Project no. 018340

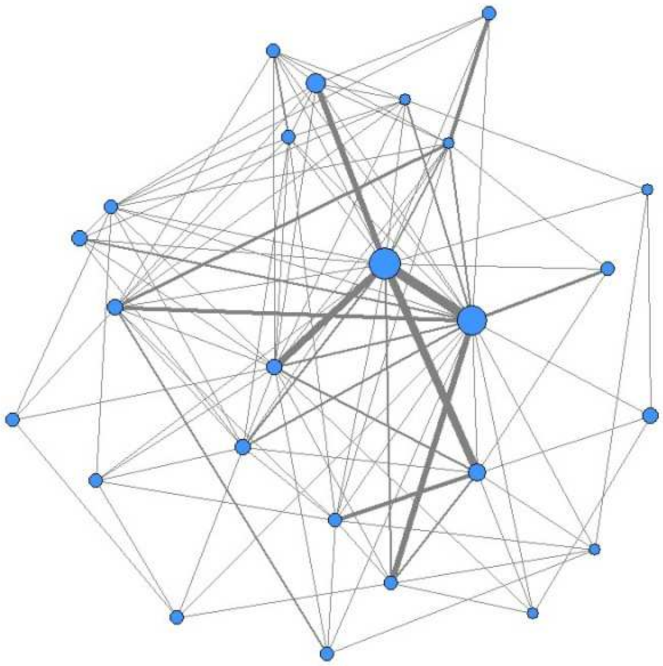
Project acronym: EDIT

Project title: Toward the European Distributed Institute of Taxonomy

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

C4.2.6 Report on social network analysis and bibliometrics to map actors in taxonomy



Due date : Month 35

Actual submission date: Month 35

Version 30/07/09

Organisation name of lead contractor for this deliverable: 2 – MNHN

Start date of project: 01/03/2006 Duration: 5 years

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of content

Preface....3

Executive summary....4

Introduction....5

1 Social network analysis and bibliometrics....5

2 Data and data processing....6

2.1 Data processing: EDIT co-authorships....6

3 Results....8

3.1 Density of the EDIT co-authorship network....8

3.2 Centrality of the EDIT co-authorship network....9

3.3 K-core partition of the EDIT co-authorship network....13

3.4 Domain partition of the EDIT co-authorship network....14

4 Discussion and next steps....15

References....17

Annex 1 – Acronyms....18

Annex 2 – Key words EDIT institutional addresses....19

Preface

Objective 5 of EDIT aims to build partnerships with the taxonomic stakeholder community. EDIT tackles this by being proactive in engaging with stakeholders directly and in parallel takes a more analytical stand and collects information about the stakeholder community. This information is to help target EDIT activities towards stakeholders and the individual EDIT partners to connect with taxonomic stakeholders. EDIT activities focus on the following stakeholder groups:

- Scientific users of taxonomic knowledge (professional and amateur taxonomists, agriculture, archaeology, conservation science, ecologists, environmental studies, genomics, genetics, molecular sciences, medicine etc.)
- Professional users (conservation management, environmental assessment industry, custom services, pest management, cosmetic and pharmaceutical industry, national governments, international governance bodies etc.)
- Taxonomic facilitators (funding bodies, educational facilitators, developers of research tools, taxonomic information facilities etc)
- The general public (nature lovers and anyone using vernacular or scientific names)
- Other networks and projects in taxonomy and biodiversity science.

This component report describes the work carried out under WP 1 and WP 4 and aims to contribute to the understanding of the EDIT community and the network in which it operates.

For comments and questions about the report contact Daphne Duin – duin@mnhn.fr

Executive summary

In this paper we explore whether and how social network analysis indicators and bibliometrics can be used to compute network maps, useful for strategic decision making purposes in EDIT. We used co-authorship statistics of EDIT members to explore the usability of the methodology and their potential to study larger networks of actors in taxonomy and taxonomy related areas.

The data we draw on in this report are scientific papers indexed in the ISI Web of Science, and co-author networks were constructed by using the social network analysis tools Netdraw and Pajek. The results show that co-authorships between EDIT members generate a "natural" network at research level. Between 2005 and 2008 every EDIT member published paper(s) with at least 4 different EDIT members and no isolated mini networks occurred in the larger network. The analysis revealed that to study the networks' characteristics it is useful to look at highest numbers of co-authorships between two institutions *and* at the 'degree centrality', the occurrences of 'cores', and 'domain' in the network.

This exercise and its follow up aim to feed into EDITs science policy activities and when applied to a lager actor network - will serve to identify possible partners form outside the taxonomic research community.

Introduction

Previous investigations of stakeholder (user) engagement in biodiversity science have demonstrated this is not a straightforward thing to do and that stakeholder involvement stays often limited to the usual suspects only [C4.1.2BIS]. Yet, EDIT aims to have a fresh look at the (potential) actors operating in the field of taxonomy and wants to develop systematically founded observations on the actors in the taxonomic fields. In this context EDIT works with the Science System Assessment Centre of the Rathenau Institute in The Hague on a network analysis of taxonomy. In particularly for EDIT strategic planning purposes it is useful to have a map of the taxonomic knowledge environment and the different actor relations within. This social network mapping exercise looks at EDIT members only. A follow up study should focus on wide network(s) and the stakeholders from outside the taxonomic research environment[1].

In this report the general question we asked ourselves was – do the EDIT institutions make up a research community? To study this question we carried out a social network analysis of the co-authorships among EDIT consortium members. This report describes the potential of a social network analysis and bibliometric methodologies to study taxonomy science networks. In this report a theoretical introduction is followed by illustrations of concrete examples from the EDIT co-author network. In the context of the EDIT objectives the mapping of affiliation networks between the EDIT institutions are of particular interest. An investigation of co-author networks is relevant because cross-institutional co-authorships are common practice in science and what is more, a co-author network reflects true professional interaction between scientists as they are genuinely acquainted with one and other (Newman, 2001, p.2). Our exercise aims above all to map research *relations* among EDIT partners; it is not an assessment on the quality of taxonomic research. EDIT and the Rathenau Institute started their collaboration in 2008 (EDIT Third JPA) and will continue to work together in 2009 (EDIT Fourth JPA).

The following paragraphs introduce theoretical and methodological concepts of the analysis, an explanation of the data, a presentation of the social network indicators with computed illustrations and we conclude with a discussion of the results.

1 Social network analysis and bibliometrics

In this section we will give a brief introduction to the theoretical perspective and the methodological approach which EDIT and the Rathenau Institute use in their forthcoming study of the taxonomic field. The study and the examples in this report draw on analytical concepts from social network analysis theory and on methodological concepts from bibliometrics.

To begin with the social network analysis, this is a distinct research perspective based on the assumption of the importance of relationships among interacting units. The linkages among units are the fundamental components of analysis. A social network is a system of a set of people or groups each of which has connections of some kind to some or all of the others. In the language of social network analysis, the people or groups are called "actors" and the connections "ties" or "paths". (cf. Wasserman & Faust, 1994; Newman, 2001). Wasserman and Faust (idem, p. 5) describe the following central principles as being important in the social network analysis perspective: 1) actors are interdependent rather than independent autonomous units; 2) relational ties (linkages) are channels through which (non)material resources flow; 3) the network environment provides opportunities for or constraints on individual action; 4) network models

[1] The elements discussed in this report are part of a forthcoming scientific study. Please make reference to this report or contact the author when using information described.

describe structure (social, economical, political, and so fourth) as lasting patterns of relations among actors. In the context of the objectives of the EDIT programme the affiliation networks between the EDIT member institutions have our special interest.

In addition, the bibliometric methodological approach needs a brief introduction. The term 'bibliometrics' is related to and has overlapping interests with 'scientometrics' and 'informetrics' and is often used synonymously for all three metrics. The three "metrics" refer "to the study of the dynamics of disciplines as reflected in the production of their literature" (Hood & Wilson, 2001, p. 291). Borgman and Furner (2002 In: Thelwell, 2007, p. 2) define bibliometrics as a study which "encompasses the measurement of properties of documents and document related process". Bibliometrics is primarily applied to science related documents. Because publications are the central output in science, bibliometrics of scientific papers offer many possibilities to study relations and dynamics of science and offer methods for systematic investigation of science networks. Bibliometrics can also be used for science evaluation purposes, but we will focus here solely on relational bibliometrics. Traditionally, bibliometric analysis is based on publications stored in databases like those of the Institute for Scientific Information (ISI). The ISI has several databases and indexes that generate information on a large part of the scientific publications published worldwide. One of ISI databases is the Web of Science (WoS) which has a number of sub-indexes encompassing a large number of science documents over a long time span. For instance, the Science Citation Index EXPANDED (SCI-EXP), has records and publications which go back to the year 1900.

Although bibliometric techniques offer many possibilities to study science networks they are limited in certain ways. Bibliometrics are restricted to categorised, indexed information. One needs very large data sets to be able to draw statistically significant conclusions; this makes the analysis a very labour intensive work practice. Furthermore, precision of bibliometrics techniques depends on the accuracy, access and organisation of the data and databases. This is something one does not always control e.g. when the databases are owned and managed by others.

One of the objectives of the exercise described in this report is to examine if we are able to compute co-authorship networks at an institutional level. We used social network analysis software tools Pajek and Netdraw.

2 Data and data processing

The data we discuss in this report are based on bibliographic information of scientific papers, downloaded from the ISI databases (SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, IC). As mentioned above, although the ISI is one of the largest scientific information providers at the moment, it is not all inclusive. An important number of peer-reviewed journals are not available in the ISI. This includes a large number of journals published by EDIT member institutions (Duin, 2009), but also other types of scientific information are not included in the ISI. This is in particular important in taxonomy as taxonomists publish new species descriptions in scientific series like Monographs, Floras, Faunas and in online databases. Besides publishing papers, taxonomists spend time on and collaborate with each other in scientific activities like: teaching; describing work processes; maintaining databases; participation in meetings and conferences; curating collections; field work etc. These activities may be listed in institutional databases or on the web but are not included in the ISI data sets. This to say that the documents in the ISI do not represent all scholarly activities of taxonomists and therefore only cover one part of their network activities as do our conclusions in this report.

2.1 Data processing: EDIT co-authorships

In the data set used for the exercise described in this report we tried to capture the collective of all co-authored papers published in the years 2005-2008 and signed by at least 2 authors with 2 different EDIT author addresses. Because the data set covers all papers published in this period it may include, next to taxonomic research, also publications from other field's represented in the institution, like anthropology, mineralogy, geology etc. To analyse the collaborations in the EDIT network we downloaded all papers from 25[2] EDIT partners that were indexed in the Web of Science (WoS).

To be able to capture all papers signed with an the EDIT address we went through several preparatory steps. The addresses in the WoS are the addresses the individual author has chosen and so a wide variety of address formulations per institutions exist. Therefore we first had to create an address list that would cover as many address spelling varieties as possible[3]. The different organisational name spelling forms that were yielded and the complexity of standardisation process can be illustrated by the following two examples[4]: 1) For EDIT partner 3 we found at least two spelling varieties: "Nat Hist Museum, Bot Garden & Museum, Copenhagen, Denmark" and in another paper "Univ Copenhagen, Nat Hist Museum, Zool Museum, DK-2100 Copenhagen, Denmark"; 2) for EDIT partner 8, "Cent Bur Schimmelcultures, Fungal Divers Ctr, NL-3508 AD Utrecht, Netherlands" and in a second publication as "CBS Fungal Biodivers Ctr, NL-3508 AD Utrecht, Netherlands". In addition, the language used for the spelling of addresses was not always standardised. The Belgian partners had papers indexed under French, Dutch and English spelled addresses, the German partners in German and English (Botanical, Botanischer, Botanic, Museum, Musée, Belgium, Belgique and België).

Other, more extensive standardisation issues had to be addressed to reach optimum between precision and recall[5]. From the EDIT partners belonging to larger scientific organisations like universities or international research centres we included only the papers with addresses from departments likely to work on natural history and taxonomic topics (e.g. for the Smithsonian only the National Natural Museum History (partner 25) and the Freie Universität Berlin and the Botanic Garden and Botanical Museum (partner 9), INRA only the publications from the department Centre de Biologie et de Gestion des Populations (partner 21). After the standardisation of research addresses we were able to locate a total of 11780 papers, signed with at least one EDIT address. With these papers we built an Access database using the ISIParser (©Rathenau Institute) and we were able to build a matrix of co-authored EDIT papers. Then, where possible, the papers were checked manually for single authors who had signed a paper with 2 different EDIT addresses. In the database this looked like a co-authorship between 2 institutions while in reality it was a paper of one EDIT author only. For instance it occurred that one author was affiliated to two different EDIT institutions at the same time or that one author had changed employers and had moved from one EDIT institution to another EDIT institution during the writing up process of the publication. In the WoS addresses field this author was indexed with two different addresses, one under "address" (ad=) and the other one in the so called "reprint address" field of the database. These "fake co-authored" records were deleted from the co-authored paper list. Then with the EDIT co-author matrix we were able to compute

---

[2] Caisses des Depots, Species 2000 and the Society for management of European biodiversity data (SMEB) did not have ISI indexed papers.

[3] Annex 2

[4] Examples are randomly chosen, similar standardisation problems exist for most addresses of the EDIT partners.

[5] *Precision* in search techniques measures how well it weeds out what one does not want, *recall* the measure of completeness

the network analysis in Netdraw and Pajek. The results of the analysis are discussed in the following section.

3 Results

Two institutions were considered connected if they (their researchers) had co-authored on one or more papers together. Then with the help of the software tools Netdraw and Pajek we were able to compute a variety of statistical properties of the EDIT co-author network. We will discuss in this report a number of network characteristics like centrality and connectedness of the actors. The actors in the network are the EDIT member institutions (also called "nodes") the relational ties among actors are the co-authorships (in the maps illustrated as "strings" between a pair of actors). These measures give us insight into the various roles and groupings of the actors in the network.

3.1 Density of the EDIT co-authorship network

The nodes in Figure 1 are the EDIT institutions that co-authored on one or more papers together between 2005 and 2008. The node size indicates the relative number of publications of the institution. The tie strength indicates the number of co-authorships between two institutions. In Table 3 we have listed an overview of the total number of indexed papers in the WoS per institution versus the total number of co-publications with another EDIT institution. All EDIT members with indexed papers in the WoS are in the network which means that between 2005 and 2008 they all produced at least one co-authored paper with one researcher from another EDIT institution.

To have statistical information of this computed relational network we started with measuring the *density* of the EDIT co-authorship network. The network density is an indicator for the general level of connectedness of the graph. We talk about a complete graph if every institution is directly connected to every other institution in the network. The density of a network is defined as the number of linkages divided by the number of nodes (institutions) in a complete graph with the same number of nodes. Wasserman and Faust describe how to quantify the connectedness of a network with a density measurement (1994, p.101). As they explain, in a network graph there are *g* nodes (loops excluded), there are $\binom{g}{2}$ =g(g-1)/2 possible unordered pairs of nodes and thus g(g-1)/2 possible lines that could be present in the network graph. The density of a graph is the portion of possible lines that are actual present in the graph, it is the ratio of the number of lines present, *L*, to the maximum possible. The density of the network graph is denoted Δ and is formulated as followed:

$$\Delta = \frac{L}{g(g-1)/2}$$

In Table 3 the figures for the EDIT co-author network are listed. Based on these figures the density of our network should be calculated as followed (note that we divided 222 linkages by 2 as co-authorship between institution A and B is considered to be the same as a co-authorship of B and A:

$$\Delta = \frac{111}{25(25-1)/2} \quad 0.37$$

A density of 0.37 means that 37 percent of all possible linkages is present. The density figure will become more meaningful when doing comparative analysis over different periods or when

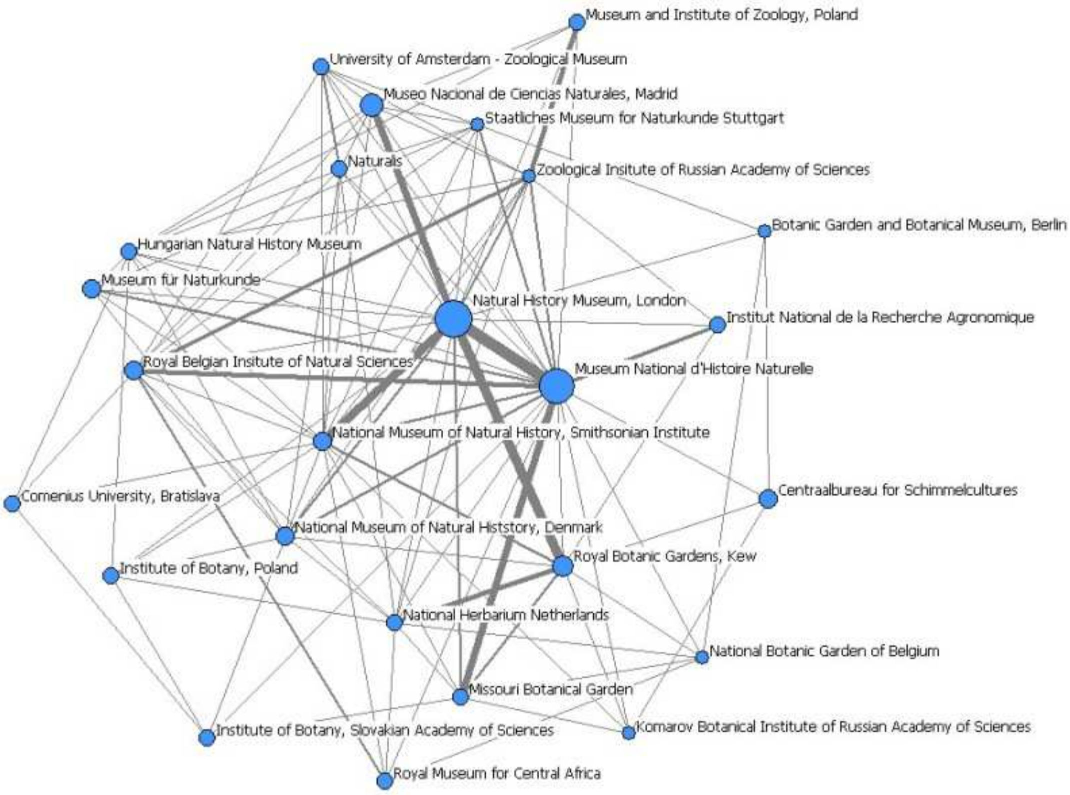


Figure 1. Co-authorships among EDIT institutions between 2005-2008. Node size indicates the total number of publications of an institution, tie strength the number of co-authorships between two actors, total number of co-authorships in network 970 (total number of papers published 11780, 25 actors, based on ISI-WoS)

comparing the EDIT network with other networks of science institutions Next to the *density* indicator, the 'centrality', the occurrences of 'cores', and 'domains' in the network were analysed.

3.2 Centrality of the EDIT co-authorship network

To understand the EDIT co-author network, we evaluated the location of EDIT institutions in the network. Measuring the network location is finding the *centrality* of a node. The centrality measure gives us insight in the various roles and groupings in a network -- who are the connectors, leaders, bridges, isolates, where are the clusters and who is in them, who is in the core of the network, and who is on the periphery (Krebs, 2008). Here we distinguish *betweenness centrality* and *degree centrality*.

Betweenness centrality "may be defined loosely as the number of times a node needs a given node to reach another node" (Otte & Rousseau, 2002, p. 443). Interactions between two nonadjacent actors might depend on the other actors in the set of actors, especially on the actors who lie on the paths between the two. Wasserman and Faust (1994) explain that "these other actors" potentially might have some control over the interactions between the two nonadjacent actors. In our co-author network betweenness centrality provides us with information which looks beyond the number of co-authorships that an institution has co-published. An actor can have a small number (fewer than the average in the network) of co-authorships relations (ties) with different actors in the network and still be one of the best located actors in the network - in between two important constituencies. The betweenness centrality figures for our network are listed in Table 1. In our example most institutions with many ties also have a high betweenness centrality (betweenness centrality is always between 0-1).

Table 1. EDIT co-authorships 2005-2008. Betweenness centrality and number of co-author pairs(based on papers from ISI)

Order number	Institution name as in EDIT's consortium agreement	betweenness centrality	Number of pairs (MNHN published with 21 institutions)
1	Muséum National d'Histoire Naturelle (MNHN)	0.2182703	21
2	Natural History Museum, London (NHML)	0.1157066	18
3	National Museum of Natural History, Smithsonian Institute, Washington (UNSM)	0.1015433	17
4	National Herbarium Netherlands (NHN)	0.0466586	12
5	Royal Belgian Institute of Natural Sciences, Brussels (RBINS)	0.0331281	12
6	Hungarian Natural History Museum (HNHM)	0.0286595	11
7	Zoological Museum, National Museum of Natural History, Denmark (UKBH-NHMD)	0.025463	12
8	Zoological Institute of Russian Academy of Sciences (ZINRAS)	0.0189876	11
9	Staatliches Museum for Naturkunde Stuttgart (SMNS)	0.0166365	9
10	Royal Botanic Gardens, Kew (RBGK)	0.0162714	9
11	Missouri Botanical Garden, (MO)	0.0140097	8
12	National Botanical Garden of Belgium (NBGB)	0.0120773	6
13	Institute of Botany, Slovakian Academy of Sciences (IBSAS)	0.0086023	5
14	Museo Nacional de Ciencias Naturales (CSIC-MNCN)	0.0073168	10
15	Centraalbureau voor Schimmelcultures (CBS)	0.0070954	4
16	Botanic Garden and Botanical Museum, Berlin (FUBGBM)	0.0067633	4
17	National Natural History Museum Naturalis (NHM)	0.0067072	10
18	Institute of Botany, Poland (IBPAN)	0.0048611	6
19	Komarov Botanical Institute of Russian Academy of Sciences (BINRAS)	0.0048309	5
20	Comenius University, Bratislava (CUB)	0.0032911	4
21	Royal Museum for Central Africa, Tervuren (RMCA)	0.0032005	5
22	Museum für Naturkunde	0.0012077	6
23	University of Amsterdam- Zoological Museum Adam	0.0010064	8
24	Institut National de la Recherche Agronomique (INRA)	0.0006039	4
25	Museum and Institute of Zoology, Poland (MIZPAN)	0	5
	TOTAL		222

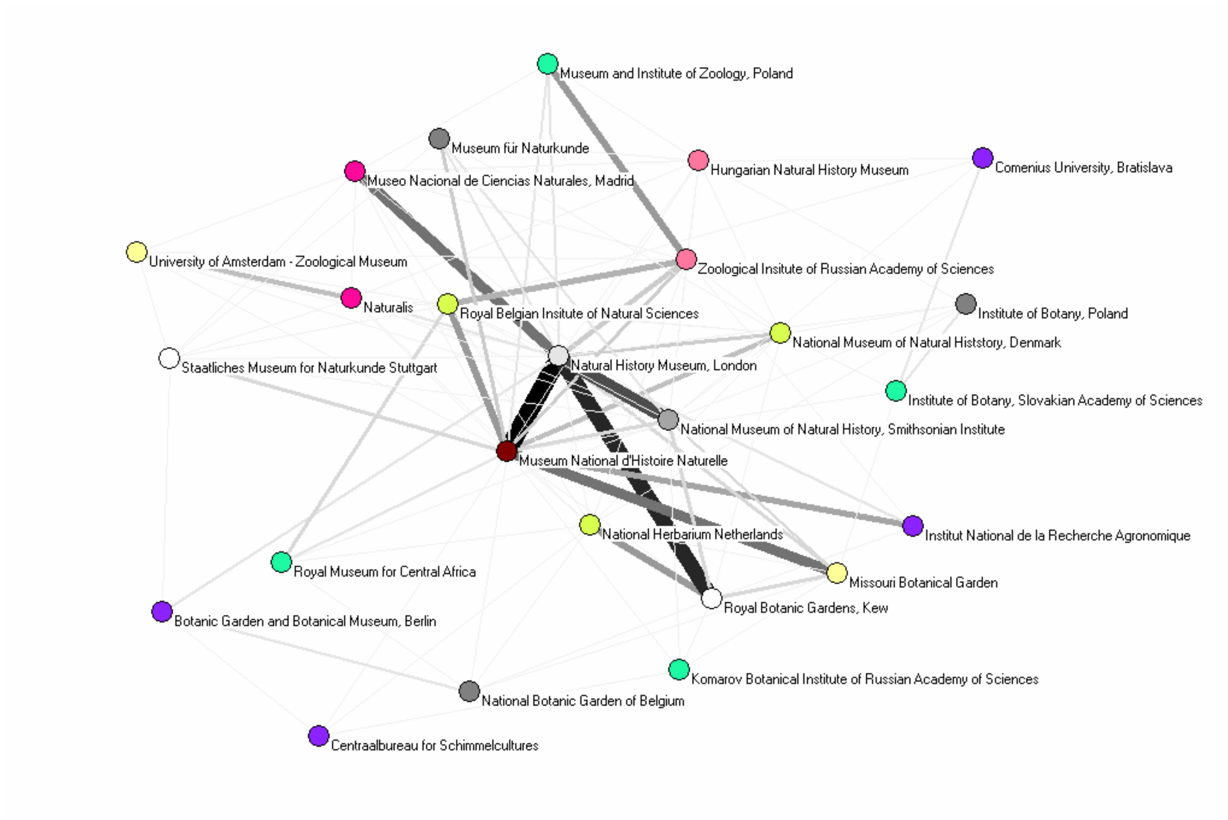


Figure 2. Degree centrality. Co-authorships in EDIT network 2005-2008. All the nodes with the same number of connections to other nodes are all coloured the same and have the same partition value (in network 25 actors, 970 co-authorships, based on ISI-WoS)

An exception might be the position of Naturalis who has 10 ties (co-author relations) and a betweenness centrality of 0.006707, less than the Centraalbureau for Schimmelcultures who has only 4 co-authorship relations in the network but who takes a more strategic position as we can learn from the betweenness centrality table. Actors with a high betweenness centrality are called (knowledge) "brokers" or "gate-keepers": because of their position they keep great influence over what flows and what does not.

Table 2. Frequency distribution of degree centrality EDIT co-author network 2005-2008 (based on ISI-WoS papers)

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
8	4	16	4	16	Centraalbureau for Schimmelcultures
10	4	16	8	32	Royal Museum for Central Africa
12	3	12	11	44	National Botanic Garden of Belgium
16	2	8	13	52	University of Amsterdam - Zoological Museum
18	2	8	15	60	Royal Botanic Gardens, Kew
20	2	8	17	68	Museo Nacional de Ciencias Naturales, Madrid
22	2	8	19	76	Hungarian Natural History Museum
24	3	12	22	88	National Museum of Natural History, Denmark
34	1	4	23	92	National Museum of Natural History, Smithsonian Institute
36	1	4	24	96	Natural History Museum, London
42	1	4	25	100	Museum National d'Histoire Naturelle
Sum	25	100			

The second centrality indicator that we discuss is the *degree centrality* measure. The degree centrality measures the most active actor in the graph, they have the most ties to other actors. Degree is the number of direct connections an actor has.

Table 3. EDIT institutions, publications and EDIT co-authorships between 2005-2008 (based on ISI-WoS papers)

Contract number	Institution name	Number of papers	Number of co-authorships in EDIT network	EDIT co-authored papers as % of total number	Number of pairs (MNHN published with 21 institutions)
2	Muséum National d'Histoire Naturelle (MNHN)	2071	153	7	21
3	Zoological Museum, National Museum of Natural History, Denmark (UKBH-NHMD)	469	36	8	12
4	Museo Nacional de Ciencias Naturales (CSIC-MNCN)	831	39	5	10
5	University of Amsterdam- Zoological Museum Adam	200	19	10	8
6	National Herbarium Netherlands (NHN)	256	35	14	12
7	National Natural History Museum Naturalis (NHM)	237	28	12	10
8	Centraalbureau voor Schimmelcultures (CBS)	358	8	2	4
9	Botanic Garden and Botanical Museum, Berlin (FUBGBM)	64	10	16	4
10	Natural History Museum, London (NHML)	2202	174	8	18
11	Royal Botanic Gardens, Kew (RBGK)	581	70	12	9
12	Staatliches Museum for Naturkunde Stuttgart (SMNS)	104	18	17	9
13	Royal Belgian Institute of natural Sciences, Brussels (RBINS)	461	48	10	12
14	Royal Museum for Central Africa, Tervuren (RMCA)	182	16	9	5
15	National Botanic Garden of Belgium (NBGB)	99	15	15	6
16	Museum and Institute of Zoology, Poland (MIZPAN)	162	28	17	5
17	Institute of Botany, Poland (IBPAN)	148	11	7	6
18	Hungarian Natural History Museum (HNHM)	235	17	7	11
19	Comenius University, Bratislava (CUB)	190	7	4	4
20	Institute of Botany, Slovakian Academy of Sciences (IBSAS)	154	11	7	5
21	Institut National de la Recherche Agronomique (INRA)	261	20	8	4
22	Society for management of European biodiversity data (SMEB)	0	0	0	0
23	Species 2000	0	0	0	0
24	Komarov Botanical Institute of Russian Academy of Sciences (BINRAS)	146	7	5	5
25	Zoological Institute of Russian Academy of Sciences (ZINRAS)	474	56	12	11
26	Missouri Botanical Garden, (MO)	384	51	13	8
27	National Museum of Natural History, Smithsonian Institute, Washington (UNSM)	1215	69	6	17
28	Museum für Naturkunde (MfN)	296	24	8	6
	TOTAL	**11780**	**970**		222

In our network being a central actor means that an institution has co-authored with many EDIT institutions. The number of co-authorships among EDIT institutions are listed in Table 3. Degree centrality can also be computed as a 2 – dimensional map of co-authorships (symbolised by nodes and ties) among EDIT institutions by using the software tool Pajek (Fig. 2).

Each node colour indicates a different level of "degree centrality". It becomes clear from Table 3 for example that the institution with the highest number of co-authored papers in the network (NHML 174 papers) is not the most central actor. Although the MNHN has less co-authored papers (153) in the period 2005-2008 it has co-authored with a higher number of other EDIT institutions (22 for the MNHN versus 19 for the NHML). The MNHN (dark brown, 22) is most central, NHML second (grey, 19) and the Smithsonian National Museum of Natural History is third (khaki, 17) central in the network. Table 3 gives an overview of the number of co-authorships per institution and should help interpretation Figure 2.

Degree measures do not tell us whether actors with a high degree are clustered or scattered all over the network. Below we applied a k-core partition and a domain partition to investigate if our network has actors with different spread of connectedness.

3.3 K-core partition of the EDIT co-authorship network

A computed map of a networks' k-core partition gives us information about the minimum number of linkages of a *range* of actors in the network. In social network analysis language "partition" is used for the classification or clustering of the actors in the network such that each actor is assigned one class or cluster.

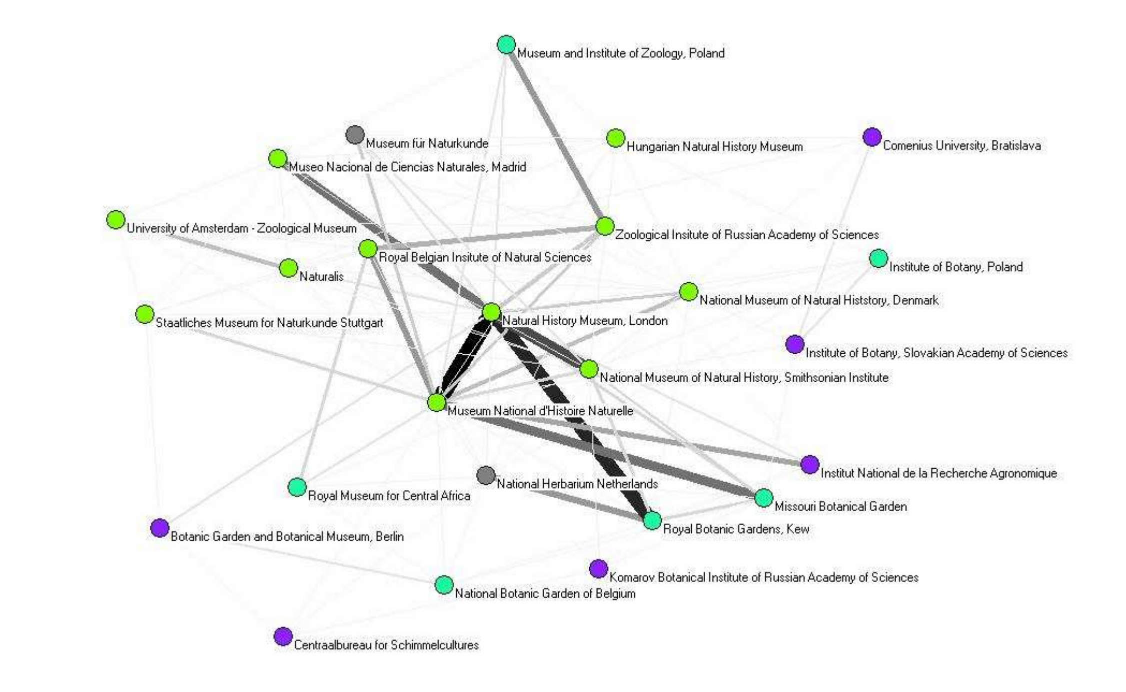


Figure 3. K-core partition. Co-authorships in EDIT network 2005-2008. Different colours indicate 4 different clusters with a shared minimum k-core. (25 actors, 970 co-authorships, 111 co-author pairs, based on ISI-WoS).

The *k* indicates the minimum number of linkages (degree) of each actor within a core. A k-core identifies relatively dense sub-networks, so they help to find cohesive subgroups. However a k-core is not necessarily a cohesive subgroup itself (De Nooy et all, 2007).

Table 4. Core partition, frequency distribution of co-authorships in EDIT network 2005-2008

Cluster	Freq	Freq% cum	Cum freq	Cum freq%	Representative
8	6	240.000	6	240.000	Centraalbureau for Schimmelcultures
10	6	240.000	12	480.000	Royal Botanic Gardens, Kew
12	2	80.000	14	560.000	National Herbarium Netherlands
14	11	440.000	25	1.000.000	Museum National d'Histoire Naturelle
Sum	25	1.000.000			

The frequency distributions table should be read as followed:
Cluster 8 = 6 institutes with a 4 core
Cluster 10= 6 institutes with a 5 core
Cluster 12 = 2 institutes with a 6 core
Cluster 14 – 11 institutes with a 7 core

When using a core partition on the network one is able to identify highly connected clusters, which help to find cohesive subgroups of nodes. This partition allows breaking down the network into smaller mini-networks depending on the maximum number of connections each node has. This differs from the degree partition as a degree partition will result in nodes with only one partition value, not a range as found in a core partition.

3.4 Domain partition of the EDIT co-authorship network

A domain partition of co-authorships allows us to see if there are isolated networks between a small number of institutions within the network. It is not uncommon to see isolated networks

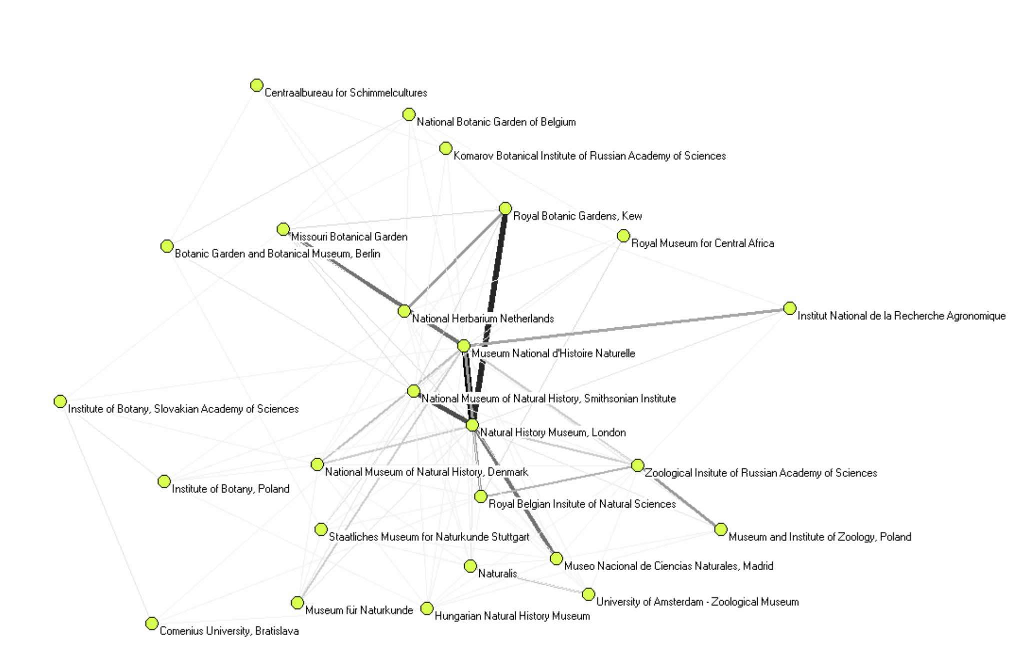


Figure 5. Domain partition. Co-authorships in EDIT network 2005-2008 (25 actors, 970 co-authorships, 111 co-author pairs, based on ISI-WoS**)**

Table 5. Domain partition, frequency distribution of co-authorships in EDIT network 2005-2008

	Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
	24	25	1.000.000	25	1.000.000	Museum National
Sum		25	1.000.000			

between geographically bound institutions or between institutions with similar scientific expertise. However our domain partition shows that this is not the case for the EDIT co-authorship network, there are no isolated co-authorship relations. Like with the degree centrality partition we can conclude the occurrence of a "natural" network of co-authorships

4 Discussion and next steps

We reported in this paper on a EDIT co-author network mapping exercise carried out in the context of a larger study on the mapping of taxonomic networks. The results discussed above provided us with rich information on the co-author relations among EDIT members. The data discussed has not been collected to serve a research quality assessment exercise. Based on our data we do not discuss or compare general research performances of actors, but only their roles in this specific network and the characteristics of the network as a whole. This exercise and its follow up aim to feed into EDITs science policy activities and when applied to a lager actor network - will serve to identify possible partners form outside the taxonomic research community.

We saw that from the figures related to the number of actors, number co-author publications and co-author pairs (Fig. 1, Table 3) that a "natural" co-author network occurred. Each institution had at least 7 EDIT co-authored papers with at least 4 different institutions. The network density measure indicates that the co-author map is moderately dense (Δ=0.37). The betweenness centrality demonstrates that in our network the actors that have many relational ties (pair linkages) have also a high betweenness centrality and therefore fulfil as well a central position also a strategic (gatekeeper) position in the network. Exceptions are Naturalis and to a lesser extent the Institute of Botany, Poland, who occupy a relatively low betweenness centrality. The degree centrality told us that the number of ties an actor has with different other actors should be used as an additional "actor activity measure" next to the total number of co-authorships an actor has in the network. By applying the degree centrality we filtered out those actors who have a high number of co-authorships but only with few actors. The Royal Botanic Gardens, Kew for example has published 70 EDIT co-authored papers with only 9 different EDIT institutions (from the 70 publications 15 are with NHN and 28 with NHML). The k-core partition measurement was applied to examine the occurrence of ranges of highly connected sub-clusters in the network. The domain partition told us that the network did not have isolated sub-networks. This supports our first observation of the occurrence of a "natural" EDIT co-author network. Although all actors have relational ties with several different actors, some are significantly more linked than others.

Hence, the theoretical approach and methodologies described provided us with rich material on the various roles and groupings of the actors in the co-author network. But the data set and the selection which we used mean that we were only able to map a limited range of activities in and relations among EDIT members and in taxonomy in general. As mentioned above, the co-authorships in our analyses also included anthropological and mineralogical papers. In a next mapping phase a more strictly taxonomy focus will be important. Moreover an investigation of the EDIT network in the global co-author network of taxonomic papers may generate valuable, additional information about the roles of actors and the flows of knowledge. We stressed the occurrence of a "natural EDIT co-author" network. A next step would be to know if this observation will persist when we look at a global level of taxonomic co-authorships. Extending the scope of the network analysis should also help us to analyse the various roles of stakeholders in taxonomy from outside the EDIT network and possible also outside the taxonomic research environment.

EDIT and the Rathenau Institute will continue in 2009 to investigate these and other social network questions related to dynamics of the taxonomic knowledge network.

References

C4.1.2BIS. *Stakeholder report 1. Stakeholder engagement in biodiversity and environmental projects* (2007). European Distributed Institute of Taxonomy

Duin, D. (2009). *Scientific titles published by or in name of EDIT institutions*. European Distributed Institute of Taxonomy. URL http://www.e-taxonomy.eu/journals [4/4/09]

Hood, W.W. & Wilson C.S. (2001). The literature of bibliometrics, scientometrics and informetrics. *Scientometrics, 52,* 2, pp. 291-314.

Krebs, V. (2008). *Social Network Analysis. A Brief Introduction*. URL: http://www.orgnet.com/sna.html [4/4/2009]

Newman., M.E.J (2001). *Scientific collaboration networks. I. Network construction and fundamental results. Physical review E, 64*, pp. 016131-016138

Nooy, W. de, Mrvar, A. & Batagelj V. (2007). Exploratory social network analysis with Pajek. New York: Cambridge University Press

Otte, E. & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science, 28*, 6, pp. 441–453.

Thelwell, M. (2007). Bibliometrics to webometrics. *Journal of information sciences, 34,* 4, pp. 1-18.

Wasserman, S. & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press. URL: http://books.google.nl/books?hl=nl&lr=&id=CAm2DpIqRUIC&oi=fnd&pg=PR21&dq=%22social+network+analysis%22+&ots=HtItAb3EKg&sig=lnJvHnlQLHM1B4O35S7wf842Qm0#PPA11,M1 [4/4/09]

Annex 1 – Acronyms

A&HCI	Arts & Humanities Citation Index
AD	Address
CPCI-S	Conference Proceedings Citation Index- Science
CPCI-SSH	Conference Proceedings Citation Index- Social Science & Humanities
EDIT	European Distributed Institute of Taxonomy
IC	Index Chemicus
INRA	Institut National de la Recherche Agronomique
ISI	Institute for Scientific Information
JPA	Joint Periodic Activity
MNHN	Muséum National d'Histoire Naturelle
NHML	Natural History Museum, London
NHN	National Herbarium Netherlands
SCI-EXPANDED	Science Citation Index Expanded
SSCI	Social Sciences Citation Index
WoS	Web of Science

Annex 2 – Key words EDIT institutional addresses

EDIT contract number	Name EDIT member institution	Keywords used for EDIT institutional addresses search in WoS
2	Museum National d'Histoire Naturelle (MNHN)	ad=(Museum Nat* Hist Nat same Paris)
3	Zoological Museum, National Museum of Natural History, Denmark (UKBH-NHMD)	ad=(Bot* Museum same copenhagen) or ad=(Geol* Museum same copenhagen) or ad=(Nat* Hist* Museum same Copenhagen) or ad=(bot* garden same copenhagen) or ad=(Zool* Museum same Copenhagen) or ad=(Museum Nat* same copenhagen) or ad=(Nat* Museum same copenhagen) or ad=(Bot* Museum same Kobenhavn) or ad=(museum bot* same Kobenhavn) or ad=(Geol* Museum same Kobenhavn) or ad=(Nat* Hist* Museum same Kobenhavn) or ad=(bot* garden same Kobenhavn) or ad=(Zool* Museum same Kobenhavn) or ad=(Museum Nat* same Kobenhavn) or ad=(Nat* Museum same Kobenhavn)
4	Museo Nacional de Ciencias Naturales (CSIC-MNCN)	ad=(Nat* Museum same Madrid) or ad=(Museo Nacl Ciencias Nat* same Madrid) or ad=(csic same mncn)
5	University of Amsterdam-Zoological Museum Amsterdam	ad=(ZOOL* MUSEUM same AMSTERDAM) or ad=(MUSEUM zool* same amsterdam)
6	National Herbarium Netherlands (NHN)	ad=(Nat* Herbarium Nederland) or ad=(Nat* Herbarium Netherlands)
7	National Natural History Museum Naturalis (NHM)	ad=(Nat* Museum same Leiden) or ad=naturalis
8	Centraalbureau voor Schimmelcultures (CBS)	ad=(CBS Fungal same Utrecht) or ad=(Schimmelcultures same Utrecht)
9	Botanical Garden and Botanical Museum, Berlin (FUBGBM)	ad=(Bot* Gar* same Berlin) or ad=(Bot* Museum same Berlin)
10	Natural History Museum London (NHML)	ad=(Nat* Hist* Museum same london)
11	Royal Botanical Gardens Kew (RBGK)	ad=(Royal Bot* G* same surrey) or ad=(Royal Bot* G* same Sussex) or ad=(kew same england)
12	Staatliches Museum for Naturkunde Stuttgart (SMNS)	ad=(STAAT* MUSEUM NAT* same Stuttgart)
13	Royal Belgian insitute of natural Sciences, Bruxelles (RBINS)	ad=(Royal Belgi* Inst* Nat* Sci* same belgi*) or ad=(INST* ROYAL SCI* NAT* same belgi*) OR ad=(Museum Nat* same belgi*)
14	Royal Museum for Central Africa, Tervuren (RMCA)	ad=(Royal Museum Central Africa same belgi*) or ad=(Royal Museum Cent* Africa same belgi*) or ad=(Africa Museum same belgi*) or ad=(Musee Royal Afr* Cent* same belgi*)
15	National Botanical Garden of Belgium (NBGB)	ad=(Nat* Bot* G* same belgi*) or ad=(jar* bot* same belgi*)

16	Museum and Institute of Zoology, Poland (MIZPAN)	ad=(Museum Inst* Zool* same Polish* Acad* Sci*) or ad=(Zool* Museum same Polish* Acad* Sci*) or ad=(Dept* Systemat* Zoo* same Polish* Acad* Sci*) or ad=(Inst* Zool* same Polish* Acad* Sci*)
17	Institute of Botany, Poland (IBPAN)	ad=(W* Szafer Inst* Bot* same poland) or ad=(Polish Acad Sci same Inst* Bot*) or ad=(PL-31512)
18	Hungarian Natural History Museum (HNHM)	ad=(Hun* Nat* Hist* Museum)
19	Comenius University, Bratislava (CUB)	ad=(paleo* same Bratislava same Comenius Univ*) or ad=(zool* same Bratislava same Comenius Univ*) or ad=(geol* same Bratislava same Comenius Univ*) or ad=(Bot* same Bratislava same Comenius Univ*)
20	Institute of Botany, Slovakian Academy of Sciences (IBSAS)	ad=(Slova* Acad* Sci same Inst* Bot*)
21	Institut National de la Recherche Agronomique (INRA)	ad=F-34988
22	Society for management of European biodiversity data (SMEB)	XXX
23	Species 2000	XXX
24	Komarov Botanical Institute of Russian Academy of Sciences (BINRAS)	ad=komarov or ad=(Inst* Bot* same Petersburg) or ad=(Bot* Inst* same Petersburg)
25	Zoological Insitute of Russian Academy of Sciences (ZINRAS)	ad=(Inst* Zool* same Petersburg) or ad=(Zool* Inst* same Petersburg)
26	Missouri Botanical Garden, (MO)	ad=(Missouri Bot* Garden)
27	National Museum of Natural History, Smithsonian Institute, Washington (UNSM)	ad=(NMNH same Washington) or ad=(Smith* Inst* same Paleo* same Washington) or ad=(Smith* Inst same Bot* same Washington) or ad=(Nat* Museum Nat* Hist same Zool* same Smithsonian Inst* same Washington) or ad=(Smith* Inst* same Nat* Herbarium same Washington) or ad=(Smith* Inst* same Entomol* same Washington) or ad=(Smith* Inst* same Mineral* same Washington) or ad=(SMITH* INST* same NAT* MUSEUM NAT* HIST* same Washington)
28	Museum für Naturkunde	ad=(Mus* Naturkunde same berlin) or ad=(Nat* Hist* Mus* same berlin) or ad=(Mus* Nat* Hist* same Berlin)