DATA FITNESS FOR USE FOR IAS: PROJECTS AND MAIN TOPICS

Generic and conceptual questions to be considered further:

PROJECT 1a. Annotation of species using categorical status information from other sources (e.g. IUCN Red List, country watchlists) and links to these (potential problem with real time updates from other sources).

*** [task] We need a general statement/recommendation on how GBIF should deal with this. [**MM, SP, QG, GR**]

Give suggestions to GBIF about if and what of these sorts of derived data needs to be considered for IAS? Data could be published at metadata level as a possible solution. GBIF may not want to hold these data/not ne most appropriate for holding these data, nonetheless, to maximize connectivity and functionality for researchers and the IAS community at large, making these links via GBIF is useful. This will end up being a recommendation in the report. Harmonisation has a multiplier effect. So what are these derived attributes that need to be recommended? E.g. GBIF could regularly harvest information on plant pest organisms from the EPPO webservice through a machine-to-machine process. This information could be displayed on GBIF species pages, but could also be a useful filter on GBIF data.
****Flag this as a key issue in the interim report [MM]. And then expand on this for the final report about how it might work and what the resources are.*

WHY: increase data usefulness and data availability for IAS research

2. Should and if so how should GBIF deal with impact and impact-relevant information, Shyama scheme used with UAE, EICAT, other approaches? Are there constituent parts of impact that GBIF could provide data on? Will also end up as a recommendation (probably separate database from GBIF, but nonetheless what data from GBIF would facilitate estimation of impact). Action item: recommendation to be developed [MM, GR]

[[PROJECT: Country guidelines for application (SP, **MM**) [Needs to be done, this is a potential solution] [outside of current task]]]

3. PROJECT: Collaborative project/paper on objective list of priority species [Note: uses existing data to characterize/classify distribution-based classification] [**MM ,**GR, SP, **QG].** *Not do-able in TG timeframe, flag it in the report, good model platform for GBIF to consider, stepwise development of ideas based on thinking developed for report, use case study species examples to illustrate*.

4. PROJECT: Species interactions, e.g. biocontrol ** (Manual of Biocontrol Agents: A World Compendium, USA [LINK]). Should GBIF be a platform for interaction data? We feel GBIF has to deal with it at the occurrence level. How can interactions between species be captured in species observations? For example, host-parasite, plant-pollinator and substrate-epiphyte relationships. Develop a case study? QG

WHY: E.g. a whole suite of insects have invaded the UK because of the introduction of oaks. If the interactions were known, then invasion process and invasion risk would be understood. Fungal pathogens can only be understood ito range dynamics via their host plants. Parasites and pathogens use other species as a ‘habitat’. Species interactions (with a focus on disease, parasites, invasives, pathogens] are key to understanding IAS impacts, also for biocontrol species (and potentially the longer term ecological outcomes of their introduction). *Recommendation to GBIF but also TDWG, flag in interim report.*

WHY:

5. PROJECT: Reviewing Darwin core recommendations, **QG [**1];

Summarise a recommendation for invasion biologist to use **DwC** and recommendations on how to accommodate data not well dealt with by DwC and how to build on DwC for the purpose of alien species data; recommend changes to vocabularies, develop a few case studies [[event data sets are opportunity to capture collocated data (such as habitat etc), also may need controlled vocabulary]]. How does this relate to Humboldt Core? [**QG**.... (gr interested in participating, but maybe focused mostly on opportunity for co-collected/recorded data such as habitat, environmental data, reproductive status, etc. and especially manifest in component below)

Recommendations to add additional new term and vocabulary and vocabulary for the old Darwin Core (DwC) terms should be presented as soon as possible from the group through the GBIF, as it is strategically important for GBIF in the future (VP).

6. PROJECT extension: --- Possibly produce a model "**event core**" for selected taxa that include different ecosystem types (terrestrial, marine, freshwater, aerial) --- to provide exemplars. (a) *To develop a two-pager for us to use to promote to the community and for GBIF, examples*.
(b) Make a table of DwC terms relevant to invasion and add comments next to each about its suitability etc. GR, **QG**

7. PROJECT: Pathways unresolved, matching /mapping exercise needed, especially for marine: see GISP (2003) book , scaleability of pathways and multiple pathways by individual species, also look at freshwater, terrestrial and marine envs using a few case studies, vector activity reticulate, data layered by other providers, usefulness of hierarchical approach lends itself to classification. Are there any vectors missing? [GR, SP, **QG]** Related to Project 1 - desire to have this info, and also how to accommodate such data. *Basic work to do comparison and based on result develop some recommendations.*

1b. PROJECT: Potential operational structures for GBIF versus other partners, recommend some pilot /exploratory projects that are scaleable [what is important for GBIF to fund - cost and timelines] [GR, SP]. Related to 1 but bigger than. *Task is to go to GBIF partners page, identify relevant partners and explain how they should/could interact with GBIF. First skype with Donald, then do…..[ ].*

8. PROJECT: Alien vs Native species record reporting dynamics. Comparison of GBIF occurrence records for model 'populations' (subsets) of native vs non-native species. [Note: This could potentially be combined with Gap analysis (below), but I see this as a separate dimension, that focuses on reporting dynamics and behavior, instead of a Gap-driven analysis.] GR. How to capture range dynamics, including new establishments and eradications [**MM, QG** ….]
Example cases / subset of taxa native and non-native from same taxonomic group and compare data publishing patterns in space and time to understand what the relative dynamic looks like. DS to check if it is feasible. Need to select species? Into report but could also be published (reporting dynamics). *MM&GR to identify criteria. How often entered, downloaded, number of records, …...native versus invasive range. Formulate a technical task for developers.*

9. PROJECT: Gap analysis for GBIF data on Alien Species X Taxa X Ecosystem as possible collaborative project/paper [Consider using Meyer et al. 2016 approach - space, time, taxonomy, stratify by ecosystem and taxonomic group, compare with country checklists, Species status information index (SSI, Meyer, EBV recommendation)] One way would be to ask what the provenace of IAS are in a spatially explicit way compared with country level for example. From this recommendations could arise about how to prioritise work. Could use various 'checklists' as background for queries. Possibly develop proposal for pilot with GBIF. [GR, SP
Just for your information we have used Terrestrial, Freshwater, Marine, Brackish (and combinations of the four) and Host to describe

Environment/System in GRIIS- these status have been applied for the complete 195 country inventories]. Task is to determine how we are going to take this forward.

10. PROJECT- list out existing and discuss and list potential web-services that GBIF could develop to support invasive species research community SP. Think about this from a needs perspective, what and why.

OUTLINES AND DETAILS ON FOUR TOPIC HEADINGS

1. What data are needed / do we need?

(i) Description of topic and motivation of relevance to DFFU_IAS (~50-200 words)

Rationale: *To provide guidance to GBIF on how to improve its existing service to the user community (with a primary focus on the research community that works with biological invasions) and to potentially expand the range of tools and services it provides , it is necessary to identify the range and types of data and variables that are needed to study, report on and manage biological invasions.*

(ii) Details/body (including summarised views from survey)……

Data variables and types identified by Task Group:
The following data/variable priorities necessary to capture biological informations that were identified during the first DFFU_IAS Task Group Meeting:

Data variables and types identified via Survey:

Related initiatives and developments: The 'Essential Variables for Invasion Monitoring and Reporting' Project (invasionevs.org) identified and published three essential variables, as well as a number of supplementary variables required to capture what we need to know about invasions. [[This could serve as a source to populate this part of the report, and or against which to compare our recommendations and the findings of the survey]].

(iii) Roles and relationships across relevant role players, relationship to GBIF, GBIF.org users and data providers

(iv) Recommendations (these will be collated and prioritised at the end -priority and doability)

	Recommendation	**Relevance to users, providers, GBIF**	**From Person/Group (name or TG etc)**
1	Prioritize and perfect taxonomic information (incl. synonyms and common names), for all species on GRIIS and any other authoritative sources of alien and invasive species information	GBIF	MM
2	Prioritize gap filling for species on GRIIS, including clarification of native and introduced ranges	GBIF, Providers	MM
3	Provide appropriate key links from species pages to related/supplementary information sources	GBIF plus...	MM
4	Data providers to find mechanisms to reduce the time lag between observation and mobilization.	Providers	QG
5	Checklists are made available for each country and all taxonomic groups with the native or introduced status of that taxa.	Providers, GBIF	QG
6	All invasive/alien species observations are provided at the highest temporal and spatial resolution possible.	Providers	QG
7	Habitat data (categorical type) associated	GBIF	

	with occurrence records. To include "substrate" type --- inanimate or living. To include elevation/depth.		GR
8	Key environmental data associated with specific occurrence records (eg, in marine systems includes salinity and temperature)	Providers	GR
9	Indication whether occurrence record(s) is on temporary structure (car, plane, ship, train, etc) versus "natural" habitat --- and ultimately whether considered established or "transient" in a region or not --- not just migratory species but also human transported species. The latter is derived by some rule set, but the core data and topic are a "need".		GR
10	Species traits (eg, life history, trophic mode, growth form, mobility, etc) to evaluate patterns spread and function in space and time.		GR
11	Resource (whether GBIF or outside) for transport mechanism(s) by species. This requires refinement of approach in other standards --- which are conflate various different processes/dynamics.		GR
12	Resource for impact data to evaluate space and time-dependent indicators of invasion effects.		GR
13	Native, non-native, and cryptogenic "range" of each species. Should consider novel genotypes in this as well. Requires regular updates to reflect changes in knowledge.		GR

14	Gaps analysis for key target or indicator taxa --- and possibly used for focused data collection campaigns for species or interest or of high indicator value --- stratified by ecosystem type.		GR
15	Filters (and possible visualization) to flag or rank observations that are unexpected or potential outliers.	GBIF	GR
16	Possible alert system for particular species of interest to individuals and organizations by region --- as automated system.	GBIF	GR
	Abundance data: Categorical measure of abundance with observation(s) --- 1, multiple, log scale --- as a starting point. Possibly transition to more extensive later or for target species campaigns or demonstration models.		GR
17	Inclusion of key location types and codes that are relevant to the invasive species issue such as islands, protected areas and other recognised areas of high biodiversity value such as Important Plant areas, Important Bird Areas etc.	GBIF	SP
18	Flagging of provenance of species -inclusion of native range/ alien range	GBIF	SP
19	Identification of the pathways and Invasiveness of introduced species,	USER	VP
20	To minimize the introduction of invasive alien species in the future	USER	VP
21	Inclusion of annotation of typical use and 'reason' for introduction eg -Aquaculture, Mariculture, Forestry, Biofuel production	GBIF	SP

22	Dynamic updates and prioritization of data for new species added to GRIIS	GBIF	MM
23	Determination of the taxonomic groups of introduced species;	USER	VP
24	Absence data, inferred absence data (confidence of absence) and associated date		MM
25	Links to Risk assessments and key sources of IAS data/ EICAT rank	GBIF	SP
26	Can any abundance information of available be included from the Verbatim info?	GBIF	SP
27	Dates of 1st introduction, time series information at and across localities ☐		MM

NOTES

GR and MM Core for data structure, what data exactly would be useful to enhance use GBIF for occurence records. Especially to understand spatial and temporal assessment, and policy factors, such as impact. Big picture, understanding.

SP What data types are needed?

VP What type of spatial and temporal assessment, such as pathways are needed?

QG Why this data are needed, product types? Part of survey: risk assessment (incl. Species), action plans, horizon scanning, management and rapid responce.

GR Identification of user groups

VP ...and recommendations to those.

DS identify audiences, give numbers on how big, where they are

MM can be done with WoS

GR researchers, practicioners... can be very many
MM literature mining

How can this be delivered? Are solutions avalable? What realistically can be done? What is the best way to do?

2. Data structure

(i) Description of topic and motivation of relevance to DFFU_IAS (~100-200 words)
The data structure of GBIF is based upon the Darwin Core (DwC) standard managed by the Biodiversity Information Standards organization (TDWG). DwC has no specific fields specifically for data related to alien and invasive species. Certain data elements are required to inform on the presence of alien species and how they became introduced to new locations. Even if such data fields are made available in DwC it is still necessary that they are adopted by data providers and that GBIF facilitates their use by providing access to them.

(ii) Details/body (including summarised views from survey)……

(iii) Roles and relationships across relevant role players, relationship to GBIF, GBIF.org users and data providers

(iv) Recommendations (these will be collated and prioritised at the end -priority and doability)

	Recommendation	Relevance to users, providers, GBIF	From Person/Group (name or TG etc)
1	To allow the recommendation of particular controlled vocabularies for Darwin Core fields establishmentMeans and occuranceStatus	GBIF providers	QG
2	To add a new field to Darwin Core to express the nativeness of a taxa in	GBIF providers	QG

	checklists (i.e. origin).		
3	Recommended database structure consist of 4 blocks (georeference, taxonomy,alien, references) GEOREFERENCE BLOCK occurenceID – numerical field (Long integer); Continent – text field (String 40); Country - text field (String 40); countryCode - text field (short String 10); stateProvince – text field (String 40); county - text field (String 40); municipality - text field (String 50); locality - text field (String 100); TAXONOMY BLOCK Kingdom -– text field (String 40); Phylum - text field (String 40); Class- text field (String 40); Order- text field (String 40); Family- text field (String 40); Species - text field (String 40); scientificName - text field (String 100); Common name (vernacularName) - text field (String 50); ALIEN BLOCK Year of introduction – Dat; Natural distribution area (Origin) - text field (String 250); Type_of_introduction- text field (String 250) Pathway_of_introduction - text field (String 50) Donor area - text field (String 250) Habitat- text field (String 250) Status - text field (String 25) Frequency - text field (String 25) Invasiveness - text field (String 25) Impact - text field (String 25) References – BLOB (string)	GBIF	VP

	Revise standard vocubularies/fields for vector (pathways) to make this more useful.		GR
	Review and establish standard vocabularies/fields for traits.		GR

NOTES

a. Ideal structure

Four blocks:

1. georeference data - DwC
    a. Occurrence ID, number long integer
    b. Continent
    c. CountryCode
    d. State, province
    e. … details by e-mail from VP pasted to the end of this doc [LINK]
2. taxonomy block - DwC
3. alien block
4. source and evidence
    b. Capture opportunity

3. What GBIF should to do with the data?

(i) Description of topic and motivation of relevance to DFFU_IAS (~100-200 words)
Data providers provide a diverse array of data from different kinds of survey and from a wide variety of organisms, habitats and places. Likewise users of biodiversity data are equally varied in their need for data. GBIF, as a mediator between these two groups plays an important role in facilitating the flow of data between these two groups. This facilitation

(ii) Details/body (including summarised views from survey)……

(iii) Roles and relationships across relevant role players, relationship to GBIF, GBIF.org users and data providers

(iv) Recommendations (these will be collated and prioritised at the end -priority and doability)

	Recommendation	Relevance to users, providers, GBIF	From Person/Group (name or TG etc)
1	Express accumulation of records for species in GRIIS*, e.g. temporal trends in records accumulated per species and per country, trends in number of species and number of 'invasive' species listed on GRIIS *Wherever GRIIS is mentioned other key sources of known alien and invasive species could be added, but GRIIS should be the priority	GBIF	MM

2	Summary statistics (and maps where relevant) for GRIIS species, e.g. number of species, number per taxon, geographic region, country	GBIF	MM
3	Global and country reports for alien and invasive species	GBIF	MM
4	Query functionality to identify species present in trading partners of country A that are not present in it	GBIF	SP
5	Functionality to receive and export alien and invasive species data with relevant databases such as World Database of Protected areas, World Register of Introduced Marine Species	GBIF	SP
	Enable generation of alien and invasive species checklists for various governance scales (e.g. countries, states, protected areas)	GBIF	MM
	User friendly and simple access that does not require IT experts	GBIF	SP
	Enable visualisation of dynamic range (record) expansion maps	GBIF	MM
	Review data resources available for key information types identified above (#1 and 2) to either export or development analytical tools to combine data for analyses and visualization.		GR
	Use GBIF records to provide automated alerts to other groups and also tools to compare (project) new records against known ranges (and regional checklists) to identify new “outbreaks” for validation. This is one type of “filter” that could be used to		GR

	detect outliers. This can include citizen science groups, resource managers, etc. Could also facilitate targetted “detection campaigns”.		
	Dynamic visualization and update of non-native species trends and “gaps”, by region, taxa, and groups. Would be interesting to do this by comparison with “native” species.		GR

NOTES

What data uses are most strongly aligned with GBIF and IAS community. E.g. circulating the gaps info. Creating simple indicators. Niche modelling, external use. Add tools or filters for DM. Example of application.

a. Accessibility = how easily I can find / extractability = how can I extract
b. Quality
    i. Citizen sci DQ - get is all, but flag non-research grade
c. Checking filters
d. Automatic alerts for species and alerts,
e. Generating metrics and stats (potentially also indicators)
f. Status and trends outputs for specific audiences
g. IAS add-on for data publishers

GR Recommendation candidate - show checklist sources associated with names. Use of keys and depth of knowledge. DQ flag for DQ e.g. mismatches for the country lists.
DS Fix the feedback system is a big flagin GBIF, recent post by Rod Page

4. Participation by user and producer community/ Encourage data publication through credit/incentives

(i) Description of topic and motivation of relevance to DFFU_IAS (~100-200 words)
Potential users of GBIF mediated data find that GBIF does not contain all the data they need. Sometimes data elements are not collected, but often they are just now provided to GBIF for a range of reasons. One wayu to motivate data mobilization is to ensure data providers are credited for the data they provide. Therefore, there should be simple mechanisms for users to cite data in meaningful ways that can be used by providers to demonstrate their usefulness and impact. GBIF itself can mediate this process with its use of DOIs and supplying metrics of data usage.

(ii) Details/body (including summarised views from survey)……

(iii) Roles and relationships across relevant role players, relationship to GBIF, GBIF.org users and data providers

(iv) Recommendations (these will be collated and prioritised at the end -priority and doability)

	Recommendation	Relevance to users, providers, GBIF	From Person/Group (name or TG etc)
1	Profile some (which?) priority species and ask data provider community to prioritize (speed up) delivery of data on these species	GBIF plus?	MM
2	GBIF to provide usage statistics at a data	GBIF	QG

	provider level. This will encourage data providers to monitor their data usage.		
5	Data users to respect persistent identifiers on data and wherever possible cite the sources of the data they use	Users	QG
	Identify “indicator” species for data collection campaigns with providers (eg, citizen science, resource managers, scientists, etc). Establish repeated campaigns perhaps over time, with pulsed frequencies.	Providers	GR
			GR

NOTES

Reproducibility and repeatability? QG

Show uses next to source

Having and maintaining unique ID through the products.

SP Primary vs. secondary source citation. DAISIE bad example.

5. Community integration with other IAS related players

(i) Description of topic and motivation of relevance to DFFU_IAS (~100-200 words)
Data on invasive and alien species are held by many different organizations and much of this data is not suitable for incorporation within GBIF. However, invasive species biologist want to be able to combine data from these multiple sources to create new information. These organizations should work together to facilitate interoperability between data types. Furthermore, to ensure interoperability it is necessary that institutions and subject domains collaborate on data standards and controlled vocabularies.

(ii) Details/body (including summarised views from survey)……

(iii) Roles and relationships across relevant role players, relationship to GBIF, GBIF.org users and data providers

(iv) Recommendations (these will be collated and prioritised at the end -priority and doability)

	Recommendation	Relevance to users, providers, GBIF	From Person/Group (name or TG etc)
1	Consider publishers who require archives as source of occurrence records --- as possible service to authors and increase data on AIS.		GR
2	Outline ideal framework for partnerships and data exchanges--- > with an aim to creating functional tools and products, to		GR

	avoid redundancy in increase utility. Show the functional interactions and 'unique' space vs overlap in roles.		
3	Implement Outreach / socialization plan to approach these respective partners and build a functional system.		
4	Even though DwC is an established standard, there is still widespread ignorance of it outside the field of Biodiversity Informatics. GBIF and other key partners could take a more proactive role in promoting, training and facilitating the use of DwC, particularly among ecologists.	GBIF	QG

NOTES

Acronym soup from white board. Specify key external, also non IAS resources (think GenBank, WorldCLIM, trade). Co-occuring, physical, chemical data.
Identifying the users, providers, providers+users that ineteract with GBIF or IAS key players that can complement and be relevant in provideing addition information and feedback to GBIF activities. Who else can do this? Desried flow of information and data to achieve.

GBIF could push emotional arguments for citizen scientist.
DS science ambassadors
SP and MM citizen science for aliens have potential additional bias
GR network approach and reporting warning. Trade relevant taxa to be considered. Dynamic…

Summary of the day: five topic areas, common understaning. Common structure for each topic

Topic
Motivation
Relationship to GBIF
Elaboration
Recommendation(s)
Driven by / user

Prioritisation tabel in the end by importance and doability

Hot topics:

- Data structure, DwC - QP, VP, GR (pathway)
- Capturing the dynamics of species distribution - MM
- Attribution of A and I status, impact - categorical attributes to occurence record.

Two dominant definition in literature - biogeographic and impact (CBD) definitions
Definision of alien and invasive species - CBDs and IUCNs agree.
Both definitions are needed. QG Canadian geese example.
MM and GR - elaborate on attribute. Might recommend terms based on imact.
SP evidence of impact is literature reporting on impact and other evidece of impact
MM Systematic decidion making flowchart leading to impact / not and evidence. Level of uncertainty. Use adopted scheme with a measure of uncertainty and terminology.
QG one can indicate more that one scheme and catogories from this scheme
MM's approach builds on Blackburn 2014
SP keystone, ecosystem engineers?

Collation of recommendations

Once all recommendations are in - collate and rank according to importance, do-ability, who should be involved (role and relevance for GBIF) and how should they contribute (GBIF, other data sources/providers, user etc). Identify problems and think about solutions…., Also flag relevance (priority) to research, policy and/or management. Feasibility analysis for all recommendations

Data analysis of survey results

1. How well do responses align with our own views/recommendations?
2. Add any new recommendations to our list
3. Weight/prioritise recommendations based on survey results
4. Look at data by region and by interest group and tailor recommendations to specific users/regions
5. Use it to identify potential partners in data use and provision
6. Stratify analysis by user group and by country/continent, discipline