

Pilotprojekt Open Research Data - Pilot ORD@CH

Schweizer Publikationsplattform für offene Forschungsdaten

Realisiert vom 1.7.2014 bis 31.12.2015 im Rahmen des Programmes SUK
2013-2016 P-2 "Wissenschaftliche Information: Zugang, Verarbeitung und
Speicherung"

Schlussbericht

Version 1.0

André Golliez, Golliez Open Data Consulting

Januar 2016

Inhaltsverzeichnis

1. Zusammenfassung	3
2. Auftrag, Ziele und Ablauf des Projektes	4
2.1 Auftrag	4
2.2 Projektziele	4
2.3 Projektverlauf	5
3. Projektergebnisse	7
3.1 Umfang Plattform (Gesamtarchitektur)	7
3.2 Die Web-Frontends und CKAN-Instanzen	8
3.3 Die Datenkataloge	15
3.4 CKAN Extensions und Harvester	17
3.5 Nutzerstatistiken im Pilotbetrieb	18
3.6 Aufbau und Betrieb der Plattform	20
3.7 Kommunikation und Interaktion mit der Open Data Community	20
4. Nutzerbefragung	22
5. Beurteilung der Projektergebnisse und Ausblick	27

1. Zusammenfassung

Das Schweizer Kompetenzzentrum Sozialwissenschaften FORS (antragstellende Institution) hat zusammen mit dem Digital Humanities Lab der Universität Basel und den ETH Scientific Services (Partner-Institutionen) im Rahmen des Programmes SUK P-2 „Wissenschaftliche Informationen: Zugang, Verarbeitung und Speicherung“ vom 1. Juli 2014 bis zum 31.12.2015 das Pilotprojekt Open Research Data (Pilot ORD@CH) durchgeführt.

Die Ergebnisse des Projektes können grundsätzlich positiv beurteilt werden. Das Pilotprojekt ORD@CH hat alle operativen Ziele innerhalb des vorgegebenen zeitlichen und finanziellen Rahmens erreicht:

- Alle CKAN-Instanzen sind live und über die jeweiligen Web-Frontends zugänglich.
- Sowohl das Harvesting der Metadaten aus den fachspezifischen Datenquellen auf die lokalen CKAN-Instanzen (NESTAR, SALSAH, OpenBIS) als auch der Transfer der Metadaten zur zentralen CKAN-Instanz konnten erfolgreich implementiert werden.
- Der Betrieb der gesamten Plattform ist stabil und kann ohne grösseren zusätzlichen Aufwand von ETH SIS weitergeführt werden.

Für die Überführung der Plattform in den Normalbetrieb und den schrittweisen Ausbau des Datenangebotes sind allerdings die folgenden Aspekte zu berücksichtigen:

- Ausbau Datenangebot:
Der Nutzen der Plattform ist direkt vom Umfang der publizierten Daten abhängig – je mehr Forschungsdaten zugänglich sind, desto grösser ist der Nutzen der Plattform, für Forschende, Studierende und weitere Interessierte.
- Weiterentwicklung User-Interfaces:
Die Web-Frontends wurden pilotmässig und mit minimalem Aufwand realisiert. Im Hinblick auf einen längerfristigen Betrieb der Plattform, sind diese auf die Bedürfnisse der Benutzer auszurichten und gründlich zu überarbeiten.
- Einbettung in einen Research Data Publishing Service:
Damit die Publikation der Forschungsdaten für die Forschenden mit minimalem Aufwand verbunden ist, sind die Funktionen „Archivierung“, „Identifikation (zwecks Zitierbarkeit)“ (z.B. mit DOI) und „Publikation“ (über eine oder mehrere Metadaten-Kataloge) als integrierter Service anzubieten.
- Kommunikation mit den Daten-Produzenten und –Nutzern:
Da die Publikation der Forschungsdaten noch lange keine Selbstverständlichkeit darstellt, braucht es eine intensive Kommunikation mit den Daten-Produzenten, um diese zur Veröffentlichung ihrer Daten zu motivieren, als auch mit (potenziellen) Daten-Nutzern, um ihnen das Potenzial der Wiederverwendung von Forschungsdaten zu demonstrieren.
- Positionierung im internationalen Kontext:
Die Plattform ist im internationalen Kontext zu positionieren, wo bereits vergleichbare Plattformen existieren oder aufgebaut werden (z.B. Zenodo.org, Datacite.org oder figshare.com).

2. Auftrag, Ziele und Ablauf des Projektes

2.1 Auftrag

Der Antrag für das Pilotprojekt Open Research Data (Pilot ORD@CH), den das Schweizer Kompetenzzentrum Sozialwissenschaften FORS (antragstellende Institution) zusammen mit dem Digital Humanities Lab der Universität Basel und den ETH Scientific Services (Partner-Institutionen) am 7. März 2014 im Rahmen des Programmes SUK P-2 „Wissenschaftliche Informationen: Zugang, Verarbeitung und Speicherung“ eingereicht hatten, wurde am 26. Juni 2014 vom Lenkungsausschuss SUK P-2 definitiv bewilligt. Der Auftrag des Projektes lautete gemäss bewilligtem Projektantrag wie folgt:

„Mit dem Pilotprojekt Open Research Data (Pilot ORD@CH) wird über einen Zeitraum von 18 Monaten eine Publikationsplattform für Forschungsdaten aus verschiedenen Wissenschaftsbereichen in der Schweiz aufgebaut und in Betrieb genommen. Die Plattform besteht im Kern aus einem Metadaten-Katalog der Datenbestände der partizipierenden Institutionen (FORS, Digital Humanities Lab der Universität Basel, ETH Scientific IT Services / SIB Swiss Institute of Bioinformatics), welcher für eine Volltextsuche via Web der Öffentlichkeit zur Verfügung gestellt wird. Der Metadaten-Katalog enthält Angaben über Inhalt, Autoren, Formate, zugehörige Publikationen sowie weitere Aspekte der einzelnen Datensätze und weist den Weg, um diese Daten zu nutzen (via direkten Download, über API's und Webservices oder über zuständige Kontaktstellen). Die Suchfunktion von ORD@CH ermöglicht es den Forschenden, Lehrenden und Studierenden aus den Sozial- und Geisteswissenschaften sowie aus den Life Sciences, Daten über die Grenzen der einzelnen Forschungsinstitutionen und –projekte hinweg zu finden und auf diese zuzugreifen. Als technische Plattform wird das Open Source-Frameworks CKAN eingesetzt, welches sich in zahlreichen Open Data-Projekten weltweit bewährt hat, u.a. auch als Basis des nationalen Open Government Data-Portales der Schweiz (opendata.admin.ch). Der föderale Aufbau der CKAN-Infrastruktur erlaubt es, die Datenbestände weiterer Institutionen, wie z.B. der ETH-Bibliothek (ETH Data Archive), schrittweise zu erschliessen und deren Metadaten in den zentralen Katalog zu integrieren. Die Plattform kann nach Abschluss des Pilotprojektes ohne grösseren Aufwand weiter betrieben und zu einem umfassenden Metadata Hub für alle Forschungsdaten der Schweiz ausgebaut werden.

Die Umsetzung des Projektes entspricht der „implementation action“ WE-2 im Bereich „publications“ des White Papers zum Programm SUK 2013-2016 P-2 (Seite 46): „Specifying and implementing a search solution for scientific publications and research data with a metadata hub and search engine, preferably as an extension to an existing solution (includes EP-11 and DM-6).“ (Projektantrag Pilotprojekt Open Research Data, 7.3.2014, Seite 3).

2.2 Projektziele

Die Ziele des Pilotprojektes ORD@CH wurden im Projektantrag wie folgt definiert:

- Eine metadatenbasierte Publikationsplattform für Forschungsdaten aus verschiedenen Wissenschaftsbereichen (Sozialwissenschaften, Geisteswissenschaften, Life Sciences) in der Schweiz wird schrittweise aufgebaut und für mindestens 9 Monate pilotmässig betrieben.

- Die Plattform besteht im Kern aus einem Metadaten-Katalog („metadata hub“) der Datenbestände der partizipierenden Institutionen (FORS, Digital Humanities Lab der Universität Basel, ETH Scientific IT Services / SIB Swiss Institute of Bioinformatics), welcher mit einer Volltextsuche („metadata search solution“) via Web von den Forschungsgemeinschaften sowie der breiteren Öffentlichkeit benutzt werden kann.
- Der Metadaten-Katalog wird aus den Datenarchiven und –verwaltungssystemen der Partnerinstitutionen sowie weiterer Forschungseinrichtungen (z.B. ETH Data Archive) schrittweise aufgebaut und ständig aktualisiert („metadata harvesting“). Wo möglich und sinnvoll, werden Metadaten mit anderen Datenkatalogen ausgetauscht (z.B. DataCite.org oder Zenodo).
- Die Primärdaten bleiben in den fachspezifischen Datenarchiven und –verwaltungssystemen und werden dort für die Nutzer der Plattform so einfach wie möglich zugänglich gemacht (via direkten Download, über API's und Webservices oder über zuständige Kontaktstellen).
- Die Plattform steht nach Abschluss des Pilotprojektes ohne grösseren zusätzlichen Aufwand zum Weiterbetrieb und Ausbau bereit.
- Die Erfahrungen der Datenproduzenten, der Datennutzer und der Plattformbetreiber sind am Ende des Pilotprojektes erhoben und dokumentiert. Der Schlussbericht zum Pilotprojekt ORD@CH bewertet diese Erfahrungen und umfasst Empfehlungen für den weiteren Betrieb und Ausbau der Plattform.

2.3 Projektverlauf

Das Pilotprojekt ORD@CH wurde vom 1.7.2014 bis 31.12.2015 gemäss Projektplan in folgenden Schritten durchgeführt:

Termin	Projektschritt	Ergebnis
1.7.2014	Start des Projektes	Projektorganisation operativ
bis 31.12.2014	Konzeptarbeiten	Konzept und Proof of Concept (PoC)
1.1. bis 31.5.2015	Realisierung	Zentrale und lokale CKAN-Instanzen und Frontends
Mai 2015	Go live	Alle CKAN-Instanzen live
5./6.6.2015	Open Research Data Hackdays	Über 100 Teilnehmende in Zürich (ETHZ) und Lausanne (FORS), über 10 Projekte realisiert
1.6. bis 31.12.2015	Pilotbetrieb	
1.6. bis 31.12.2015	Weiterentwicklung	Anpassungen Frontend, Ausbau Datenangebot
4.11. bis 31.12.2015	Nutzerbefragung	
bis 29.2.2016	Projektabschluss	Schlussbericht, Auswertung Nutzerbefragung, Reporting

Der Steuerungsausschuss Pilot ORD@CH hat sieben Mal getagt und das Projekt an seiner Sitzung vom 20. Januar 2016 formell abgeschlossen.

Alle operativen Ziele des Projektes ORD@CH wurden innerhalb der Zeitplanung und mit den vorhandenen finanziellen Mitteln erreicht.

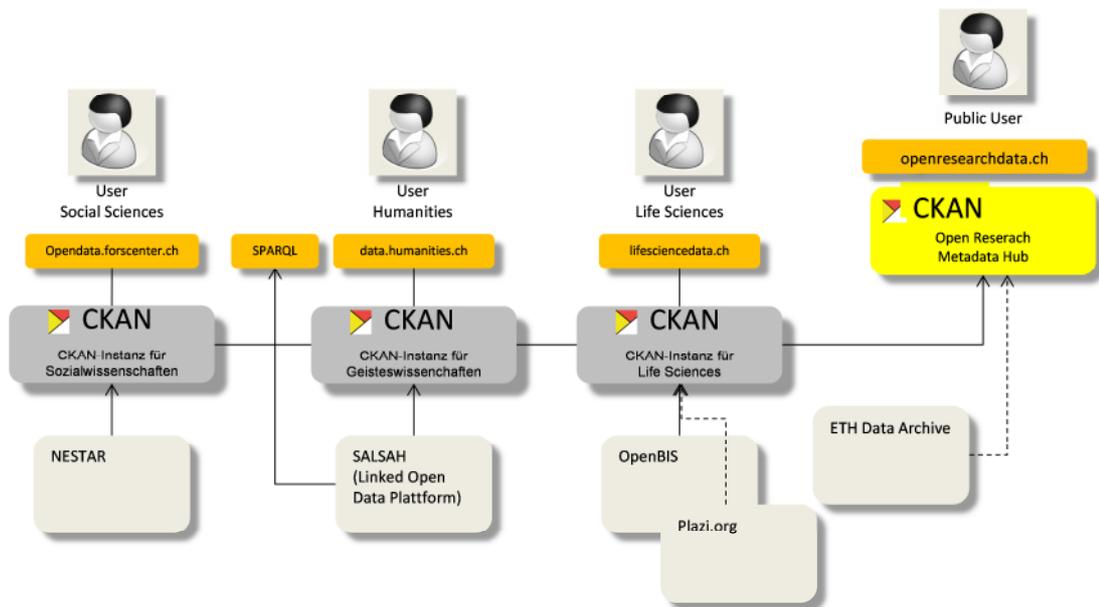
3. Projektergebnisse

In den folgenden Abschnitten werden die operativen Ergebnisse des Pilotprojektes ORD@CH dokumentiert

- Die Web-Frontends und CKAN-Instanzen (Screenshots, kurze Beschreibung der Funktionen)
- Die Datenkataloge (lokal und zentral)
- CKAN Extensions resp. Harvester
- Nutzerstatistiken (1.5.2015 – 16.1.2016)
- Aufbau und Betrieb der Plattform
- Kommunikation und Interaktion mit der Open Data Community (Open Research Data Hackdays)
- Nutzerbefragung (Kapitel 4)

3.1 Umfang Plattform (Gesamtarchitektur)

Per 31.12.2015 sieht die Plattform Open Research Data Schweiz wie folgt aus:



Die Metadaten aus dem ETH Data Archive sowie von Plazi.org werden bis Ende Februar auf die zentrale CKAN-Instanz (Web-Frontend www.openresearchdata.ch) resp. die lokale CKAN-Instanz für Life Sciences (Web-Frontend www.lifesciencedata.ch) geharvestet.

3.2 Die Web-Frontends und CKAN-Instanzen

Nutzerbedürfnisse und Funktionalitäten

In einer Konkurrenzanalyse (Q4-2014) wurden die Funktionalitäten von existierenden Angeboten zur Publikation von (Forschungs-)Daten erhoben. Folgende Dienste wurden analysiert:

Dienst	Slogan
B2FIND - b2find.eudat.eu	Find research data.
DataCite - datacite.org	Helping you to find, access, and reuse research data.
OpenAIRE - openaire.eu	Start sharing, reusing, and linking research to advance open science.
zenodo - zenodo.org	Research. shared.
GitHub - github.com	Build software better, together.
figshare - figshare.com	Credit for all your research.
Dataverse - dataverse.org	Share, Cite, Reuse, Archive Research Data – Scientific data for reproducible research.
DataHub - datahub.io	The easy way to get, share and use data.
opendata.cern.ch	–

Die erhobenen Funktionalitäten wurden nach Bereichen kategorisiert und nach Vorkommen aufgelistet:

Funktionalität	Bereich	Dienste
DOIs	Citation	DataCite, zenodo, opendata.cern.ch, figshare
News feed	News	DataCite, DataHub, figshare, Dataverse
Popular tags	Browsing	DataHub, opendata.cern.ch, figshare, Dataverse
Statistics on registered datasets	Statistics	DataCite, OpenAIRE, figshare
Upload publications as well as data	Publications	OpenAIRE, zenodo, Dataverse
Assign datasets to research "communities"	Browsing	B2FIND, zenodo
Slider for communities on frontpage	Browsing	B2FIND, Dataverse
OAI-PMH data provider	Export	DataCite, zenodo
Filter by type (dataset, publication, image...)	Searching	zenodo, opendata.cern.ch

Integrated with EC funding reports	Browsing	OpenAIRE, zenodo
GitHub integration	Programming	zenodo, opendata.cern.ch
REST API	Programming	zenodo, figshare
Blog	Community	Dataverse, figshare
Copy/paste citation	Citation	Dataverse, figshare
Forum	Community	DataCite
Test environment	Programming	DataCite
Content negotiation	Programming	DataCite
DOI citation formatter	Citation	DataCite
Glossary and tutorials on open data/publication	Publications	OpenAIRE
Choice of licenses	Browsing	zenodo
DropBox integration	Data	zenodo
Featured groups	Browsing	DataHub
Virtual Machines	Analysis	opendata.cern.ch
Visualizations of data	Visualization	opendata.cern.ch
Learning Ressources / Documentation	Analysis	opendata.cern.ch
Diffs to compare data revisions	Browsing	GitHub
Interaction with community	Community	GitHub (wiki + issue tracker)
Twitter integration	Community	Dataverse
Apps section	Community	Dataverse
Publications of studies	Publication	Dataverse
Export as MARXML	Export	opendata.cern.ch
Export as DDI	Export	Dataverse
Search for variables in tabular data	Searching	Dataverse
View & Share counter	Browsing	figshare
Category browser (inkl. description, image etc.)	Browsing	figshare
Preview of datasets (PDF,	Visualization	figshare

Images, Plaintext)

Comments	Community	figshare
Most shared / most views dataset section	Browsing	figshare
Store data not just metadata	Filesharing	zenodo

Auf dieser Grundlage wurden mit den am Projekt beteiligten Institutionen die Nutzerbedürfnisse erhoben und gewichtet. Daraus ergaben sich die folgenden Funktionalitäten für die vier Web-Frontends:

Funktionalität	Central	Humanities	Life Sciences	Social Sciences
Metadata search (full text)	X	X	X	X
Projects (to organise Datasets)	X	X	X	X
Mission statement	X	X	X	X
Datasets (most viewed / latest updated)	X	X	X	X
Categories (to organise Datasets)			(MeSH terms)	X
Tags (to organise Datasets)				X
Advanced metadata search interface			X	
Publications			X	
Multilanguage interface	X	X		X
About, Contact, How-to, FAQ	X	X	X	X
Logo, Title, Claim	X	X	X	X
Link to ,Central ,portal		X	X	X
Links to ,Research domain' portals	X			

Davon ausgehend wurden die Web-Frontends der vier CKAN-Instanzen konzipiert und entwickelt. Wo nötig wurden die Funktionalitäten dabei den während der Entwicklung zur Verfügung stehenden Daten angepasst.

Central - openresearchdata.ch

The screenshot shows the homepage of the Open Research Data Platform Switzerland. The page has a red header with a hexagonal pattern. At the top, there are navigation links for 'EN DE FR' and 'Mandated by swissuniversities'. Below this, there are links for 'Datasets', 'Projects', 'About', 'FAQ', and 'Contact'. The central logo features three stacked cubes with the word 'BETA' next to it. The main heading reads 'Open Research Data Platform Switzerland' with the tagline 'search. explore. reuse.' and a search input field. Below the search bar, there are links for 'Open Humanities Data', 'Open Life Sciences Data', and 'Open Social Sciences Data'. The main content area is divided into two columns. The left column lists 'Most used Datasets' and 'Last updated Datasets', with a red button for 'All Datasets'. The right column lists 'Projects with more Datasets' and 'Newest Projects', with a red button for 'All Projects'. The featured datasets include 'Reise ins Heilige Land' (art history), 'Eurobaromètre' (1 dataset), 'Selects' (3 datasets), and 'MOSAICH' (4 datasets). Two scientific datasets are also featured, both related to protein quantification assays. The footer contains links for 'About', 'Contact', 'Source code on Github', 'CKAN API', and 'Open Knowledge', along with the text 'Operated by ETH Zürich' and 'Mandated by swissuniversities'.

EN DE FR Mandated by swissuniversities

Datasets Projects About FAQ Contact

BETA

Open Research Data Platform Switzerland

search. explore. reuse.

Search

Open Humanities Data | Open Life Sciences Data | Open Social Sciences Data

Most used Datasets | Last updated Datasets **All Datasets** Projects with more Datasets | Newest Projects **All Projects**

Reise ins Heilige Land
kunstgeschichte
File formats: JPEG
Bilderfolgen Basler Frühdrucke

A repository of assays to quantify 10,000 human proteins by SWATH-MS - K_TRIP...
humans mass-spectrometry proteome proteomics
File formats: HTML

A repository of assays to quantify 10,000 human proteins by SWATH-MS - L_TRIP...
humans mass-spectrometry proteome proteomics
File formats: HTML

Eurobaromètre
Number of Datasets: 1

Selects
Number of Datasets: 3

MOSAICH
Number of Datasets: 4

About Contact Source code on Github Developed with CKAN CKAN API Open Knowledge Operated by **ETH Zürich** Mandated by swissuniversities

Humanities - data.humanities.ch

The screenshot shows the homepage of data.humanities.ch. At the top, there is a browser window with the URL 'data.humanities.ch'. Below the browser, a navigation bar includes language options (EN, DE, FR) and the text 'We are part of openresearchdata.ch'. The main header features a logo of three stacked cubes and the text 'Open Humanities Data' with the tagline 'search. explore. reuse.'. A search bar is prominently displayed. Below the search bar, there are several filter buttons: 'Most used Datasets', 'Last updated Datasets', 'All Datasets', 'Projects with more Datasets', 'Newest Projects', and 'All Projects'. The main content area displays three dataset cards on the left and two project cards on the right. Each card includes a title, a category tag (e.g., 'kunstgeschichte'), file formats, and a description. The footer contains links for 'About' and 'Contact', source code information (GitHub, CKAN API, Open Knowledge), and logos for 'DHLab', 'ETH SIS and FORS', 'ETH Zürich', and 'swissuniversities'.

data.humanities.ch

EN DE FR We are part of openresearchdata.ch

Datasets Projects About FAQ Contact

Open Humanities Data
search. explore. reuse.

Search

Most used Datasets Last updated Datasets **All Datasets** Projects with more Datasets Newest Projects **All Projects**

Itinerarius sive peregrinarius Beatissime Virginis Marie
kunstgeschichte
File formats: JPEG
Bilderfolgen Basler Frühdrucke

De generacione Christi
kunstgeschichte
File formats: JPEG
Bilderfolgen Basler Frühdrucke

Eyn blechlin sancti Methodij mar/trens. vnd bischoffs zuo partinentz. in kri...
kunstgeschichte
File formats: JPEG
Bilderfolgen Basler Frühdrucke

Bilderfolgen Basler Frühdrucke
Number of Datasets: 20

Bilddatenbank Bibliothek St. Moritz
Number of Datasets: 2001

About Contact Source code on GitHub Developed with CKAN CKAN API Open Knowledge Powered by **DHLab**, in collaboration with **ETH SIS** and **FORS** Operated by **ETH Zürich** Mandated by **swissuniversities**

The screenshot shows a web browser window with the URL `lifesciencedata.ch`. The page features a navigation bar with links for `Datasets`, `Projects`, `About`, `FAQ`, and `Contact`. A central logo consists of three stacked cubes. Below the logo, the text reads "Open Life Sciences Data" with the tagline "search. explore. reuse." and a search input field. A secondary navigation bar includes filters for `Most used Datasets`, `Last updated Datasets`, `All Datasets` (highlighted), `Projects with more Datasets`, `Newest Projects`, and `All Projects`. The main content area displays three dataset cards, each titled "A repository of assays to quantify 10,000 human proteins by SWATH-MS - L_TRIP..." and "SWATH-MS - J_TRIP...". Each card includes tags for `humans`, `mass-spectrometry`, `proteome`, and `proteomics`, and notes "File formats: HTML". The footer contains links for `About` and `Contact`, information about source code on GitHub and CKAN API, and mentions of funding and operation by ETH Zürich.

Social Sciences - opendata.forscenter.ch

The screenshot shows the homepage of the Open Social Sciences Data portal. At the top, there is a navigation bar with language options (EN, DE, FR) and the text "We are part of openresearchdata.ch". The main header features a logo of three stacked cubes and the title "Open Social Sciences Data" with the tagline "search. explore. reuse.". A search bar is prominently displayed. Below it, "Most used tags" are listed: "donnees-quantitatives", "valeurs", and "participation". The main content area is divided into sections for "Most used Datasets", "Last updated Datasets", "All Datasets", "Projects with more Datasets", "Newest Projects", and "All Projects". Two dataset entries are visible: "Voxit : Projet 973 du 30.11.2008" and "Sécurité 2005: Evolution des opinions sur la sécurité, la politique extérieur...". A sidebar on the right lists "Eurobaromètre", "Selects", and "Swiss Electoral Studies (Selects)" with their respective dataset counts. The footer contains contact information, source code links (GitHub, CKAN API, Open Knowledge), and acknowledgments to FORS, ETH SIS, and DHLab.

opendata.forscenter.ch

EN DE FR We are part of openresearchdata.ch

Datasets Projects About FAQ Contact

Open Social Sciences Data

search. explore. reuse.

Search

Most used tags:

donnees-quantitatives valeurs participation

Most used Datasets Last updated Datasets All Datasets Projects with more Datasets Newest Projects All Projects

Voxit : Projet 973 du 30.11.2008

comportement-politique connaissance-des-ob... decision-de-vote
donnees-quantitatives estimation-des-enjeux formation-de-l'opinion
interet-pour-la-pol... mode-de-vote moment-de-la-decision
mots-d'ordre opinion-politique opinion-sur-les-arg... participation
partis-politiques preference-pour-un-... utilisation-des-medias valeurs
votation vote

File formats: HTML

Voxit : enquêtes post-votation standardisées

Sécurité 2005: Evolution des opinions sur la sécurité, la politique extérieur...

alternatives-au-ser... attitude-envers-l'armee confiance-envers-le-...
donnees-quantitatives engagements-militai... evaluation-de-l'ave...
extremisme-politique neutralite opinion-sur-les-tac...
perception-de-la-me... securite-nationale sentiment-de-securite
terrorisme

Eurobaromètre Number of Datasets: 1

Selects Number of Datasets: 3

Swiss Electoral Studies (Selects) Number of Datasets: 4

About Contact Source code on GitHub Developed with CKAN CKAN API Open Knowledge Powered by FORS in collaboration with ETH SIS and DHLab Operated by ETH Zürich Mandated by swissuniversities

3.3 Die Datenkataloge

Datenkatalog	Projekte (Anzahl Datensätze)	Datensätze ohne Projekt
openresearchdata.ch	Bilderfolgen Basler Frühdrucke (20) Eurobaromètre (1) Eurobaromètre ISSP (4) European Social Survey (5) MOSAiCH (4) Sécurité (18) Selects (3) Swiss Electoral Studies (Selects) (4) Voxlt (23)	
data.humanities.ch	Bilddatenbank Bibliothek St. Moritz (2001) Bilderfolgen Basler Frühdrucke (20)	
lifesciencedata.ch	–	A repository of assays to quantify 10,000 human proteins by SWATH-MS - K_TRIPLE-TOF 2014-02 A repository of assays to quantify 10,000 human proteins by SWATH-MS - K_TRIPLE-TOF 2014-01 A repository of assays to quantify 10,000 human proteins by SWATH-MS - J_TRIPLE-TOF 2013-10 A repository of assays to quantify 10,000 human proteins by SWATH-MS - J_TRIPLE-TOF 2013-08 A repository of assays to quantify 10,000 human proteins by SWATH-MS - L_TRIPLE-TOF 2013-08 A repository of assays to quantify 10,000 human proteins by SWATH-MS - L_TRIPLE-TOF 2013-07 A repository of assays to quantify 10,000 human proteins by SWATH-MS - J_TRIPLE-TOF 2013-07 A repository of assays to quantify 10,000 human proteins by SWATH-MS - J_TRIPLE-TOF 2013-06 A repository of assays to quantify 10,000 human proteins by SWATH-MS - L_TRIPLE-TOF 2013-06 A repository of assays to quantify

10,000 human proteins by SWATH-
MS - L_TRIPLE-TOF 2013-03

A repository of assays to quantify
10,000 human proteins by SWATH-
MS - L_TRIPLE-TOF 2013-01

A repository of assays to quantify
10,000 human proteins by SWATH-
MS - L_TRIPLE-TOF 2012-11

A repository of assays to quantify
10,000 human proteins by SWATH-
MS - L_TRIPLE-TOF 2012-08

A repository of assays to quantify
10,000 human proteins by SWATH-
MS - L_TRIPLE-TOF 2012-07

A repository of assays to quantify
10,000 human proteins by SWATH-
MS - L_TRIPLE-TOF 2012-06

opendata.forscenter.ch Eurobaromètre (1)
Eurobaromètre ISSP (4)
European Social Survey (5)
MOSAiCH (4)
Sécurité (18)
Selects (3)
Swiss Electoral Studies (Selects)
(4)
Voxit : enquêtes post-votation
standardisées (330)

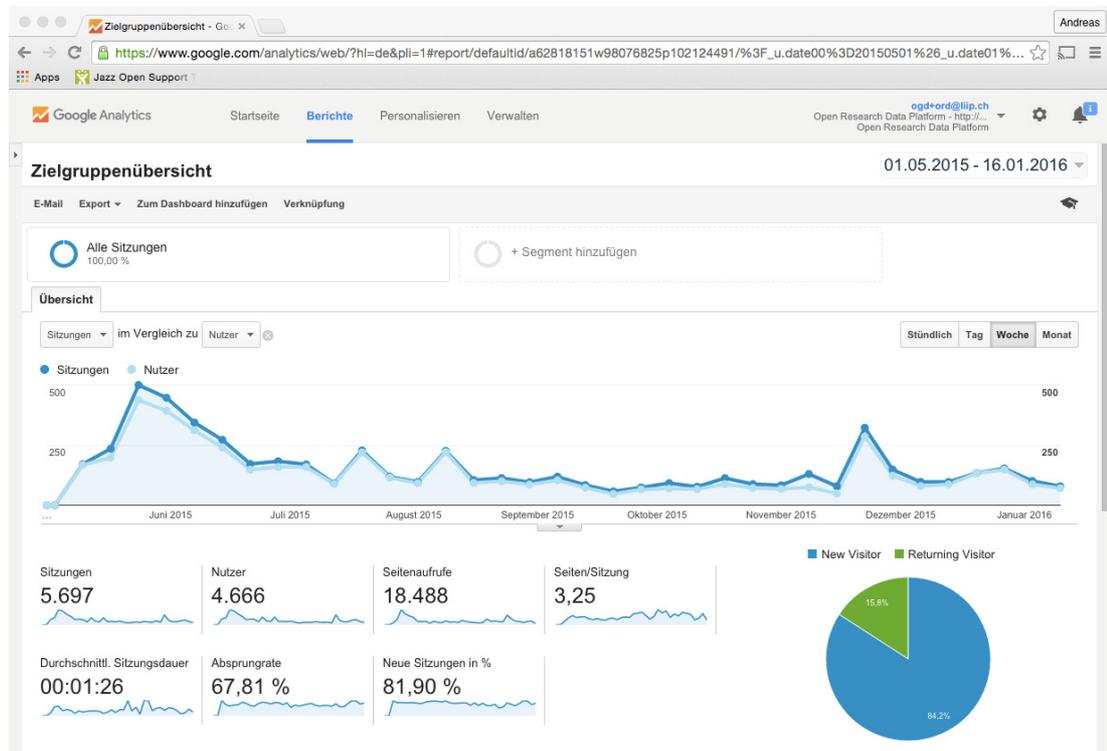
3.4 CKAN Extensions und Harvester

Datenkatalog	CKAN Extensions	Beschreibung	Letztes Update
openresearchdata.ch, data.humanities.ch, lifesciencedata.ch, opendata.forscenter.ch	ckanext-ordbase-theme	CKAN base theme for Open Research Data	08.01.2016
	ckanext-pages (forked from ckan/ckanext-pages)	simple cms	21.07.2015
	ckanext-ord-hierarchy	An extension to add a hierarchical structure to datasets	19.06.2015
	ckan (forked from ckan/ckan)	CKAN is an open-source DMS (data management system) for powering data hubs and data por- tals. CKAN makes it easy to publish, share and use data. It powers datahub.io, catalog.data.gov and data.gov.uk among many other sites.	26.06.2015
data.humanities.ch	ckanext-salsah	Harvester for salsah.org	26.06.2015
lifesciencedata.ch	ckanext-plazi	CKAN Harvester for Plazi	14.01.2016
	ckanext-oaipmh	OAI-PMH harvester for CKAN	14.01.2016
	ckanext-openbis	OpenBis harvester based on ckanext-oaipmh	03.04.2015
opendata.forscenter.ch	ckanext-nesstar	CKAN harvester for NESSTAR based on OAI-PMH	24.06.2015

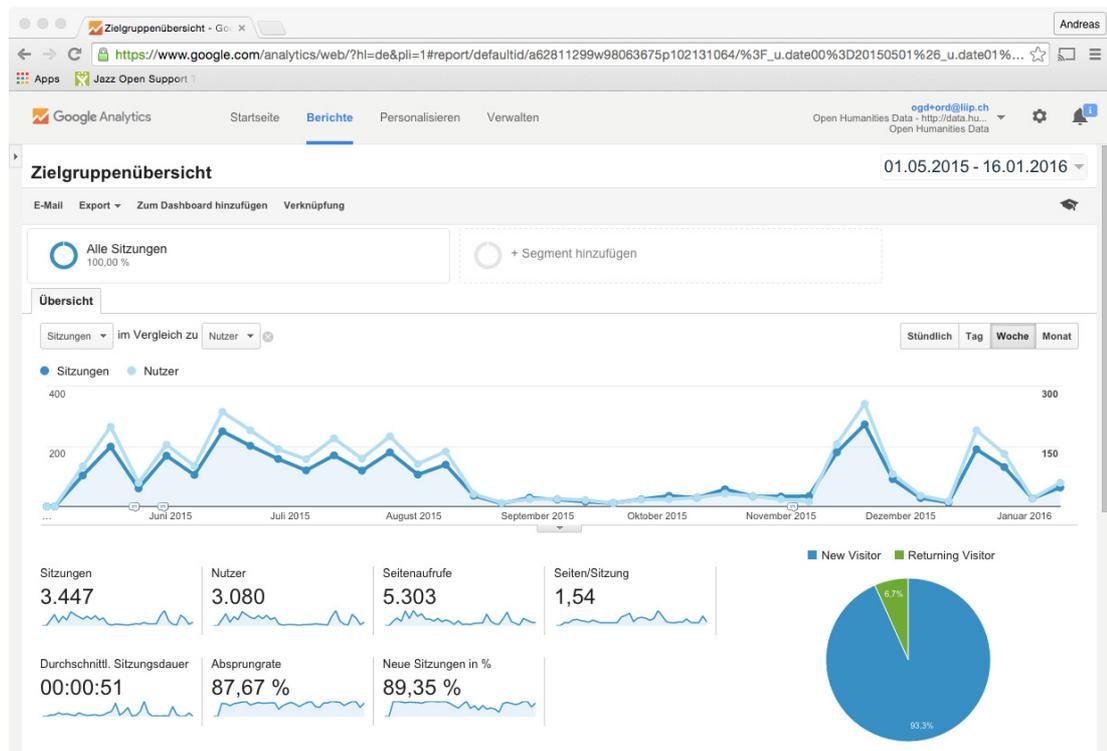
Der Quellcode und die Dokumentation der aufgeführten CKAN Extensions sind verfügbar unter:
<https://github.com/openresearchdata>.

3.5 Nutzerstatistiken im Pilotbetrieb

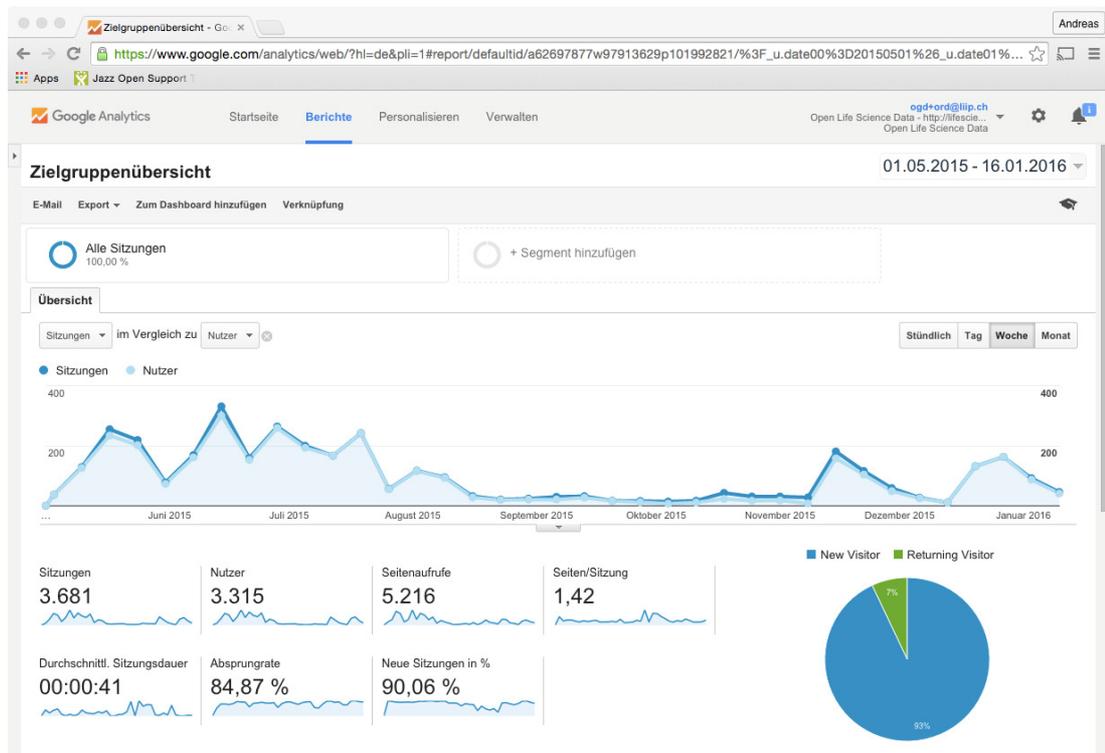
openresearchdata.ch



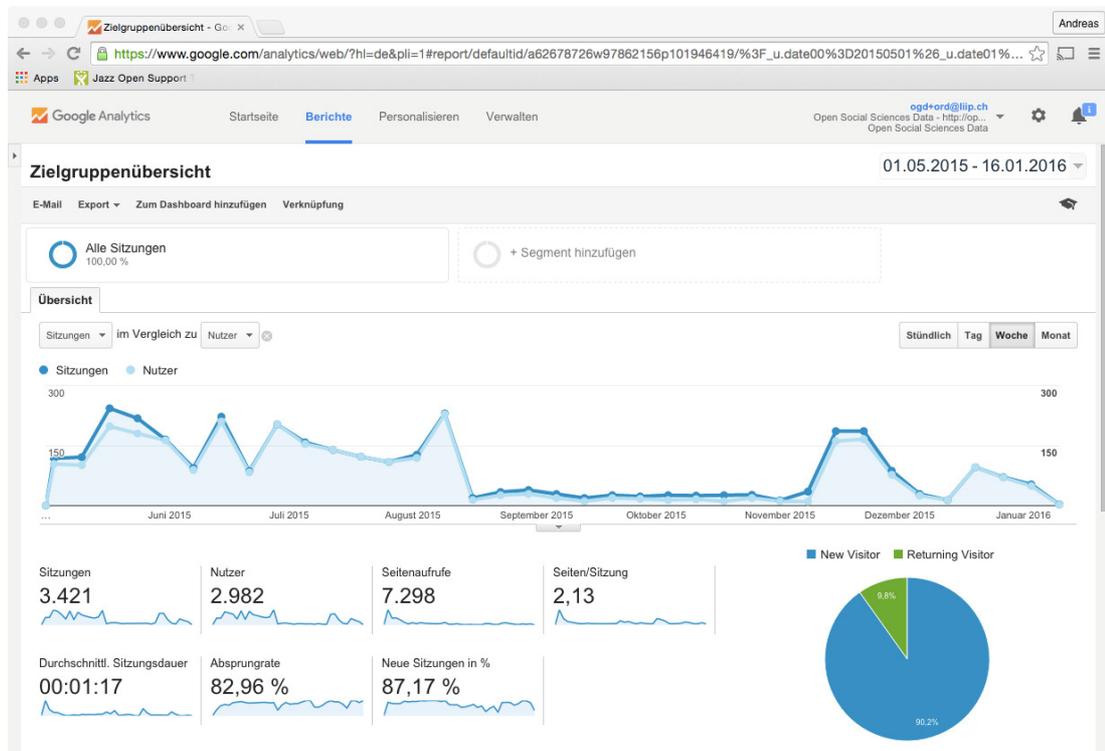
data.humanities.ch



lifesciencedata.ch



opendata.forscenter.ch



3.6 Aufbau und Betrieb der Plattform

Die vier CKAN-Instanzen werden von den Informatikdiensten der ETH Zürich innerhalb der hauseigenen virtuellen Serverinfrastruktur betrieben. Hosting auf virtuellen Servern bietet den entscheidenden Vorteil, dass die physische Server-Hardware (CPU, RAM, Speicherbedarf) effizient und zeitnah den Bedürfnissen angepasst werden kann. So lassen sich die Server bei hohen Zugriffsraten (z.B. im Rahmen von Research Data Hackdays) innerhalb kürzester Zeit aufrüsten oder bei entsprechend niedriger Auslastung auch wieder abrüsten, z.B. um Kosten zu sparen. Um den nahtlosen Betrieb und die gleichzeitige Weiterentwicklung der Plattform zu ermöglichen, wurde von ETH SIS ein technisches Betriebskonzept entwickelt, das für jede der vier CKAN-Instanzen sowohl einen produktiven als auch einen staging (Test) Server vorsieht:

CKAN Instanz	WWW Adresse	Produktiver Server	Staging Server
Central	<i>openresearchdata.ch</i>	<i>ord-ch-central.ethz.ch</i>	<i>ord-ch-central-stage.ethz.ch</i>
Social	<i>opendata.forscenter.ch</i>	<i>ord-ch-social.ethz.ch</i>	<i>ord-ch-social-stage.ethz.ch</i>
Life	<i>lifesciencedata.ch</i>	<i>ord-ch-life.ethz.ch</i>	<i>ord-ch-life-stage.ethz.ch</i>
Humanities	<i>data.humanities.ch</i>	<i>ord-ch-humanities.ethz.ch</i>	<i>ord-ch-humanities-stage.ethz.ch</i>

Im Rahmen der Entwicklungsarbeiten zum Aufbau der Plattform wurden sämtliche Konfigurationen zuerst auf den Staging Servern implementiert und getestet. Erst nachdem der stabile Betrieb der Staging Server sichergestellt ist, wird die entsprechende Konfiguration auf die produktiven Server übertragen. Ähnlich wird auch im laufenden Betrieb verfahren: alle Änderungen an den Servern (System Updates, CKAN Upgrades, Installation neuer Harvester etc) werden zuerst ausgiebig auf den Staging Instanzen getestet und dann auf die produktiven Server übernommen. Durch dieses Betriebskonzept lassen sich Wartungsarbeiten im laufenden Betrieb mit nur minimalem Unterbruch des Service durchführen (normalerweise nur wenige Minuten). Um die Anfälligkeit der Plattform gegenüber Sicherheitsrisiken zu reduzieren, sind die Staging Server zudem nur innerhalb des ETH Netzes erreichbar.

3.7 Kommunikation und Interaktion mit der Open Data Community

Am 5. Und 6. Juni 2015 hat der Verein Opendata.ch (www.opendata.ch) zusammen mit dem Pilotprojekt ORD@CH die ersten Open Research Data Hackdays in der Schweiz durchgeführt. Über 100 Personen beteiligten sich an den Hackdays, welche gleichzeitig an der ETH Zürich und an der Universität Lausanne stattfanden.

Die Resultate der Hackdays (9 Projekte) wurden auf dem Wiki [make.opendata.ch](http://make.opendata.ch/wiki/event:2015-06) publiziert und sind dort der Öffentlichkeit zugänglich (<http://make.opendata.ch/wiki/event:2015-06>).

An den Election Hackdays des Vereins Opendata.ch, welche am 4. Und 5. September 2015 in Zürich (Redaktion NZZ) und in Lausanne (Redaktion Le Temps) stattfanden, wurden die Daten zur SELECTS-Befragung 2011 von FORS (siehe <http://forscenter.ch/en/our-surveys/selects/>) für die Entwicklung der Visualisierungs-Applikation „Was wäre wenn Wahlen?“ genutzt (siehe

[http://make.opendata.ch/wiki/project:was waere wenn wahlen](http://make.opendata.ch/wiki/project:was_waere_wenn_wahlen)
<https://benjaminschlegel.ch/charts/waswaerewenn/>).

und

4. Nutzerbefragung

Vom 04.11.2015 bis 31.12.2015 nahmen 35 Personen an der frei im Web zugänglichen Nutzerbefragung teil. Die Befragung war verfügbar in Englisch (18 Teilnehmer), Französisch (3) und Deutsch (14).

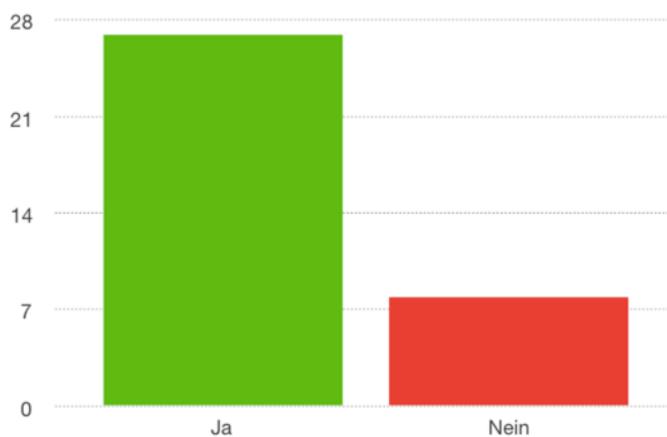
Die Teilnehmer konnten fakultativ Angaben zu ihrem beruflichen Hintergrund machen (24 Teilnehmer). Folgende Hintergründe wurden genannt:

- Forschungsinstitution (10 Teilnehmer)
- Bibliothekswesen (9)
- Informationswissenschaften (3)
- Informatik (2)

Die Nutzerumfrage umfasste die folgenden fünfzehn Fragen.

Kennen Sie openresearchdata.ch?

Zwei Drittel der Teilnehmer kannten openresearchdata.ch, bevor sie eine Einladung zur Nutzerbefragung erhielten.



Wie wurden Sie auf das Portal aufmerksam?

Zehn Teilnehmer wurden via eine SUK-Mailingliste auf das Portal aufmerksam, neun via Open Data Community-Aktivitäten, fünf durch eigene Recherche und die restlichen acht via verschiedene andere Kanäle.

Welchen Zweck hat das Portal?

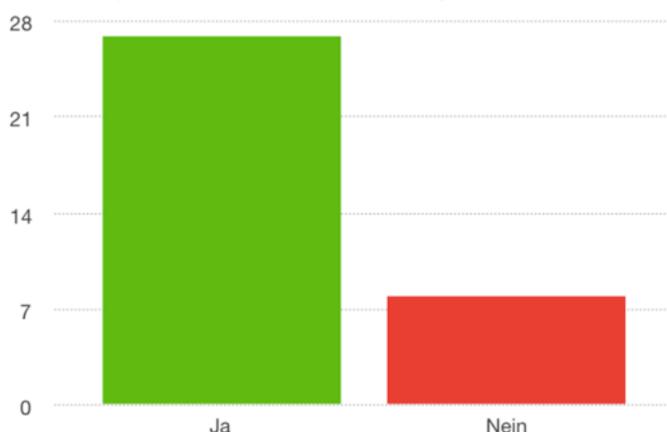
Mehr als zwei Drittel der Teilnehmer haben den Hauptzweck „Finden und Wiederverwenden von Forschungsdaten“ genannt. Eine interessante weitere Nennung ist „Langzeitarchivierung“ (2 Teilnehmer). Lediglich ein Teilnehmer gab an, dass für ihn der Zweck klar nicht ersichtlich sei.

Was können Sie auf dem Portal tun?

Die Antworten decken sich mit den Antworten zur Frage 4.3; wichtige Nennungen sind, dass auf den Portalen „noch zu wenige Daten zu finden“ (3 Teilnehmer) und die Suchmöglichkeiten zu verbessern sind (2 Teilnehmer).

Nehmen sie wahr, dass das Portal ein gemeinsames Projekt verschiedener Forschungsinstitutionen ist?

Für zwei Drittel der Teilnehmer ist ersichtlich, dass openresearchdata.ch ein forschungsinstitutionsübergreifendes Projekt ist.

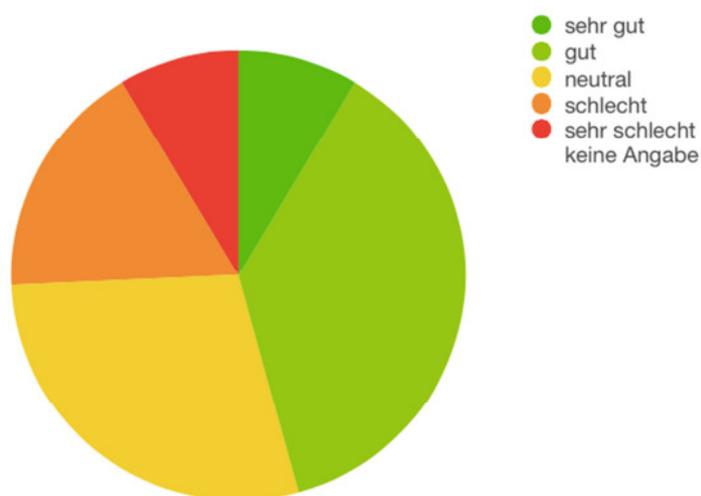


Wer gehört zur Trägerschaft des Portals?

Die Teilnehmer haben die an der geteilten Trägerschaft beteiligten Institutionen überwiegend erkannt.

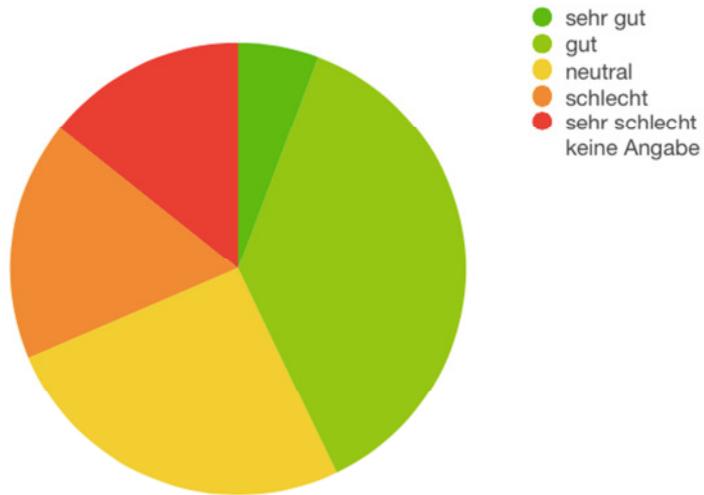
Wie empfinden Sie die Navigation auf dem Portal?

Knapp die Hälfte der Teilnehmer empfindet die Navigation als gut bis sehr gut, etwas mehr als ein Viertel als neutral und ein Viertel als schlecht bis sehr schlecht.



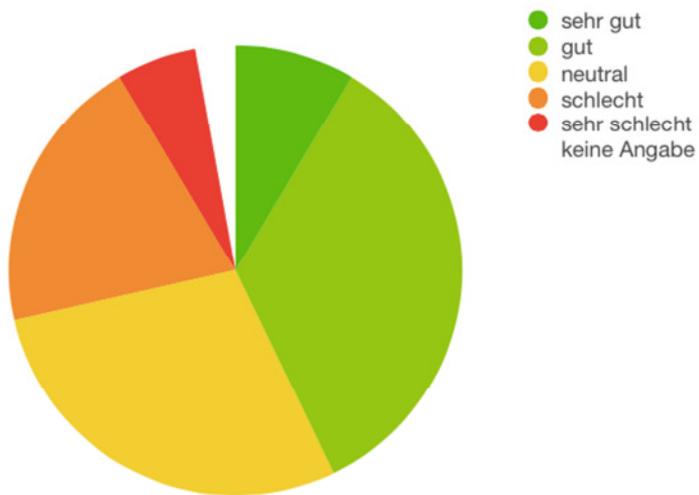
Wie beurteilen Sie die Möglichkeiten zur Suche von Datensätzen?

Knapp die Hälfte der Teilnehmer bewertet die Suchmöglichkeiten als gut bis sehr gut, ein Viertel als neutral und etwas mehr ein Viertel als schlecht bis sehr schlecht.



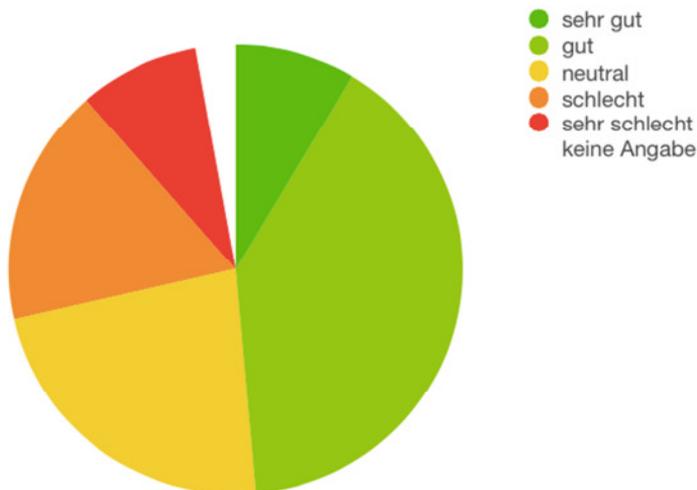
Wie beurteilen Sie die Darstellung der Suchresultate?

Knapp die Hälfte der Teilnehmer empfindet die Darstellung der Suchresultate als gut bis sehr gut, etwas mehr als ein Viertel als neutral und ein Viertel als schlecht bis sehr schlecht.



Wie beurteilen Sie die Darstellung und Beschreibung der einzelnen Datensätze?

Die Hälfte der Teilnehmer empfindet die Darstellung und Beschreibung eines Datensatzes als gut bis sehr gut, knapp ein Viertel als neutral und ein Viertel als schlecht bis sehr schlecht.



Welche Daten fehlen auf dem Portal?

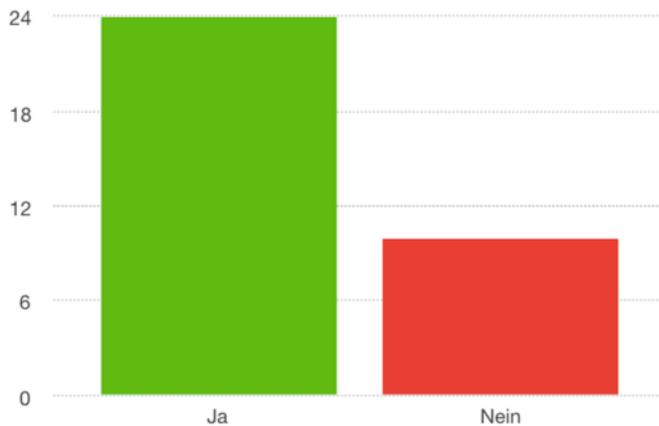
Einige Teilnehmer vermerken zurecht, dass auf den Portalen noch sehr wenige Daten verfügbar sind, was denn auch eine Aussage, welche Daten fehlen, erschwere. Folgende Forschungsbereiche wurden genannt:

- Geisteswissenschaften (2 Teilnehmer)
- Digital Humanities (2)
- Naturwissenschaften (2)
- Medizin (2)
- Statistische Daten (2)
- Life Sciences (1)
- Archäologische Daten (1)
- Bildung (1)
- „Technology“ (1)
- Wirtschaftswissenschaften (1)
- „micro economics data“, „experimental economics“ (1)
- Medien- und Kommunikationswissenschaften (1)
- Informations- und Kommunikationstechnologie (1)
- „Public Health“ (1)
- Physik (1)

Ein Teilnehmer weist speziell auf das Open Data Portal des CERN (Physik) hin. Ein Teilnehmer moniert, dass „no households per street and nr“ und „no type of house (single, block, twohouse....) per street and nr“ (Statistische Daten) verfügbar seien.

Sind Metadatenkataloge zu Forschungsdaten für Sie von Nutzen?

Zwei Drittel der Teilnehmer erklären, dass Metadatenkataloge zu Forschungsdaten für sie von Nutzen sind.



Haben Sie Bedürfnisse, bei denen das Portal Ihnen Hilfe bieten könnte?

Wiederum wurde zurecht genannt, dass noch zu wenige Daten auf den Portalen verfügbar sind.

Interessante Nennungen, wozu die Portale den Teilnehmern dienen könnten, sind:

- Für den „Re-use“ von Daten zu sensibilisieren;
- Bereich mit Apps/Showcases zu sehen, die den „Re-use“ klar vor Augen führen;
- Statistiken zu den verfügbaren Daten zu sehen;
- Finanzierungsquellen der Forschungsprojekte sichtbar zu machen;
- Datensätze zu bewerten zu („Rating“);
- Suchabfragen zu speichern und Benachrichtigungen zu abonnieren;
- Fehlende/weitere Daten anzufragen („Request new data“).

Haben Sie weitere Bemerkungen und Anregungen für uns?

Die Teilnehmer geben folgende interessante Hinweise:

- Metadatenstandard DCAT-AP Switzerland;
- Mehrsprachigkeit;
- Lizenzen der Datensätze sind klarer ersichtlich zu machen;
- Ebenfalls privatwirtschaftliche Forschungsinstitutionen als Datenlieferanten aufnehmen;
- Eine (oder mehrere) komplementäre Data Repository-Lösung(en) für Primärdaten.

Was fehlt dem Portal, damit es Ihnen dienen könnte.

Die Teilnehmer nennen folgende Verbesserungen:

- Open Research Data Governance;
- Endbenutzer können „related datasets“ angeben (und andere Endbenutzer wiederum von diesen weiterführenden Informationen profitieren);
- Datenlieferanten als Kategorien ausweisen;
- Domain-spezifische Begriffsglossare.

5. Beurteilung der Projektergebnisse und Ausblick

Die Ergebnisse des Projektes können grundsätzlich positiv beurteilt werden. Das Pilotprojekt ORD@CH hat alle operativen Ziele innerhalb des vorgegebenen zeitlichen und finanziellen Rahmens erreicht:

- Alle CKAN-Instanzen sind live und über die jeweiligen Web-Frontends zugänglich.
- Sowohl das Harvesting der Metadaten aus den fachspezifischen Datenquellen auf die lokalen CKAN-Instanzen (NESTAR, SALSAH, OpenBIS) als auch der Transfer der Metadaten zur zentralen CKAN-Instanz konnten erfolgreich implementiert werden.
- Der Betrieb der gesamten Plattform ist stabil und kann ohne grösseren zusätzlichen Aufwand von ETH SIS weitergeführt werden.

In unmittelbarer Zukunft ist ein Ausbau der Plattform in verschiedenen Bereichen geplant oder bereits im Gange:

- Die Erschliessung weiterer Datenquellen (ETH Data Archive, Plazi.org) wird bis spätestens März 2016 realisiert.
- Innerhalb der ETH besteht Interesse an weiteren fachspezifischen CKAN-Instanzen, z.B. in den Bereichen „Physical Sciences“ und „Climate Modelling“, um auf diesem Weg Forschungsdaten zu publizieren.
- FORS plant die Erschliessung des eigenen Datenarchives, das per 22. Februar 2016 in einer neuen Version live geht (FORSbase 2.0) via CKAN. In diesem Zusammenhang soll auch ein Open Repository aufgebaut werden, welches ausschliesslich offene Forschungsdaten aus den Sozialwissenschaften umfasst.

Der pilotmässige Aufbau und Betrieb einer Publikationsplattform für offene Forschungsdaten in der Schweiz im Sinne eines Metadaten-Hubs auf der Basis von CKAN hat sich bewährt. Der operative Betrieb sowie der schrittweise Ausbau des Datenangebotes durch die Erschliessung weiterer Datenquellen kann auf dieser Basis weitergeführt werden.

Dabei sind allerdings die folgenden Aspekte, welche auch in der Nutzerbefragung zum Ausdruck kommen, zu berücksichtigen:

- **Ausbau Datenangebot:**
Der Nutzen der Plattform ist direkt vom Umfang der publizierten Daten abhängig – je mehr Forschungsdaten zugänglich sind, desto grösser ist der Nutzen der Plattform, für Forschende, Studierende und weitere Interessierte.
- **Weiterentwicklung User-Interfaces:**
Die Web-Frontends wurden pilotmässig und mit minimalem Aufwand realisiert. Im Hinblick auf einen längerfristigen Betrieb der Plattform, sind diese auf die Bedürfnisse der Benutzer auszurichten und gründlich zu überarbeiten.

- Einbettung in einen Research Data Publishing Service:
Damit die Publikation der Forschungsdaten für die Forschenden mit minimalem Aufwand verbunden ist, sind die Funktionen „Archivierung“, „Identifikation (zwecks Zitierbarkeit)“ (z.B. mit DOI) und „Publikation“ (über eine oder mehrere Metadaten-Kataloge) als integrierter Service anzubieten.
- Kommunikation mit den Daten-Produzenten und –Nutzern:
Da die Publikation der Forschungsdaten noch lange keine Selbstverständlichkeit darstellt, braucht es eine intensive Kommunikation mit den Daten-Produzenten, um diese zur Veröffentlichung ihrer Daten zu motivieren, als auch mit (potenziellen) Daten-Nutzern, um ihnen das Potenzial der Wiederverwendung von Forschungsdaten zu demonstrieren.
- Positionierung im internationalen Kontext:
Die Plattform ist im internationalen Kontext zu positionieren, wo bereits vergleichbare Plattformen existieren oder aufgebaut werden (z.B. Zenodo.org, Datacite.org oder figshare.com). Dazu sind einerseits die Aufgaben der Plattform im nationalen Rahmen zu definieren (Dienstleistungen in Zusammenhang mit dem Data Life Cycle Management) und andererseits der Datenaustausch mit anderen Plattformen ins Auge zu fassen (z.B. auf Basis des international verbreiteten Metadaten-Standards DCAT-AP).

Das Folgeprojekt „researchdata.ch“ ist auf diese Aspekte ausgerichtet.