TECHNICAL ANNEX

1. S&T EXCELLENCE

1.1. Challenge

1.1.1. Description of the Challenge (Main Aim)

Biodiversity is an essential part of our lives; sustainability, well-being and happiness depend on our knowledge, appreciation and taking the right decisions regarding Biodiversity. This has been perceived in many ways and as a result a large number of national and international conservation and management programmes are being launched to assess ecological integrity and help establishing sustainable ecological conditions. These initiatives are central instruments for the implementation of international commitments and legislations, such as the Convention on Biological Diversity and its associated Aichi biodiversity targets (http://www.cbd.int/sp/targets/), or the European Biodiversity Strategy for 2020 (COM(2011)0244). (Gärdenfors & al. 2014). In this context it is only logical that seeking integration and interoperability in biodiversity information should be a priority for Europe. This has been reflected in the number of EU projects carried out in recent years around biodiversity information and biodiversity informatics (e.g. ENVRI, BioVel, EUBON, VIBRANT, pro- iBiosphere, PESI, Creative- B, etc.)

This is also the case in COST, with nearly 30 actions built with biodiversity in focus, many of which aspire to "integration", "linking" and "harmonisation". As pertinent as these efforts are, they are only providing limited progress, as the integration they build is restricted to (e.g.) marine environment, or modeling, or pollinators, or invasive species or other sectors of biodiversity. Almost all European initiatives, particularly those funded under EU Research Programmes, are time restricted so there is a need to provide a repository where data can be stored in the long term, and made available for reuse.These initiatives also often neglect to recognize that many European countries have already put in place a number of biodiversity information national nodes, as part of the Global Biodiversity Information Facility (GBIF, the largest biodiversity data network in the world, comprising 54 country members, of which 21 are European: www.gbif.org). The GBIF, as a distributed research infrastructure provides services such as long-term data storage and access to data initiatives and funding bodies in an open and standardised way.

This action aims to provide an "integration of integrations" for biodiversity information in Europe; it will focus on content, processes and capacities rather than tools. Thus, it will provide a solid pan European approach to how biodiversity information is captured, documented (metadata), archived, processed (validation), made available (standards, LOD) and used. In order to reach these goals, this action will cooperate with short-term initiatives which have components on biodiversity information – such as other COST actions, and EU projects – and with long-term initiatives and mechanisms such as the EEA (European Environment Agency) and LifeWatch (www.lifewatch.eu). This action will gather together a core team of partners constituted by national GBIF Nodes. These nodes already comprise a wide range of capacities, approaches and levels of development, with much potential for complementarity and leverage. As coordinators of national biodiversity information networks GBIF Nodes are not only in direct contact with those projects and entities that generate and administer biodiversity information, but also with those who use biodiversity information for research and decision making.

This action will result in more coordinated and focused developments for understanding and managing biodiversity in Europe, and for understanding and improving the way biodiversity information is used, to bring added value to existing data holdings and initiatives, a better return-on-investment in this field, and improved linkages with biodiversity information networks in other regions and at a global scale.

1.1.2. Relevance and timeliness

This COST proposal is strategic in a number of ways:

In recent times we have observed an explosion of tools and services (developed by EU projects or others) relying on biodiversity "available data". However, these "available data", are far from being readily usable: lack of traceability, lack of a coherent semantic framework, lack of means to assess data quality, etc. Filtering and disregarding data as well as methodological assumptions are made routinely in order to obtain results, which are limited by data shortcomings.

The ultimate goal of this COST Action is to significantly alleviate this bottleneck. Some activities such as more data in digital form are beyond the scope of the action, but even in this area a substantial impact can be attained by spreading and adopting the more advanced tools and procedures among partners. Other aspects can be tackled directly by the COST Action, as those that can be labeled as "intelligent openness" as defined by the United Kingdom's Royal Society ("Science as an Open Enterprise" p. 7. 2012): "Data must be accessible and readily located; they must be intelligible to those who wish to scrutinise them; data must be accessible so that judgments can be made about their reliability and the competence of those who created them; and they must be usable by others".

Improved access to usable data is a requisite for better science and management. In practice this means that the answers we get by analysing the vast quantities of data available through the GBIF network and other sources will be more accurate and decisions taken more valid. This will also expand the community ability to exploit these data, and thus improve return of investment at several levels (in digitalisation, in identifying and selecting suitable data for specific purposes, etc.). A harmonised, well-connected network of GBIF nodes will bring data publications and long-term maintenance to research projects, (national, EU, or others), monitoring programs, other initiatives (Citizen science, etc.). These go beyond archiving, since data are accessible, retained under common standards, associated with good metadata, traceable, and citable.

This COST action will also leverage investment countries are already making in their GBIF Nodes by achieving true integration of biodiversity data, across countries, and across disciplines and areas of interest (e.g., pollinators, invasive species, crop wild relatives, migratory routes, climate change, indicators).

This COST proposal is timely. A number of initiatives pursuing the identification of trends or providing answer on which to base unrelated decisions have emerged in recent times. Many, if not all, require representative, current and scientifically validated biodiversity data. With few exceptions, biodiversity data acquisition, provenance and quality has not been the focus of these initiatives. Gathering, harmonising and publishing data is costly (we are far from having automatic sensor for recording biological species, and human observation and expert knowledge is still needed). Without good usable data, any service or analysis built on top is at the risk of becoming just an academic exercise. This action will allow services and tools to be put in place to provide better results.

This COST action will have an impact in current and past EC projects that deal with biodiversity (better and more data will produce better answers, provide a dynamic usable repository for data gathered); will support infrastructures such as Lifewatch, will contribute effectively to the reporting and information gathering processes of the EEA, and serve to better respond to pan-European monitoring schemes and reporting (some of these legally binding: i.e. IPBES, CBD, CITES, RAMSAR). Being able to express GBIF mediated data under the INSPIRE (Infrastructure for Spatial Information in the European Community, DIRECTIVE 2007/2/EC) specifications is relevant here.

The technological moment is ripe to be exploited through capacity transfer and implementation of emerging technologies (global identifiers such as DOIs or LSIDs, semantic interoperability, standard licenses, cloud storage and computing, mobile devices, etc.), and these and other technologies are being explored and used by one or a few projects, agencies or GBIF national nodes, but none is using all or most of them. The potential for expansion and leverage through training and dissemination is tremendous.

In a wider context, having the actors and the processes to acquire, maintain and make "intelligently open" and harmonised biodiversity data – coming from virtually every relevant source via the national GBIF nodes and other data suppliers – is going to be a pillar in a world advancing towards more data intensive science, more linked data, more big data.

1.2. Objectives

1.2.1. Research Coordination Objectives

1. To improve data reliability and usability by expanding the use of persistent identifiers, and other semantic web components; and collectively improve the tools to handle them.
2. To harmonise and standardise tools and practices for data capture, quality control, Web services, persistent identifiers, and other semantic web components; provide input to standardisation bodies such as the “Biodiversity Information Standards” (This organization – also known as the ”Taxonomic Databases Working Group” or TDWG – is the global reference in development of standards for the exchange of biological/biodiversity data; www.tdwg.org).
3. To improve data quality by sharing quality control tools and practices, training new staff on the use of these tools and practices, and joining forces to improve, update and maintain the tools;
4. To increase collection data mobilisation by sharing techniques and practices for massive digitisation of collections and specimen labels data capture, and joining forces to improve data capture methods and tools, combining OCR and citizen science data curation;
5. To increase impact of national efforts through tools sharing, co-development, and good practice relating to dissemination;
6. To enhance sustainability of tools through open source libraries, enabling collective long term update, improvement and maintenance of common tools.

1.2.2. Capacity-building Objectives

1. Bridging projects and Actions with common objectives and harmonised approaches in data acquisition, management and publication;
2. To implement a training resource repository compiling materials developed by projects and Action's partners on key aspects of biodiversity information management and use for capacity building and transfer, dissemination and outreach;
3. To fast-track countries (Europe, NNC, IPC) aiming to build or to improve their biodiversity data networks by bringing them into the Action and exposing them to the more advanced and successful procedures and tools via training schools, STSM, and other Action activities;
4. To train young scientists and others in the use of already mobilised data; in collaboration with universities when feasible.
5. To foster experience, know-how and knowledge exchange through the WG workshops;
6. To expand the capacity in quality control, via document dissemination and training sessions throughout the network;
7. To increase the critical mass of capable partners in biodiversity data management and network coordination by pairing experienced partners with "emerging" partners

1.3. Progress beyond the state-of-the-art and Innovation Potential

1.3.1. Description of the state-of-the-art

At the moment (cf. GBIF Annual report 2014) there are 21 National GBIF Nodes in Europe out of 47 worldwide. European countries are among the leading data publishers around the world, contributing c. 224M records out of 550M records currently available online. All these records, made available through the GBIF Infrastructure (central portal, national nodes, participant collections and projects), are under a common format, centrally indexed, queryable and suitable for analysis (using APIs and workflow tools), and enriched with metadata.

These figures might look impressive, and ten years ago we could only dream of having such information at the tip of our fingers. However, this is not the ultimate picture of biodiversity in Europe:

- Being Europe, overall a well-known territory, regarding data online, there are still important gaps, temporal, geographical or taxonomic, specially at detailed scales.
- Different data categories (point data, area data, multimedia, organism level, molecular data, etc.) are minimally integrated; integration efforts are limited and not interconnected.
- Biodiversity data available online is not even representative of our knowledge on biodiversity; there are vast amounts of data (in collections, in the scientific literature, etc.) unavailable, non standardised, and isolated but with large potential to be applied for research and management in biodiversity, environment and global change.
- A number of initiatives and projects (including a number of COST actions) working in data integration, resulting in "fragmented integration". Some themes found in current COST Actions related to biodiversity are:
- In the past decade a number of successful EC projects were carried out in the area of "Biodiversity informatics", the following non-exhaustive list illustrates this:

ENVRI	Tools
BioVel	Workflows
EUBON	Building the European Biodiversity Observation Network: tools, integration, analysis
VIBRANT	Virtual research environment
pro-iBiosphere	Coordination and policy development
4D4LIFE	A coherent classification and species checklist of the world's species
PESI	Pan-European checklist
EDIT	European Distributed Institute of Taxonomy: bring together the leading taxonomic institutions in Europe
OpenUP!	Connecting Natural History data and multimedia object to Europeana
Creative- B	Coordination of Research e-Infrastructures Activities Toward an International Environment for Biodiversity
EBONE	European contribution on terrestrial monitoring to GEO BON

- Most of these projects aimed to produce services or tools, or coordination, on the basis of data available; in most cases again, "data available" referred to GBIF-mediated data or LTER (Long Term Ecological Research) data; however in practice, it is almost exclusively GBIF as LTER data are not standardized. In all cases these are limited in time, resulting in non-curated datasets (or tools) which access, quality and relevance degrade over time.
- Despite being the best and largest source of biodiversity data online, records currently available via GBIF have some important shortcomings: quality is heterogeneous and difficult to assess, semantic identifiers are not stable; searches are made by names and not by concepts; annotations are not possible; many datasets lack standardised use licenses.

1.3.2. Progress beyond the state-of-the-art

Improved usability of data already in digital form and available online.

Quality. Coordinated efforts on data quality/validation, data cleaning, fitness-for-use indicators and data annotation mechanisms will greatly increase the value of already mobilised data.

Semantic framework. Use of persistent identifiers, controlled vocabularies, Linked Open Data approaches, data output as RDF, for example, from the start of the "data life cycle" would make data much more easy to use, provide better traceability, be easier to combine and to aggregate knowledge (e.g. annotations).

Licenses. Lack of clarity about what uses are permitted limits data usability; expanding the use of standardized licenses will increase data usability.

Standardization and harmonization (e.g. LTER, Consortium for the Barcode of Life, & GBIF). Aligning and bringing together data formats and concepts of the largest biodiversity data avenues will result in richer and more meaningful data applicable to a wider range of issues.

Improving usability of biodiversity data within the frameworks of the EEA and the INSPIRE directive. Making geo-referenced biodiversity data coming from research and citizen science compliant with the INSPIRE directive, working for and at the end to bring science, society and administration closer.

Increased data usage. Through a number of coordinated tasks involving standatisation, training in data analysis, visualisation, workflows and other skill related to data use; with special attention younger researchers and professionals, and reaching out for groups where potential biodiversity data usage is larger (administrations, citizen science, private sector).

Increased quantity of data available.

Increased data mobilisation activities. By establishing activities aiming to bring additional partner to the network, in combination with capacity building activities around data capture, the mass of institutions, and initiatives involved in digitalising and publishing biodiversity will increase and in turn the amount of biodiversity data available.

Massive data acquisition. For instance, following the path set by a number of countries in Europe and outside Europe, which have recently undertaken a massive effort to digitise (scan) their collections: making millions of high definition images of scientific specimens available (e.g. *http://www.webdoc-herbier.com/#!91*) with a great opportunity to capture the wealth of information recorded on the specimens labels, by combining the use of adapted OCR systems with a citizen science approach to error correction, drawing on experience already gained in countries like Finland, France, Germany, Australia and the USA.

Citizen science. CS is being recognised and as an element that can complement more formal data acquisition and handling activities. Identifying, disseminating and adopting the best developments as well as working with SC to integrate their results in initiatives such as the Biodiversity Information System for Europe (*http://biodiversity.europa.eu/*) or GBIF wibll make the European biodiversity data landscape not only larger but more current and representative; and thus applicable to relevant areas such as phenology or invasive species.

New data types

Publication. The bulk and the core of biodiversity data online --potentially suitable for analysis-- is "point" data; is presence data. Standards, procedures and access points --such as web portals, application program interfaces (APIs), etc,-- for other types of data (polygons, plot-based, absent data, multimedia) are needed, so better knowledge can be distilled from that information. Some developments exist in this regard, but they need to be refined, and mainstreamed, A network as proposed in this Action is the most suitable approach for achieving these.

Integration. Having different types of biodiversity data sharing a core standard, following Semantic web specifications and available online, will enable seamless integration and reduce the costly and redundant pre-analysis data harmonisation and filtering processes so common now.

Dissemination and adoption of best methods and tools available from all to all.

The Action will be a key instrument in this area by working in:

Initiative bridging. Data harmonisation occurs too much at the end of the "data life cycle" making that work hardly reusable and inefficient. Harmonizing data from the start in an open standardised way is much better. That requires first communication, to lead to coordination (avoiding gaps and duplication) then to cooperation among data providers and users, coming from various context (administrations, research, society). Progress in this area will be in-line with the EU Data management plan and Data Pilot required for all data related H2020.

Facilitation. Identifying shortcomings and solutions among partners, working to leverage the network capacity using the COST networking tools.

Cross-pollination. Good approaches, tools and ideas appear within any context; a culture of open data, open source multiplies the impact and benefit of such developments. By promoting adoption and co-development of the best tools and practices, the Action will provide sustainability to developments, visibility to developers and partners, and increased general efficiency.

1.3.3. Innovation in tackling the challenge

The action aims to produce innovation by promoting, refining and supporting developments in the following areas:
High-throughput Biodiversity Data capture and digitalization methods and techniques.

- Semantic aware technologies for capturing "silent knowledge" from collections and sharing it to enhance data capture (for example RDF models and vocabularies for narrowing the scope in the transcribing process [data capture] based on known locations and time periods
- Citizen science techniques (moderation, qualification, motivation, quality control, etc.) adapted to amateur naturalists communities for transliteration of scanned specimen images
- Processes and devices for the digitization of specimens, adapted to the various kinds of specimens (insects, plants, fungi; individual or collective; 2D or 3D, etc).
- OCR techniques adapted to labels manuscript writing with specialised controlled vocabularies.

Data management, exchange, and publication:

- Semantic aware technologies for interoperability, traceability, error detection and correction, fitness for use evaluation; and related tools
- Best practices and training programmes for using Persistent identifiers
- Trans-discipline/domine multilingual controlled vocabularies

Biodiversity data quality framework:

- A common language for Biodiversity data quality f
- Sustainable annotation systems for biodiversity data published on the web
- Data quality control library (programming code and services) for biodiversity data

Data integration to open biodiversity data to non-specialists, as the GPS did with geographic information so in can be used well beyond its initial intent (e.g. Ecotourism, Species identification, divulgation pPhenological analyses and predictions, Invasive species management, etc.). attainable through:

- Data more integrated (via common standards, Semantic technologies and via APIs).
- Data more in context (metadata, annotations, traceability)

1.4. Added value of networking

1.4.1. In relation to the Challenge

A number of activities, serving the same purposes and using similar techniques, are performed independently by GBIF Nodes, data holders and data users, with too little coordination. These activities cover a diversity of themes and fields: e.g. training, data portals, online maps, various software development for quality control, citizen science, data capture, etc. They will all greatly gain in effectiveness, cost saving, outreach, impact and sustainability, by being conducted within a network with a minimum level of coordination and mutual capacity building, where the various actors can share experience, harmonize their practices, train each other, and join skills and efforts for common tools development, update, improvement and long term maintenance.

1.4.2. In relation to existing efforts at European and/or international level

The following table lists the most relevant long term initiatives in the area of biodiversity data with comments on how the Action can add value in relation to them:

Initiative	*added value of the Action*
EEA	Improve the “workable knowledge” of biodiversity in Europe by contributing more data -- current and historical -- to the EEA reporting procedures. Work with national focal points and policy makers to lower the barriers to use and integrate data coming from academia and citizen science
LifeWatch	Better data, better documented to enable easier integration and more powerful analysis and uses under the LifeWatch infrastructure (services, virtual labs, etc.)
BISE	Increased participation of biodiversity data holders
GBIF	Strengthen collaboration among Nodes on core issues in biodiversity information, expand and leverage overall capacity of the network, improve integration and use of “GBIF data products” beyond GBIF community
IPBES (Intergovernmental Platform on Biodiversity and Ecosystem Services)	Assisting IPBES in data, expertise and capacity building needs
EUDAT (collaborative Pan-European infrastructure providing research data services, training and consultancy)	Provide community-based standards and procedures for data exchange and documentation; foster use of EUDAT services among the biodiversity data communities
EU-BON	Reinforce Eu-BON-GBIF existing collaboration to strengthen GBIF position within GEOSS
CETAF	Advance capacity of natural history museums in Europe by refining and testing CETAF recommendations and good practices; contribute to the implementation of these in a wider community
TDWG	Contribute to TDWG’s objectives, by providing expertise and feedback on developing standards, and expanding their impact through training events
Catalogue of Life	Enhance taxonomic expertoice of the CoL network, provide feedback on taxonomic concepts reconcilaition
CBOL	Improve integration and cross analysis capabilities of molecular and other biodiversity data (occurrences, traits, etc.).
GRBio (Global Registry of Biodiversity Repositories)	Expanding the contribution and use of this initiative
LTER-Europe (a regional network of ILTER, the international Long-Term Ecological Research Network)	Improved interoperability between LTER and GBIF data and metadata standards, procedures and portals; reduced duplication in data curation and archiving
OpenAIRE	Enable OpenAIRE (https://www.openaire.eu) to work out how to organize biodiversity data within Horizon 2020 Open Research Data Pilot
Foster	Co-organize training activities (https://www.fosteropenscience.eu/)
Other COST Actions	Provide overla integration and harmonizations of several domain-specific integration and linkages currently going on in several COST actions (e.g.: Marine biodiversity observatories, alien and invasive species, forests, pollinators, biodiversity and ecosystem modelling)

2. IMPACT

3.

2.1. Expected Impact

2.1.1. Short-term and long-term scientific, technological, and/or socioeconomic impacts

On the short term, the Action will contribute to improve data quality, reliability and traceability, to standardise and improve methods and tools, and to save costs, upgrade skills, and improve practices.

On the long term, it will provide a sound basis for enhancing data mobilisation, for improving data quality, for more relevant and accurate uses of data for science, operation and policy, as well as facilities for collectively improving, updating and maintaining methods and tools, and for outreach and mutual training.

2.2. Measures to Maximise Impact

2.2.1. Plan for involving the most relevant stakeholders

Action's proposers have leading roles in organising the biodiversity data networks In their respective countries, and are often integrated in more general purpose biodiversity platforms. They work closely with data holders and data users, and with the major national programs, agencies and authorities dealing with biodiversity. They are thus in an ideal position for involving all concerned stakeholders and identify the key actors among them.

Furthermore, the workshops organized at the action's start will identify and let emerge the most relevant stakeholders for the various issues tackled by the Action..

2.2.2. Dissemination and/or Exploitation Plan

WG1 to WG3 will maximise outreach and dissemination beyond the GBIF Community and the key European institutions involved. Appropriate intective platforms and Web site sections will be dedicated to dissemination and outreach, and GBIF Secretariat, through their website and portal, their participation in all key biodiversity events and their contacts with the major actors in biodiversity information, will bring an essential contribution.

The present action will organise the feedback from users to data holders and tools developers, provide training and tools to improve data quality and traceability, and set up open source libraries to enable the network to collectively improve, maintain and use the various categories of tools for making the best of the wealth of data mediated through GBIF.

2.3. Potential for Innovation versus Risk Level

2.3.1. Potential for scientific, technological and/or socioeconomic innovation breakthroughs

To know what we know in relation to biodiversity is something we have not yet attained. The diversity of actors involved in gathering and using data at all scales for a variety of uses --that go from recreational to adaptation plans fro global change-- is overwhelming and the result is that all those ends work with a fraction of the knowledge, and duplicating efforts in obtaining information know somewhere else. This action aims to bring down the level of duplicity in efforts related to biodiversity data gathering and use, and increase significantly the return-of-investment of the many ongoing initiatives that serve and use these data. Besides, access to integrated biodiversity data will enable better decisions in relation to the environment and new opportunities, scientific as well as commercial.

4. IMPLEMENTATION

3.1. Description of the Work Plan

3.1.1. Description of Working Groups

There will be two kinds of Working Groups: “coordination” and “capacity building” WGs.
Coordination Working Groups

WG1: Linking projects and COST actions

Objectives: WG1 aims at making the most of the efforts and funding invested in other COST actions and EU projects, which have some degree of overlap and complementarity with the present COST action, by organising the most efficient possible synergy for enriching the results and reinforcing their sustainability.

Tasks: WG1 will organise the synergy with overlapping and complementary COST actions and EU projects, through the following tasks:

- WG1.1: inventory and document the COST actions and EU projects which have some overlap and/or complementarity with some of this action’s objectives;
- WG1.2: set up mailing lists and sectors in the Action’s website dedicated to the cooperation with the relevant COST actions and EU projects;
- WG1.3: organise the participation of key contacts from these actions and projects in the present action’s relevant workshops;
- WG1.4: for each task in the Action, identify relevant outputs to other actions and projects, and reciprocally relevant inputs from other actions and projects, and set up a side work plan to ensure a timely exchange of these inputs and outputs;
- WG1.5: implement this side work plan.

Milestones: Tasks WG1.1 to WG1.5 feed into each other. Their scheduling and dependencies are indicated in the Gantt Diagram presented in paragraph 3.1.2..

Deliverables: WG1 will deliver (1) a documented inventory of overlapping and complementary COST actions and EU projects, (2) interactive communication platform and sectors on the Web site dedicated to cooperation with the selected actions and projects, (3) a set of inputs and outputs to be exchanged with the selected actions and projects, and (4) contributions to the annual and final reports.

WG2: Long term initiatives

Objectives: WG2 aims at benefiting from the possible synergies with long term initiatives -e.g. EEA, GEOSS, LifeWatch, etc.-, at increasing the present COST action’s impact, and at reinforcing the long term sustainability of its results.

Tasks: WG2 will liaise with long term initiatives through the following tasks:

- WG2.1: inventory and document the long term initiatives which have overlapping and/or complementary objectives with the present action;
- WG2.2: set up mailing lists and sectors in the action’s website dedicated to the liaison with the relevant initiatives;
- WG2.3: organize the participation of key contacts from these initiatives in the present Action’s relevant workshops;
- WG2.4: for each task in the present action, identify relevant outputs to these long term initiatives, and relevant inputs from these long term initiatives, and set up a side work plan to ensure a timely exchange of these inputs and outputs;
- WG2.5: implement this side work plan.

Milestones: Tasks WG2.1 to WG2.5 feed into each other. Their scheduling and dependencies are indicated in the Gantt Diagram presented in paragraph 3.1.2..

Deliverables: WG2 will deliver (1) a documented inventory of long term initiatives with overlapping and complementary objectives, (2) mailing lists and Web site sectors dedicated to cooperation with the selected initiatives, (3) a set of inputs and outputs to be exchanged with the selected initiatives, and (4) contributions to the annual and final reports.

WG3: Outreach and training

Objectives: WG3 aims at widening the network of participants, so as to enrich the set of skills, experience, tools, data, and use cases available to this COST action, to expand the capacity building impact and to strengthen the results sustainability. Besides providing training is one of the best actions to bring new participants into the network and demonstrate value. In this respect outreach and training go hand in hand. Special emphasis will be made to bring in new researches and participants from countries not yet represented in the network.

Tasks: WG3 will organise the communication with COST Near Neighbour Countries (NNCs) and COST International Partner Countries (IPCs) and foster their participation in this COST action, through the following tasks:

- WG3.1: set up interactive platforms and sectors in the Action's website for NNCs and IPCs;
- WG3.2: consult with NNC and IPC partners and set up a plan for fostering their participation in the action;
- WG3.3: plan and organise specific sessions for NNCs and IPCs in the action's meetings and workshops;
- WG3.4: liaise closely with other WGs and apply the plan to foster the participation of NNC and IPC partners;

Milestones: Task WG3.1 feeds into WG3.2 to WG3.4; WG3.2 feeds into WG3.3 and WG3.4; WG3.3 produces a plan and reports regarding the specific NNC and IPC sessions in the action's meetings; and WG3.4 produces reports on the general participation of NNC and IPC partners in the action. Their scheduling and dependencies are indicated in the Gantt Diagram presented in paragraph 3.1.2..

Deliverables: WG3 will produce dedicated mailing lists and Web site sectors, as well as a set of plans and reports regarding the participation and training of NNC and IPC partners.

3.1.1.2 Capacity building Working Groups

WG4: Quality control

Objectives: Quality control is central for costly data to be really useful, and a sound evaluation of the data fitness for use is key to relevant data selection and processing for any particular question. WG4 aims at organising the sharing of experience and the easy identification, accessibility and use of the numerous methods and tools independently developed.

Tasks: A wealth of experience has been gained by numerous data holders, GBIF Nodes and users regarding error detection and correction, as well as fitness for use evaluation. This experience needs to be shared, and the methods and tools developed independently must be made accessible to all and collectively improved, updated and maintained on the long term.
WG4 will achieve this through the following tasks:

- WG4.1: organise a workshop gathering data holders and users with GBIF Nodes to share experience on quality control, and fitness for use evaluation;
- WG4.2: inventory and document the numerous methods and tools developed for error detection and correction, and for fitness for use evaluation;
- WG4.3: analyse the issues mentioned by users regarding data quality and fitness for use, and deduce a set of recommendations for improvement;
- WG4.4: design and set up an open source library framework for sharing tools and for collectively updating, improving and maintaining them on the long term;

- WG4.5: collectively populate the library and put it in operation;

Milestones: Task WG4.1 feeds into WG4.2; WG4.2 and WG4.3 feed into WG4.4, which provide the platform used by WG4.5. The tasks scheduling and dependencies are indicated in the Gantt Diagram presented in paragraph 3.1.2.

Deliverables: WG4 will deliver a report from the experience sharing workshop; it will deliver a documented inventory of the methods and tools, and it will produce an operational and populated open source library of tools, for error detection and correction, and for fitness for use evaluation.

WG5: Semantic Web

Objectives: WG5 aims at making the best of all Web services related to biodiversity information and overcome their present heterogeneity. It also aims at improving data interoperability by using semantic aware technologies, and at facilitating and widening the use of persistent identifiers, by sharing experience, methods and tools among the various actors handling persistent identifiers, and by training new actors on their use.

Tasks: WG5 will sort out the landscape of Web services related to biodiversity information, so as to overcome the present heterogeneity and agreeing on common principles. It will also build on the experience acquired by a number of institutions with semantic aware technologies and persistent identifiers, so as to improve the methods and tools for handling them, and to train further institutions to efficiently use them. WG5 will achieve this through the following tasks:

- WG5.1: organise a workshop gathering developers and key users of biodiversity information related Web services to compare methods and tools and share issues from developers and users experience;
- WG5.2: inventory, document and categorise the numerous Web services;
- WG5.3: develop common principles compatible with all categories, design and set up a directory framework, collectively populate the directory and put it in operation;
- WG5.4: organise a workshop gathering data holders and users with GBIF Nodes to share experience on semantic aware technologies and persistent identifiers;
- WG5.5: organise training sessions on persistent identifiers, so as to expand the use of existing methods and tools;
- WG5.6: design and set up an open source library framework for sharing tools related to persistent identifiers, and for collectively updating, improving and maintaining them on the long term; populate the library and put it in operation;

Milestones: Task WG5.1 prepares the basis for WG5.2, which feeds into WG5.3. Tasks WG5.4 to WG5.7 feed into each other. The tasks scheduling and dependencies are indicated in the Gantt Diagram presented in paragraph 3.1.2.

Deliverables: WG5 will deliver (1) reports from the workshops, (2) an organised and documented inventory of existing Web services dedicated to biodiversity information; on this basis it will then deliver (3) a directory framework for these services; and it will produce (4) an operational directory populated with all Web services developed by institutions involved in this action. WG5 will also deliver (5) a set of training materials and a training session, and (6) an operational and populated open source library of tools, for handling persistent identifiers.

WG6: Specimen data capture

Objectives: WG6 aims at making the most of the recent investment by numerous countries in massive collection digitisation initiatives and in subsequent specimen label data capture initiatives, by encouraging this effort to be continued and expanded and other innovative data capture methods (e.g literature mining) explored.

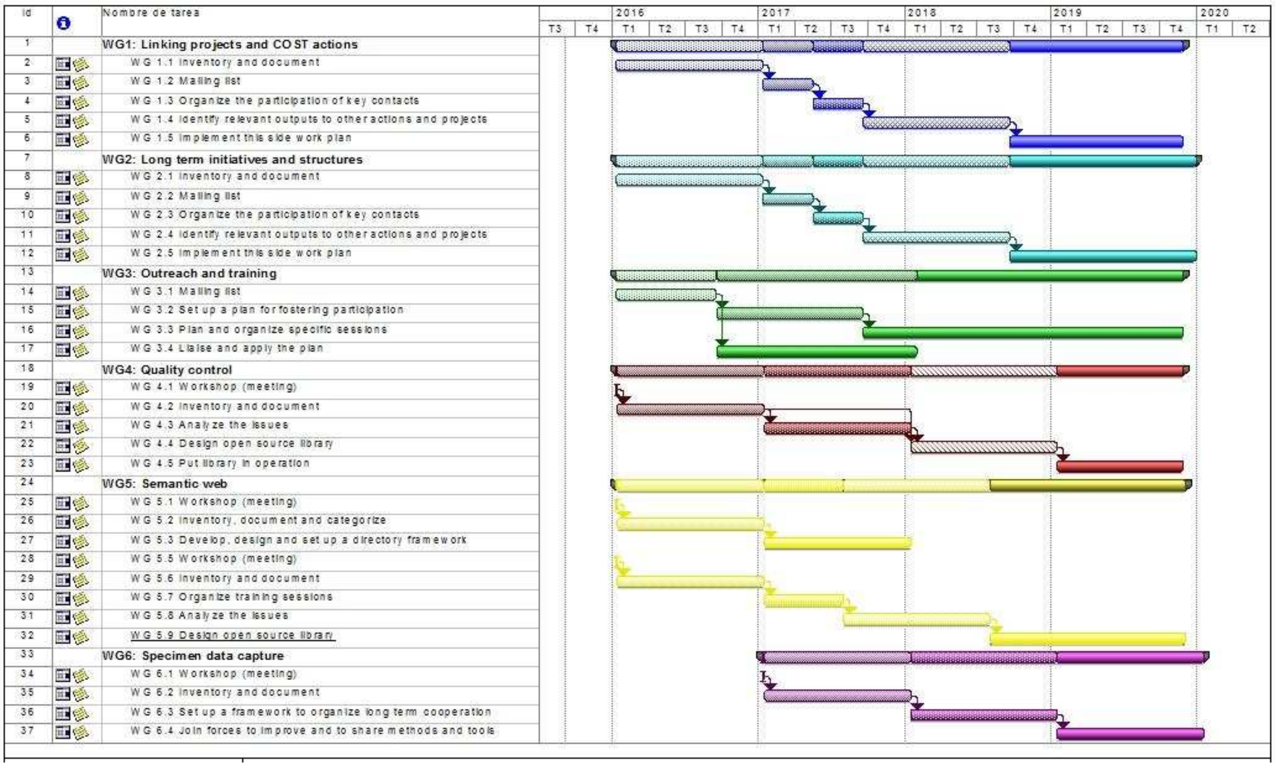
Tasks: Institutions holding large collections in several countries have undertaken to digitize them - scan the specimens-, and to use the images to capture the data recorded on the labels through a combination of OCR and citizen science approaches. WG6 will encourage expanding specimens digitisation and labels data capture through the following tasks:

- WG6.1: organise a workshop to share experience from recent and current massive collection digitisation, and specimen data capture initiatives;
- WG6.2; inventory and document the methods and tools for massive collection digitisation and for specimen data capture;
- WG6.3: set up a framework to organise long term cooperation between national initiatives in massive collection digitisation and specimen label data capture;
- WG6.4: join forces to improve methods, and to share, update, improve and maintain tools in open source;

Milestones: Task WG6.1 to WG6.4 feed into each other. The tasks dependencies and scheduling are indicated in the Gantt Diagram presented in paragraph 3.1.2.

Deliverables: WG6 will deliver documented inventories of recent and current initiatives in massive collection digitization and specimen label data capture, as well as reports from the experience sharing workshops; it will also provide a Web site sector dedicated to cooperation in collection digitization and specimen label data capture, including methods documentation and an open source platform for tools (e.g. specimen label specific OCRs, citizen science interface for collective moderated label data capture, etc.) sharing and long term collective improvement, update and maintenance.

3.1.2. GANTT Diagram



Id	ℹ	Nombre de tarea
1		**WG1: Linking projects and COST actions**
2		WG 1.1 inventory and document
3		WG 1.2 Mailing list
4		WG 1.3 Organize the participation of key contacts
5		WG 1.4 identify relevant outputs to other actions and projects
6		WG 1.5 implement this side work plan
7		**WG2: Long term initiatives and structures**
8		WG 2.1 inventory and document
9		WG 2.2 Mailing list
10		WG 2.3 Organize the participation of key contacts
11		WG 2.4 identify relevant outputs to other actions and projects
12		WG 2.5 implement this side work plan
13		**WG3: Outreach and training**
14		WG 3.1 Mailing list
15		WG 3.2 Set up a plan for fostering participation
16		WG 3.3 Plan and organize specific sessions
17		WG 3.4 Liaise and apply the plan
18		**WG4: Quality control**
19		WG 4.1 Workshop (meeting)
20		WG 4.2 inventory and document
21		WG 4.3 Analyze the issues
22		WG 4.4 Design open source library
23		WG 4.5 Put library in operation
24		**WG5: Semantic web**
25		WG 5.1 Workshop (meeting)
26		WG 5.2 inventory, document and categorize
27		WG 5.3 Develop, design and set up a directory framework
28		WG 5.5 Workshop (meeting)
29		WG 5.6 inventory and document
30		WG 5.7 Organize training sessions
31		WG 5.8 Analyze the issues
32		WG 5.9 Design open source library
33		**WG6: Specimen data capture**
34		WG 6.1 Workshop (meeting)
35		WG 6.2 inventory and document
36		WG 6.3 Set up a framework to organize long term cooperation
37		WG 6.4 Join forces to improve and to share methods and tools

3.1.3. PERT Chart (optional)

3.1.4.

3.1.5. Risk and Contingency Plans

Risk	***Assesment***	***Measures / Contingency plans***
Failure to engage the wider community involved in biodiversity data	**Likelihood** Médium **Impact on the Action** Systemic **Impact severity** High	Having a wide base of Action proposers with a good geographic spread, and a portfolio of contacts and collaborations in countries and regions initially outside the Action. Specific communication actions under WG3
Failure to maintain the network after the Action finalizes	**Likelihood** Médium **Impact on the Action** Systemic **Impact severity** Medium	Embedding Action activities in those of the relevant long term initiatives such as GBIF, EEA, or Lifewatch Taking advantage of the facilities provided by these long term initiatives for communication, and data and documentation maintenance Involving young generation of managers and researchers in training and STSMs as a priority
Lack of expertise on the target areas of the Action	**Likelihood** Low **Impact on the Action** Depending on the WP affected: WP3-WP6 **Impact severity** Variable	Having a wide base of partners with a good range of expertise and experience in the areas targeted by the Action
Lack of know-how to carry out activities	**Likelihood** Low **Impact on the Action** Depending on the WP affected: WP3-WP6 **Impact severity** Medium	Having a wide base of partners with the capacity and facilities to carry out the planned activities
Failure to engage the projects and other COST actions with common or compatible objectives	**Likelihood** Medium **Impact on the Action** WP1 **Impact severity** High	Reach out for projects and actions at an early stage with enough time to coordinate planned activities and establish collaborations. Being flexible in how coordination is managed, to avoid disturbing plans or target projects and actions, so that collaboration, and subsequent harmonization and integration are feasible Involving partners in those initiatives in this Action
Failure to engage with ongoing long term initiatives with common or	**Likelihood** Medium **Impact on the	Building an understanding of the rules and objectives followed by the targeted long term initiatives

compatible objectives	**Action** WP2 **Impact severity** High	
Failure to identify the aspects of biodiversity data management and use critical to make the many initiatives involved converge on shared standards, tools and approaches that will result in real integration of biodiversity data	**Likelihood** Medium **Impact on the Action** Systemic **Impact severity** High	Engaging policy-makers in the Action. Involve experts and practitioners from the different aspects of biodiversity data (identification, capture management, curation, dissemination, analysis, synthesis).

3.2. Management structures and procedures

At the basis of this COST Action is the existence of a network of GBIF national Nodes, which are the coordinators of biodiversity information gathering, management and publication in their respective countries. GBIF Nodes are organised in six regions, one of them is Europe. Within this framework, annual meetings are held, and contacts and collaborations among national GBIF Nodes maintained.

Taking this into account, the Action aims to align its management structure with the regional coordination procedures of GBIF in Europe. This approach will provide this COST Action with a quick start, and a lightweight deployment, easy to maintain beyond the temporal scope of the Action. This is in itself a risk control and sustainability strategy.

The Management Committee (MC), will be set up in accordance with the "COST Open Call Submission, Evaluation, Selection and Approval (SESA) Guidelines", will be holding 1-2 meetings a year, whenever possible in association with GBIF meetings in Europe.

The following WGs will be organised:

WG1: Linking projects and COSTs Actions
WG2: Long-term initiatives
WG3: Outreach & training
WG4: Quality control
WG5: Semantic web
WG6: Specimen data capture

The same lightweight philosophy depicted for the MC applies for WG; because "integrating integrations" is not about creating parallel networks or a new overarching network, but finding common understandings and solutions. Thus, WG events and meeting will be organized as much as possible in association with events of other initiatives (Projects, Actions, etc.). Accordingly, WG activities will be coordinated with those of projects, other COST actions, or entities with a focus on biodiversity data and information, through collaboration and harmonization of data and procedures. Hopefully this approach will also contribute to lighten Action's carbon footprint, and participants travel agendas.

3.3. Network as a whole

Because the complex, global and diverse nature of how biodiversity information is acquired and used, a network aiming to connect and harmonize its fundamental aspects is a most suitable approach to gain overall effectiveness in the many initiatives and uses around biodiversity data. No project can do that, it is the contacts, the building of the common understanding, the action to bring into the mainstream those that are not, and to work to sustain the connections, the open source developments, and the capacity transfer mechanisms what can make a significant impact in the long run.

The Action's proposes comprises leaders or coordinators of national biodiversity research networks, among them we find:

- Implementers of advanced data portals with state-of-the-art query, visualization and API capabilities, or have
- Strong training programs, in Biodiversity informatics
- Developers and Implementers of advanced tools and methods, such as as persistent identifiers, controlled vocabularies, RDF access, etc.
- Inclusiveness Target Countries (ITC) and partners with strong links with then
- High throughput digitisation projects of natural history collections.
- Participants of relevant projects and initiatives as those cited under section 1.4.2. (e.g. EUBON, Lifewatch, GBIF, etc.), or when not, with good linkages to those.
- Some of the largest data providers in GBIF.

The project proposers represent an excellent geographic coverage of Europe, comprising so far 16 countries of which 5 are Inclusiveness Target Countries (ITC). Furthermore, some of the proposers have contact and working relations with countries from other regions as Africa, America, and South-East Asia. Bringing international partners into the Action is pertinent as standards, and innovative tools and procedures for capturing, managing and publishing biodiversity data are needed, and developed all over the word; besides, to respond to global environmental issues, global data --at different scales are needed.

In summary, the proposers combine partners and countries that are leaders in many of the components needed to achieve true integration of biodiversity data with others with potential to acquire those capacities and then contribute effectively to the global picture of biodiversity information and benefit from the information available; to do better science and to respond more effectively to societal key challenges such as global change, invasive species, land management under multiples pressures, or maintaining ecosystem services (water, pollinators, etc.)