Terms of Reference for the GBIF LSID-GUID Task Group (LGTG)

Background

GBIF has identified the provision of stable and resolvable identifiers for biodiversity objects as one of the central challenges to developing a global bioinformatics infrastructure. One of the stated goals in the GBIF strategic plans document "*GBIF Plans 2007 – 2011 from prototype towards full operation*" is to consolidate the underlying enabling infrastructure and standardisation for global connectivity of biodiversity data and information through an activity to "develop a system of globally unique identifiers and encourage their use throughout biodiversity informatics". Milestones include using TDWG standards to "allow all data objects to be identified using standard resolvable globally unique identifiers" and provision of a GBIF web service and user interface to allow users "to locate and view any data object with a standard globally unique identifier".

Several other goals in the GBIF Work Programme 2009-2010 depend directly or indirectly on the deployment of GUIDs/LSIDs. These include, under the DIGIT work area, the need to promote development and implementation of a "technical, social and policy framework" for publication of primary biodiversity data as a "research data management chain", and to enable credits/attributions to data contributors/providers, and under the ECAT work area, the need to develop the core components of the Global Names Architecture (GNA). In addition, several components of the GBIF informatics architecture including the Global Biodiversity Resources Discovery System (GBRDS), the Integrated Publishing Toolkit (IPT) and the Nodes Portal Toolkit (NPT) rely on use of GUIDs.

Now, because of the perceived urgency and uncertainty around the future of LSIDs, and more generally, the social/institutional challenges of providing stable and persistent GUIDs, GBIF is prioritising the convening of an LSID-GUID Task Group to explore the issues and offer recommendations on the way forward, with particular reference to the GBIF network, that will enable GBIF to provide architecture leadership and best practices for implementation.

Objective

The principal objective of the group is to provide recommendations and guidelines on deployment of LSIDs and other GUIDs on the GBIF network with particular reference to the potential role of GBIF as a stable, long term provider of GUID resolution services. The task group should address issues for both immediate, near term (August to December 2009) and longer term implementation.

Terms of Reference

The task group will address the following issues:

1. Review the main GUID technologies for biodiversity informatics (in particular, LSIDs, DOIs/Handles, PURLS), their main use cases, their implementation/deployment, and consequences of adoption (social and technical)
2. Review the plans for a decentralised GBIF informatics architecture to ascertain requirements for GUID technologies; in particular, advise on the role of GUIDs in -
    1. a citation / impact factor tracking mechanism
    2. management and disambiguation of taxon names
3. Evaluate a role for GBIF in provision of LSID hosted services including proxy services; issues include -
    1. centrally managed vs distributed service; what are the benefits of a distributed service over a community-funded central service?
    2. commitment to 24/7 operation (persistence) ensuring that GUIDs are always resolvable
    3. monitoring service quality (in a distributed system) and providing warnings to providers when problems are detected
    4. ensuring back up service for LSID technologies as protocols and resolution issues change/evolve over time
4. Identify the data models/vocabularies for use in metadata returned on GUID resolution; issues include -
    1. the need for standard vocabularies/ontologies; a role for TDWG ontologies

2. associating machine readable metadata with identifiers
3. recommended schemas for all digital objects being assigned LSIDs/GUIDs for use within the GBIF network, e.g., citation, dataset description, network agent (e.g. institutions, organisations)
4. review of existing schemas to ensure that reuse is possible (*Occurrence Record* identified to a *Taxon Concept* etc.)
5. Provide applicability statements for LSIDs and other GUID technologies
6. Propose a business model for adopting LSIDs – what associated infrastructure do we need that will bring added value and reward those who breach the technical barrier to set up an LSID system?
7. Identify and suggest solutions to social issues relating to adoption and persistence of GUIDs, including -
   1. dealing with different levels of institutional I.T. support for setting up an LSID system
   2. use of proxy services to avoid the need for working with DNS records
   3. educating data publishers and biologists about use of GUIDs
   4. enforcing the social responsibility concerning commitment to persistence when issuing GUIDs
8. Identify solutions for integrating GUIDs (e.g., LSIDs, Handles, DOIs) with the Semantic Web and the Linked Data model, e.g., through HTTP resolvers and proxies
9. Identify solutions for promoting the rapid distribution and uptake of GUIDs, including linking of GUIDs to each other

Outputs

The main deliverable will be a set of practical guidelines and recommendations around the issues defined under the Terms of Reference and presented in the form of a report to the Executive Secretary of GBIF on or before September 15, 2009.

In addition, the Task Group will prepare an applicability statement for each GUID technology recommended for use with biodiversity data, and a short statement (by late May) for the eBiosphere conference in order to highlight the the critical role of GUIDs in biodiversity informatics over the coming decade, emphasising -

1. GUIDs as part of a stable, long-term cyberinfrastructure in support of biodiversity research
2. the level of commitment required
3. the need for solutions to stop/go funding while building a stable biodiversity informatics infrastructure

Timeline of Activities

Timeline	Activity
April 2009	Call for Nominations to the Task Group
May 2009	Establishment of the Task Group; prepare statement for eBiosphere
Late June 2009	Draft discussion paper on 'LSID/GUID Deployment on the GBIF Network'
Late June – Early July 2009	Distribute draft paper to seek inputs from key stakeholders; comments / inputs from key stakeholders collated
Late July or Early August 2009	Face-to-face meeting of the Task Group to draft first round report
Late August – Early Sept 2009	Solicit GBIF Participant comments on the draft discussion paper 'LSID/GUID Deployment on the GBIF Network';
Early September 2009	Revise report based on inputs
September 15, 2009	Deliver final report

Task Group: Characteristics, Timeline, Modus Operandi

The Task Group should operate from May (once group members are appointed) until mid September 2009 when the final report should be submitted. The Task Group should convene most of its business remotely through mailing-list, wiki, Skype and conference calls. Regular physical meetings are not planned but at least one is foreseen in order to synthesise inputs and formulate the final report.

The Task Group will be coordinated by, and report through, the IDA Programme Officer at the GBIF Secretariat.

Task Group: Composition

The Task Group should consist of a small number of experts with experience of implementing persistent identifiers particularly in a biodiversity informatics context.

Call for Nominations

Any individual with the required expertise and who can commit the time to contribute towards the stated objectives of the Task Group is invited to apply for membership. If you are associated with a GBIF Participant you may consider contacting your Node or Head of Delegation to support your application. Members of the Task Group will be required to devote considerable time for frequent discussions, conference calls, and timely inputs towards various activities of this exercise. No remuneration for time spent is envisaged. However, logistical/travel costs of meeting/s will be covered.
Please send your application including a brief CV outlining relevant experience by 10 May 2009 to Éamonn Ó Tuama, GBIF Programme Officer for IDA (Inventory, Discovery, Access) [Email: eotuama@gbif.org].