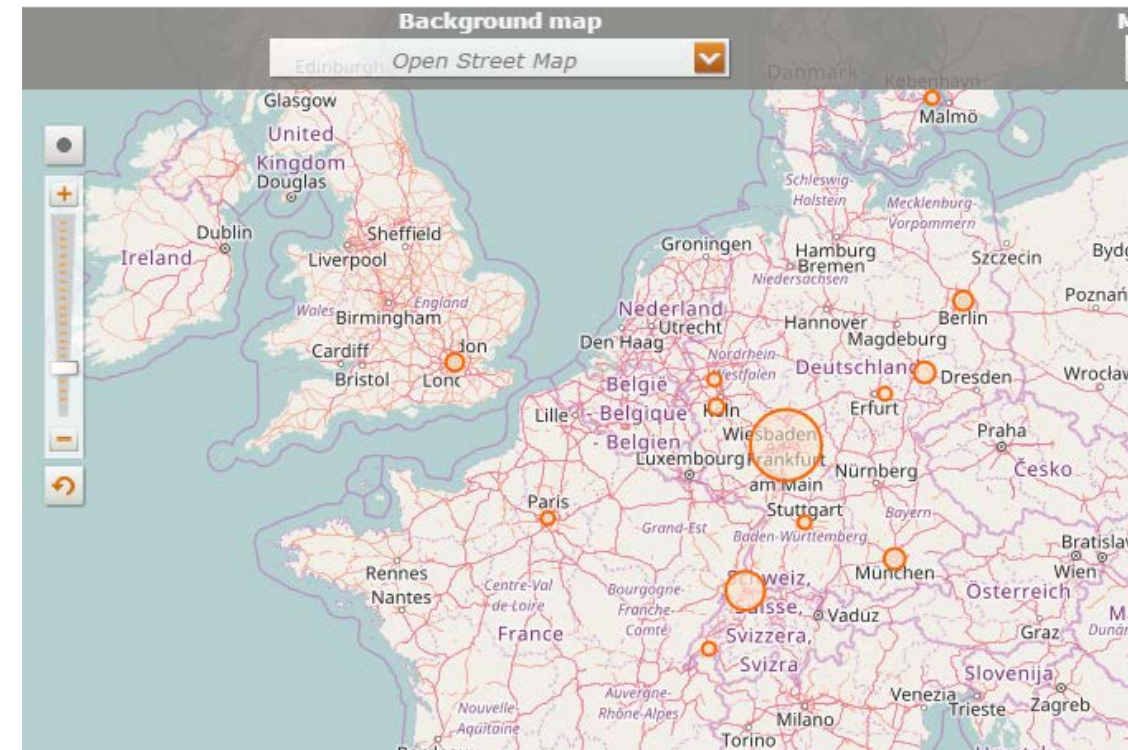
Static Data

Franz Kafka

load

Load Overlay

KML File URL



Exemple pour la prosopographie

- Les acteurs sont importants dans beaucoup de projets.
- C'est intéressant, lorsque l'on peut représenter des relations entre les personnes.
- DARIAH met à disposition une base de données de personnes permettant d'enregistrer divers aspects concernant les acteurs.
- Une base de données de personnes (Archiv-Editor) peut générer de nouvelles connaissances pour un institut.
- Les chercheurs peuvent réciproquement se mettre à disposition certaines connaissances.
- <https://de.dariah.eu/personendatenrepositorium>

Archiv-Editor 2.3.14 lite

Datei Person Aspekt Quelle Ansicht Datenbank Einstellungen Hilfe

Aktueller Benutzer
admin admin (admin)

Aspekte-Ansicht Hannover, Ernst August von Leibniz, Gottfr...

Aktuell ausgewählte Quelle: Leibniz Universität Hannover
[Note: 3.9.13]

Kriterien für Baumdarstellung
Suche nach: Personen
Kategorie/Quellentyp: Themen

Suche

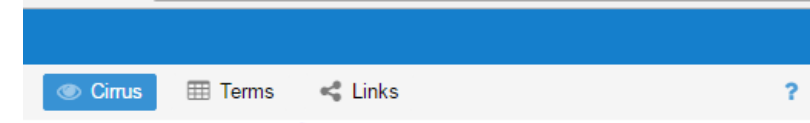
Alle Personen Alle Quellen Personen 3

- Altmitgliederverzeichnis der...
- Geburtsurkunde
- Holz, Hans Heinz: Leibniz: Das Leben eines Uni...
- Kuno, Fischer: Gottfried Wilhelm Leibniz: L...
- Leibniz Universität Hannover
- Mitgliederverzeichnis der Ro...
- Regionalgeschichte
- Sterbeurkunde
- Test, Test: dfgvfdg
- Titel
- Wikipedia

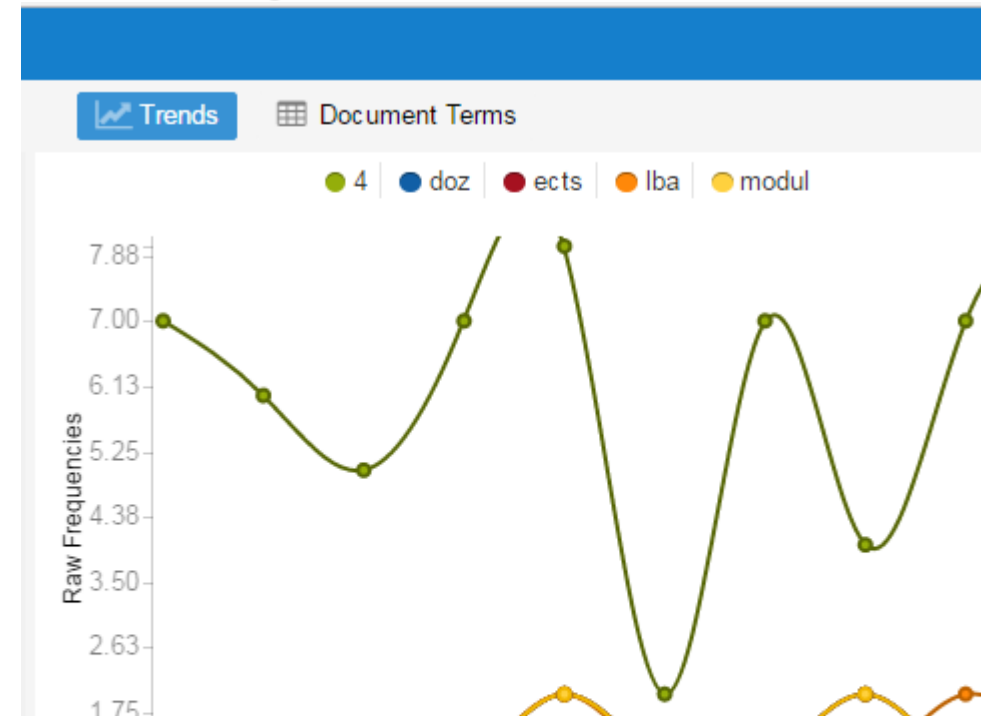
Subjekt \ Objekt	Objekt
Braunschweig-Lü...	Hannover, Ernst...
Hannover, Ernst...	
Hannover, Georg...	
Leibniz, Gottfr...	Angestellt von: 1676 Gefördert von: V...
Schönborn, Joha...	

Exemple pour l'analyse textuelle

- DARIAH met à disposition un outil simple d'analyse textuelle.
- <http://voyant-tools.org/>
- Les données de différents textes se laissent facilement traitées, par exemple grâce à Wordclouds, Bubblelines, etc.
- Ainsi des aspects très divers peuvent être abordés: utilisation de termes dans les textes littéraires, développement des discours au fil du temps, etc.



Terms:

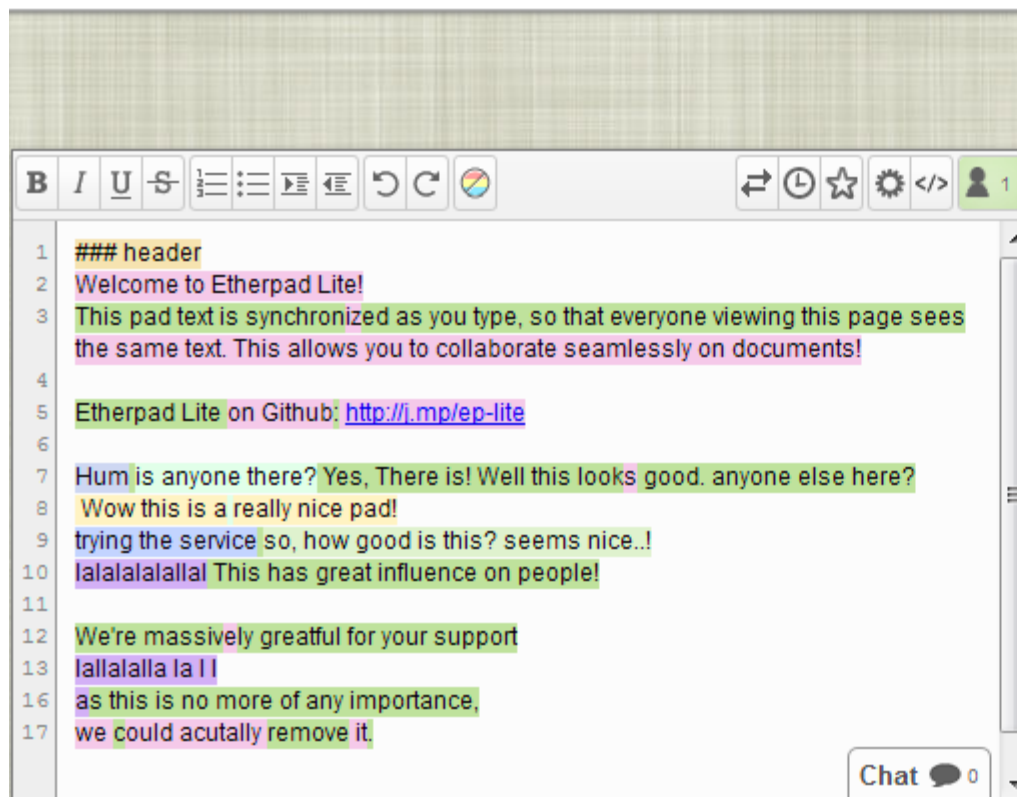


Exemple pour l'écriture collaborative

- DARIAH renvoie vers etherpad, un outil simple d'écriture collaborative.
- <http://etherpad.org/>



[About](#) [Download](#) [Contribute](#) [Links](#)



Etherpad is a highly customizable **Open Source** online editor providing collaborative editing in real **real-time.**

DOWNLOAD

Version 1.6.0



Un nouveau rôle pour les bibliothèques

- Assistance aux chercheurs dans la recherche d'outils
- Accès à des formations sur les logiciels
- Mise à disposition de certains outils génériques pour les activités de recherche

Ceci peut représenter un nouveau rôle pour les bibliothèques dans le domaine de la gestion des données de recherche

- Lié à la production des données de recherche
- Impliquant une simplification de la gestion de ces données
- Qui, sinon les bibliothèques, pourrait assumer ces services?

Connaissez-vous vos chercheurs?

- Si les bibliothèques veulent assumer un rôle actif dans la gestion des données de la recherche, elles doivent intensifier leurs relations avec les chercheurs.

Echangez:

- Connaissez-vous les chercheurs?
- Connaissez-vous leurs besoins en termes d'outils?
- Les bibliothèques doivent-elles ou peuvent-elles offrir des services pour les différents instituts et disciplines?
- A quoi cela pourrait-il ressembler?

Sommaire

Retour d'expérience

Prestations possibles en bibliothèque

Infrastructure pour la documentation de la recherche

Infrastructure pour la publication de données






Gestion des données de recherche signifie documentation de l'activité de recherche

- La gestion des données de la recherche contribue aussi à l'assurance qualité.
- Pour cela, outre les données elles-mêmes, les notes et les documents du processus de recherche doivent également être gérés.
- C'est à cela que sert normalement un cahier de laboratoire.
- Dans un environnement de recherche numérique, ce cahier doit être électronique (ELN, Electronic lab notebook)
- En sciences humaines et sociales, cet outil n'est pas encore établi, mais il est indispensable.

- Les ELN open source sont pour la plupart peu intuitifs.
- S'ils sont disponibles, les systèmes de gestion de documents peuvent alors représenter des outils de substitution. Exemples:
 - SharePoint
 - Confluence



L'implémentation de Confluence ou de SharePoint est une tâche pour les hautes écoles

- Introduire SharePoint ou Confluence est complexe.
- Ces outils ne facilitent le travail que s'ils sont bien implémentés.
- Sinon, les chercheurs se reportent à leurs propres outils.

Solutions moins complexes

- De nombreux chercheurs utilisent des outils de prise de notes déjà bien répandus, tels que:
 - Evernote
 - OneNote
 - etc.
- Fonctions: intégration de fichiers, calendrier, notes, synchronisation sur plusieurs terminaux, parties de fichier en groupes, export d'une page pour l'archivage
- Attention: ce sont des solutions cloud – les données doivent-elles restées chez Microsoft?

The screenshot displays the Microsoft OneNote application window. The title bar reads "Seminar 2016 - OneNote". The ribbon menu at the top includes "DATEI", "START", "EINFÜGEN", "ZEICHNEN", "VERLAUF", "ÜBERPRÜFEN", and "ANSICHT". On the left, the "Notizbücher" (Notebooks) pane is visible, showing a list of notebooks including "Eigenes Notizbuch", "dasErste", "Lesenotizen", "Familie", and "lod_led". The main content area shows a page titled "Seminar 2016" with a timestamp of "Dienstag, 2. August 2016 14:02". The page contains handwritten text in German:

Erfahrung: Bisher kritische Auseinandersetzung mit Daten
Besser: Wenn kleinere Einheiten
- Auseinandersetzung mit kleinen 'Häppchen',

On the right side, a search bar is labeled "Durchsuchen (Strg+E)". Below it, a "Seite hinzufügen" (Add page) panel is partially visible, listing "Die digitale seit", "Train2dacar 25", "Harvey dilza-Le", and "Seminar 2016". The user's email address "niklaus.stettler@outlook.de" is visible in the top right corner.

Solutions moins complexes

- Utiliser un logiciel de gestion de projet comme cahier de laboratoire électronique?
- Les tableaux de bord peuvent être exportés en CSV, JSON ou PDF.
- Exemple: Trello

The screenshot displays a Trello board interface for a project titled "2016_Projekt_BAR-Dateiformate". The board is organized into several columns:

- Korrespondenz:** A list of email correspondence, including messages from Niklaus, Krystyna, and wenkbruno.
- Aspekte:** A list of topics and questions related to digital preservation, such as "Container und Codecs", "Warum Audio und Video als Inhalte", and "JPEG2000: OCR Software?".
- Quellen:** A list of references and links, including "JPEG2000 and Digitisation: Expert round table" and "Using Lossy JPEG 2000 Compression For Archival Master Files".
- Arbeitspakete:** A list of tasks and work packages, including "Zuteilung der Arbeitspakete", "JPEG2000 bzw. JPEG", "OOXML bzw. Open Document", "XML / XSD", "Videoformate", "PDF/A-2", and "alle Überlegungen in Bezug zu Webseiten bringen".

The right sidebar contains a "Menü" (Menu) with options like "Hintergrund ändern", "Karten filtern", "Extras", "Sticker", and "Aktivität". The "Aktivität" section shows recent actions, such as "wenkbruno hat Mail vom 15.9. von Niklaus zu Korrespondenz hinzugefügt".

Sommaire

Retour d'expérience

Prestations possibles en bibliothèque

Infrastructure pour la documentation de la recherche

Infrastructure pour la publication de données



Où déposer les données?

- Une étude sur la gestion des données de recherche au sein d'universités du Bade-Wurtemberg a démontré que les dépôts disciplinaires sont clairement préférés dans les sciences humaines et sociales.
- Alors où déposer les données?

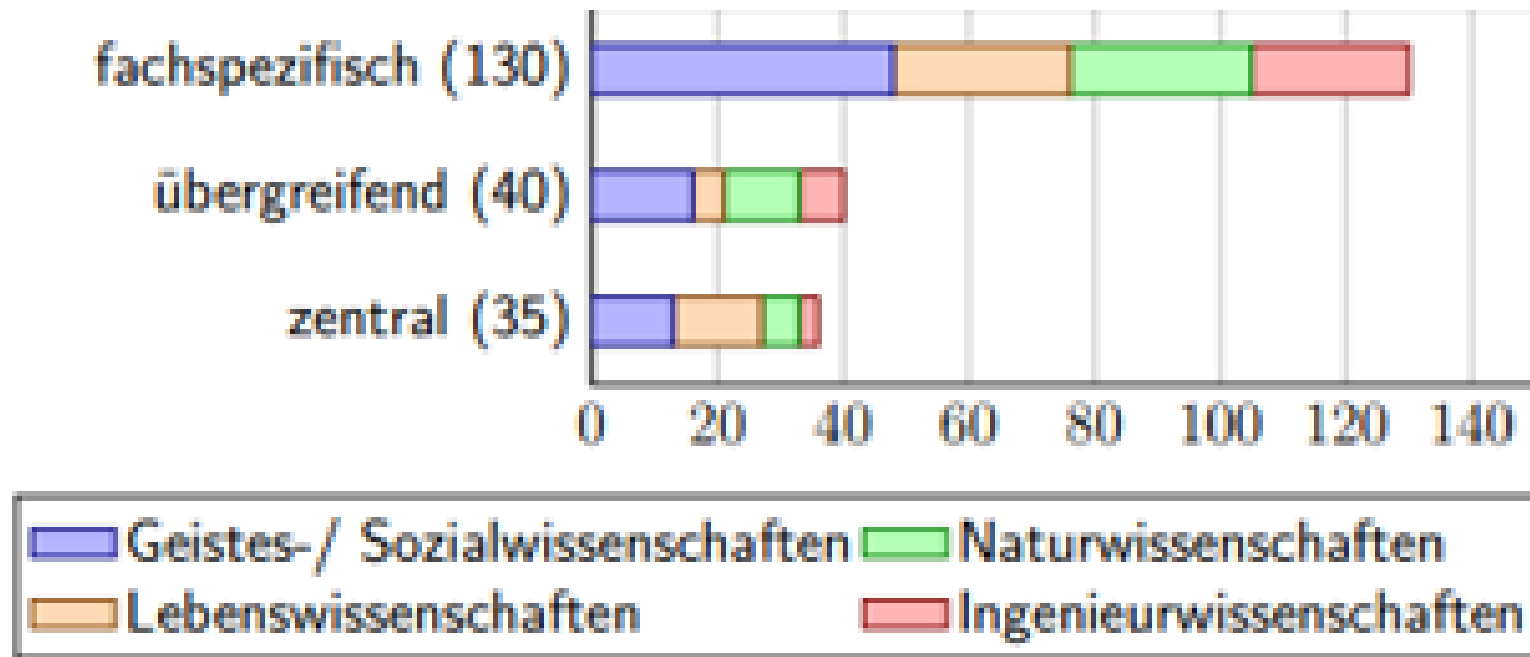


Abb. 34: Disjunkte Aufteilung nach dem Betriebskonzept des gewünschten Repositoriums in den Wissenschaftsbereichen.



Préférence pour des dépôts disciplinaires

- Pour augmenter la visibilité
- Pour augmenter la réutilisation

En Suisse, les dépôts disciplinaires pour les humanités numériques sont en cours de développement.

Dans le cadre du projet «Information scientifique» de swissuniversity, certains ont pu être créés. Mais l'horizon temporel reste relativement ouvert.

Une alternative de substitution jusqu'à l'émergence d'un service national?

- Mendeley Data – relie la gestion des références bibliographiques à la gestion des données
- Publication et données peuvent être liées
- Possibilité de créer des groupes
- Archivage à long terme des données
- Indexation des données par les moteurs de recherche

For datasets big and small
Store your research data online
Quickly and easily upload files of any type and we will host your research data for you. Your experimental research data will have a permanent home on the web that you can refer to.

Make your research data citable
DOIs and versioning following FORCE11 guidelines
Your published research data will include a FORCE11 compliant citation so that other researchers can effortlessly cite your research. We will also provide a unique DOI for each version of your dataset, so that your dataset's citation will always be valid.

Share your progress
with your colleagues and funding bodies
Privately share your unpublished data with the right people to move your research forward. You have full control over who can see and download your research data.

View and create
your dataset on any device
Mendeley Data has been created to work on a wide range of devices - from your laptop/desktop to your mobile phone. Browse datasets on the go, or take a picture in the lab and upload it instantly to your research data.

Long term preservation of your research data
Dark archive storage with DANS
We have teamed up with DANS (Data archiving and Networking Service) to ensure that your research data will

Une alternative de substitution?

Results: More than bin packing: Dynamic resource allocation strategies in cloud data centers

Published: 19 Sep 2015 | **Version 3** | DOI: 10.17632/xrpf54r78h.3

Viewed 524 Downloaded 175

















Contributor(s): [Andreas Wolke](#)

Description of this data

Data produced by simulations and experiments that were used in our paper "More than bin packing: Dynamic resource allocation strategies in cloud data centers" (Information Systems)

Experiment data files

[Download all files \(8\)](#)

 table_kmt_factors.csv	0.24 KB	
 sonar_cpu_mem_utilization.csv	0.15 KB	
 table_schedule_factors.csv	0.11 KB	
 table_schedule_factors_simulation.csv	0.07 KB	
 cbc_all.csv	108 KB	
 data---experiments.csv	26 KB	
 data---simulations-8.csv	108 KB	
 schedule_configurations4.csv	0.71 KB	

This data is associated with the following peer reviewed publication:

[More than bin packing: Dynamic resource allocation strategies in cloud data centers](#)

[Cite this article](#)



Published in:
Information Systems

Latest version

Version 3 2015-09-19

Published: 2015-09-19

DOI: 10.17632/xrpf54r78h.3

Merci beaucoup