

The organisational and technical aspects of establishing plagiarism detection processes in the Slovenian open access infrastructure

Milan Ojsteršek, Goran Hrovat

University of Maribor, Faculty of Electrical Engineering and
Computer Science

milan.ojstersek@um.si; goran.hrovat@um.si

Development financed by:



REPUBLIC OF SLOVENIA
MINISTRY OF EDUCATION,
SCIENCE AND SPORT



Investing in your future
OPERATION PART FINANCED BY THE EUROPEAN UNION
European Regional Development Fund

Partners:



University of Ljubljana



The Slovenian open access infrastructure is partly financed by the European Union, European Regional Development Fund and Ministry of Education, Science and Sport of the Republic of Slovenia within the framework of the Operational Programme for Strengthening Regional Development Potentials for Period 2007 - 2013.



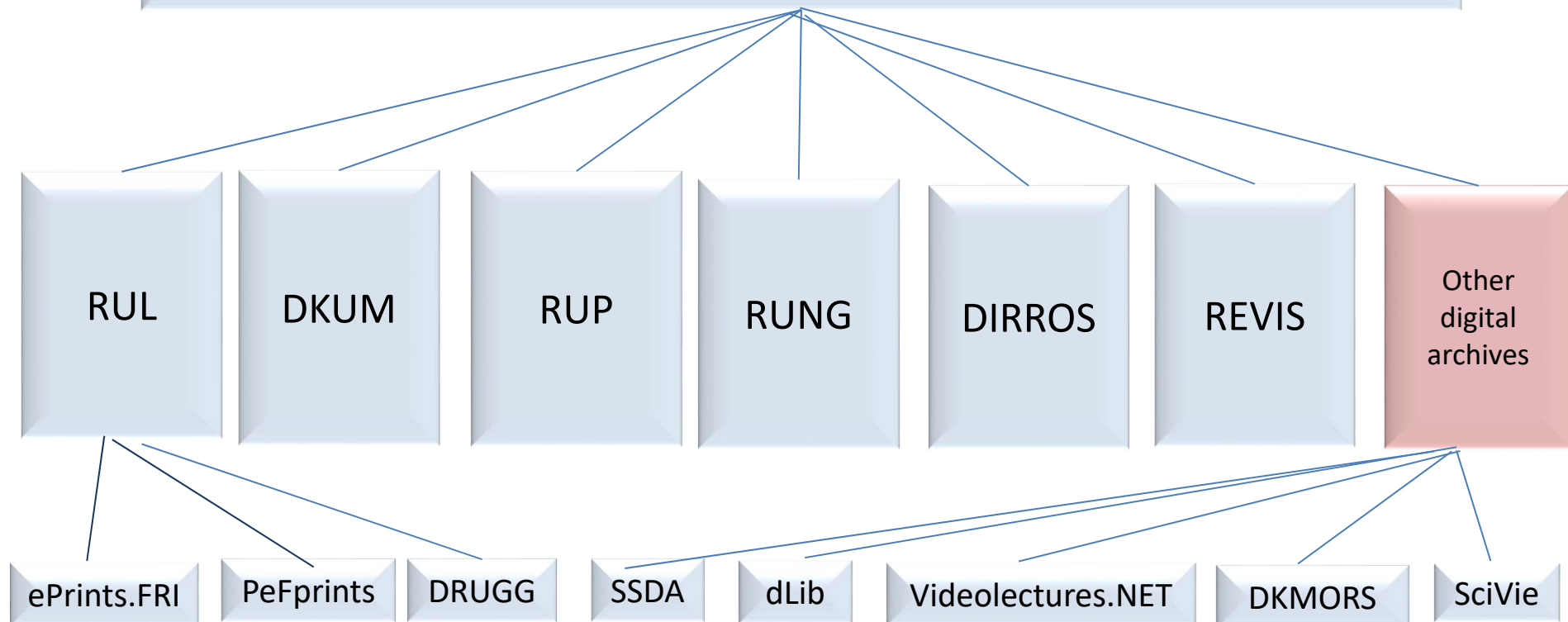
The state of open access and academic integrity in Slovenia in 2012

- More than 40 open access journals and more than 20 digital archives and repositories.
- Publications have been duplicated across different repositories or digital archives.
- Inadequate quality of metadata of research publications, research data, dissertations and theses.
- Processes for submission of final study works (theses and dissertations), research publications, research data and other academic production have not been established.
- No common search engine for theses, dissertations, research publications, research data and other academic production.
- Lack of common recommender system and plagiarism detection system.
- Lack of metadata, data, content, submission and preservation policies.
- Lack of documents for supporting academic and research integrity.

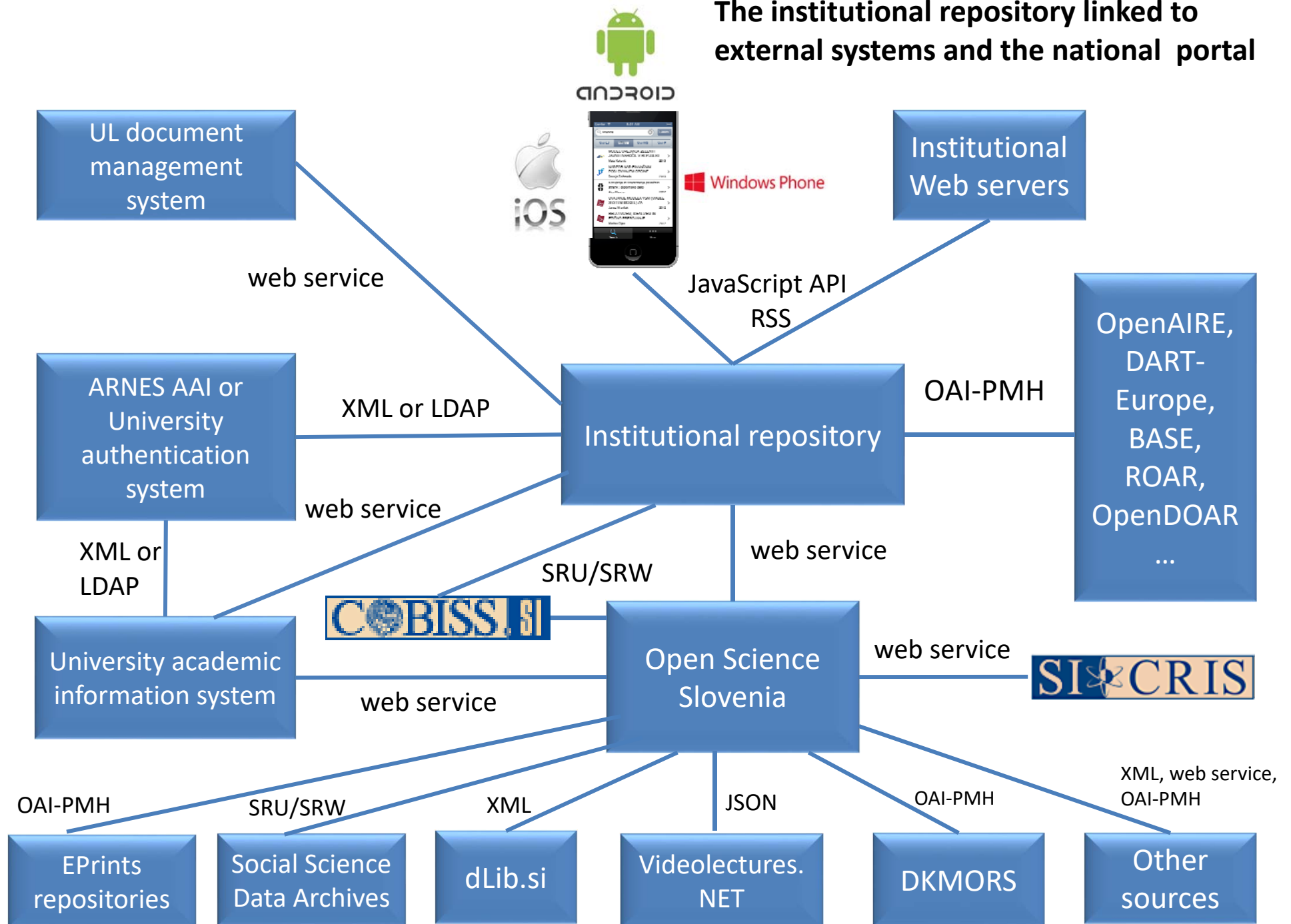
A network diagram of the Slovenian open access infrastructure

Open Science Slovenia (<http://www.openscience.si>)

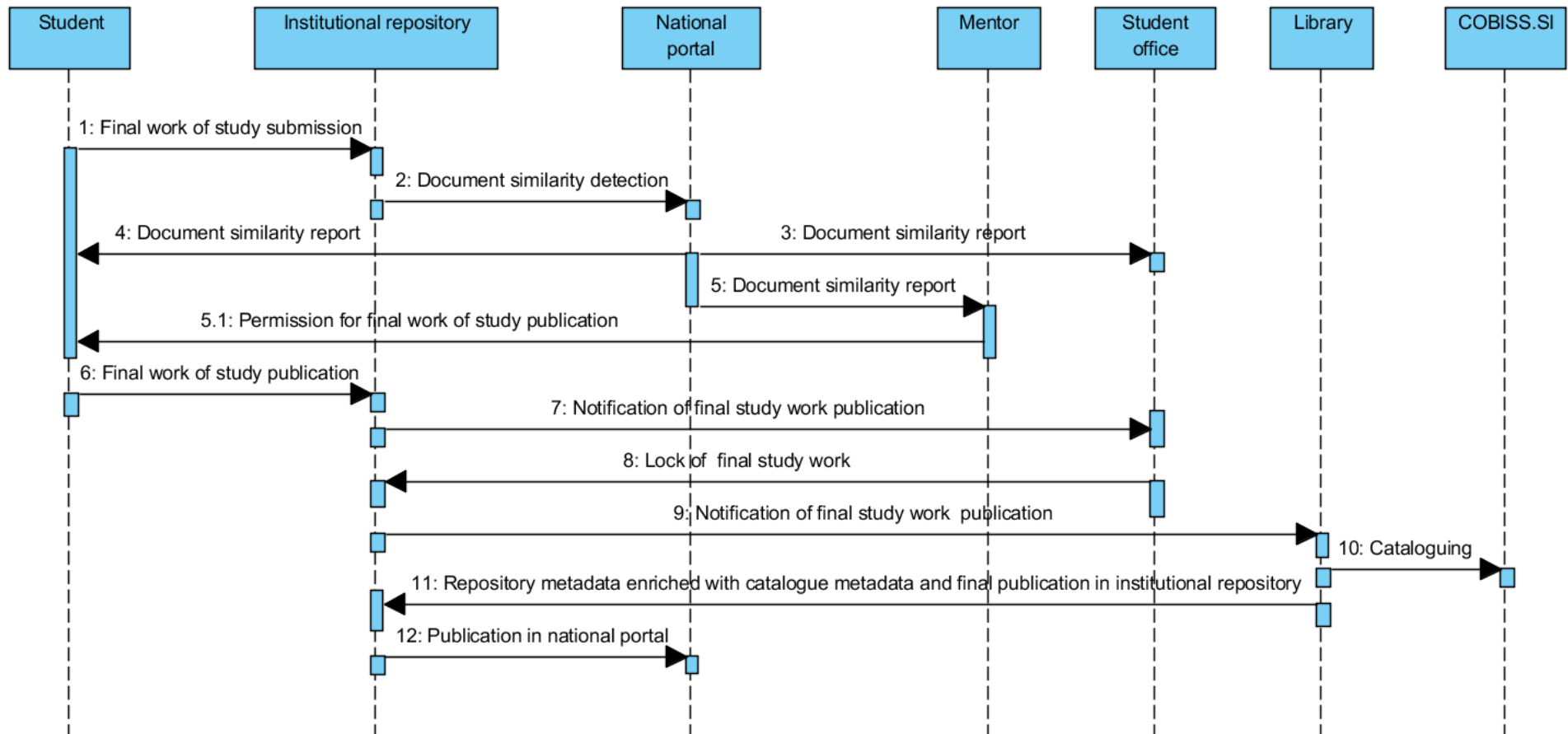
- similar content detection,
- federated search,
- recommendation system.



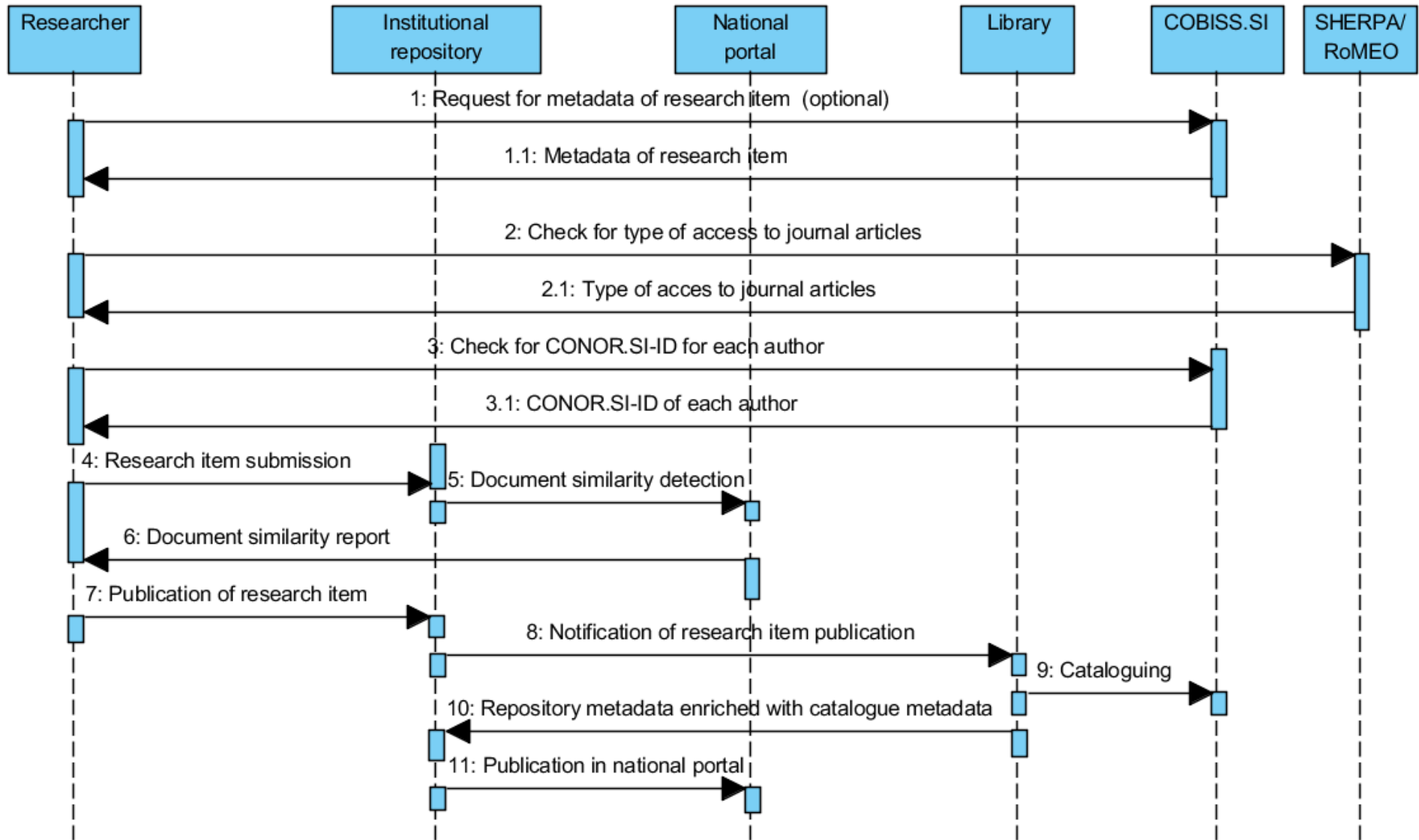
The institutional repository linked to external systems and the national portal



A sequence diagram of final study work submission and publication at the universities of Maribor and of Nova Gorica



A sequence diagram of research item submission and publication



Documents that are base for academic integrity at Slovenian universities

- Statute.
- Code of Ethics.
- The European Code of Conduct for Research Integrity.
- Rules on protection of intellectual property.
- Open access policy.
- Rules on verifying and assessing student's knowledge.
- Rules of preparation and defence of final study works.
- Instructions for checking the similarity of the content in the final study work.
- Plagiarism policy.
- Guides for plagiarism prevention and awareness.

Our test corpus of 80011 documents (average similarity is 19,29%)

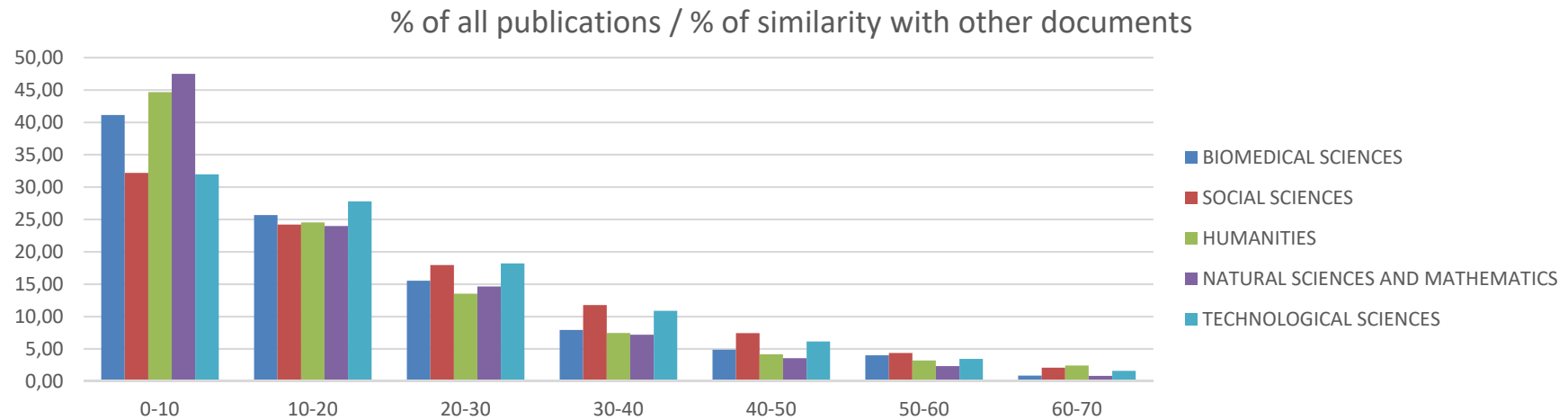
SCIENTIFIC AREA	number of publications	percent of publications	average similarity	difference from average similarity	standard deviation of % similarity
BIOMEDICAL SCIENCES	1250	1,56	16,49	2,80	15,49
SOCIAL SCIENCES	45442	56,79	20,46	-1,17	16,26
HUMANITIES	1153	1,44	15,90	3,39	16,05
NATURAL SCIENCES AND MATHEMATICS	9908	12,38	14,36	4,93	14,35
TECHNOLOGICAL SCIENCES	22258	27,82	19,43	-0,14	15,11
SUM	80011	100,00			
Publications typology	number of publications	percent	average similarity	difference from average similarity	standard deviation of % similarity
Scientific paper	3271	4,09	15,39	3,90	16,88
BSc Thesis	53327	66,65	20,74	-1,45	15,78
PhD thesis	1577	1,97	8,56	10,73	9,95
MSc thesis	9906	12,38	16,84	2,45	13,17
Other	11930	14,91	17,34	1,95	17,20

Distribution of publications with regard to their percent of similarity

SCIENTIFIC AREA / percent of similarity	0-10	10-20	20-30	30-40	40-50	50-60	60-70	SUM
BIOMEDICAL SCIENCES	514	321	194	99	61	50	11	1250
SOCIAL SCIENCES	14626	10994	8155	5352	3378	1983	954	45442
HUMANITIES	515	283	156	86	48	37	28	1153
NATURAL SCIENCES AND MATHEMATICS	4705	2376	1451	712	352	231	81	9908
TECHNOLOGICAL SCIENCES	7112	6185	4051	2420	1366	769	355	22258

SCIENTIFIC AREA / percent of publications with percent of similarity	0-10	10-20	20-30	30-40	40-50	50-60	60-70
BIOMEDICAL SCIENCES	41,12	25,68	15,52	7,92	4,88	4,00	0,88
SOCIAL SCIENCES	32,19	24,19	17,95	11,78	7,43	4,36	2,10
HUMANITIES	44,67	24,54	13,53	7,46	4,16	3,21	2,43
NATURAL SCIENCES AND MATHEMATICS	47,49	23,98	14,64	7,19	3,55	2,33	0,82
TECHNOLOGICAL SCIENCES	31,95	27,79	18,20	10,87	6,14	3,45	1,59

Distribution of publications with regard to their percent of similarity



Distribution of publications by issue year with regard to their percent of similarity

Publications with more than 20% and less than 70 % of similarity with other documents

SCIENTIFIC AREA / number of publications per year (N)	2000-2005	2006-2010	2011	2012	2013	2014	2015	2016	SUM
BIOMEDICAL SCIENCES	8	127	70	54	39	21	26	12	357
SOCIAL SCIENCES	2617	5219	2094	1767	1348	1368	1288	1802	17503
HUMANITIES	39	117	41	36	27	27	5	9	301
NATURAL SCIENCES AND MATHEMATICS	75	667	311	303	278	329	264	258	2485
TECHNOLOGICAL SCIENCES	1787	2681	730	710	537	466	476	621	8008

Publications with less than 70 % of similarity with other documents

SCIENTIFIC AREA / number of publications per year (N all)	2000-2005	2006-2010	2011	2012	2013	2014	2015	2016
BIOMEDICAL SCIENCES	65	311	134	128	130	118	111	94
SOCIAL SCIENCES	5100	10368	4211	3931	3399	3406	4038	5965
HUMANITIES	160	308	108	114	103	85	61	77
NATURAL SCIENCES AND MATHEMATICS	514	1702	798	831	873	936	1015	1282
TECHNOLOGICAL SCIENCES	2691	5437	1759	1797	1671	1558	2084	3137

Distribution of publications by issue year with regard to their percent of similarity

SCIENTIFIC AREA / (N/N all) peer year	2000-2005	2006-2010	2011	2012	2013	2014	2015	2016
BIOMEDICAL SCIENCES	12,31	40,84	52,24	42,19	30,00	17,80	23,42	12,77
SOCIAL SCIENCES	51,31	50,34	49,73	44,95	39,66	40,16	31,90	30,21
HUMANITIES	24,38	37,99	37,96	31,58	26,21	31,76	8,20	11,69
NATURAL SCIENCES AND MATHEMATICS	14,59	39,19	38,97	36,46	31,84	35,15	26,01	20,12
TECHNOLOGICAL SCIENCES	66,41	49,31	41,50	39,51	32,14	29,91	22,84	19,80

Distribution of publications by issue year with regard to their percent of similarity

Percentage of publications with more than 20% and less than 70 % of similarity with other documents / year



Where you can find more information

- National portal of open science: <http://www.openscience.si/Default.aspx>
- DKUM: <https://dk.um.si/info/index.php/eng/>
- RUL: <http://repozitorij.uni-lj.si/info/index.php/eng/>
- RUP: <http://repozitorij.upr.si/info/index.php/eng/>
- RUNG: <http://repozitorij.ung.si/info/index.php/eng/>
- DIRROS: <http://dirros.openscience.si/info/index.php/eng/>
- REVIS: <http://revis.openscience.si/info/index.php/eng/>

- Our publication:
Milan Ojsteršek , Janez Brezovnik , Mojca Kotar , Marko Ferme , Goran Hrovat , Albin Bregant , Mladen Borovič , (2014) "[Establishing of a Slovenian open access infrastructure: a technical point of view](http://www.emeraldinsight.com/doi/full/10.1108/PROG-02-2014-0005)", Program: electronic library and information systems, Vol. 48 Iss: 4, pp.394 – 412
<http://www.emeraldinsight.com/doi/full/10.1108/PROG-02-2014-0005>

Contact

Milan Ojsteršek

University of Maribor, Faculty of electrical
Engineering and Computer Science

tel.: +386 2 220 74 51, +386 40 696 538

email: milan.ojstersek@um.si