



Outputs from the SBIF Review Workshop on Data Flows

18-19 September 2017

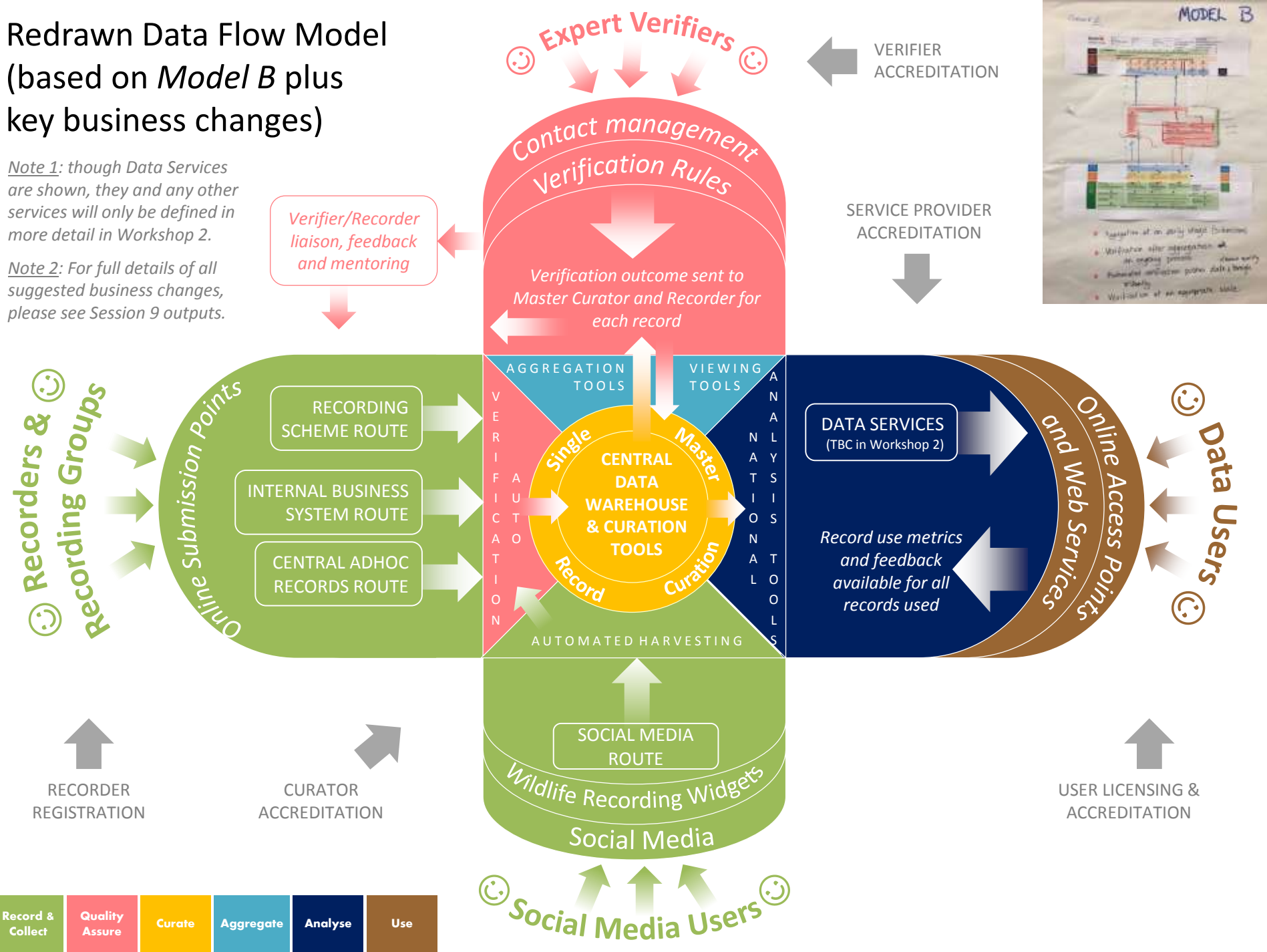
Key messages from Workshop 1

1. That overall data flow should be centralised for maximum efficiency and to facilitate the availability of records for everyone (whether local, regional, national or central).
2. That a single centralised route for the submission of *ad hoc* records should be established to accept records from anywhere in the UK.
3. That both off line and online capture of records should be accommodated, including those from social media, but formal record submission is to be online.
4. That verification happens at multiple points during the data flow pathway; auto-verification is crucial for efficiency in handling the bulk of records that can be accepted without further assessment, while allowing records that need expert verification to be flagged. Both verified and unverified records to be aggregated but with a 'quality stamp' so that they are available and are of known quality.
5. That aggregation should be facilitated as early as possible in the data flow pathway so that the aggregated records can help verification decisions and so that all data are available in context (with sensitive data restricted appropriately) for all users.
6. That analysis tools are required at a national level for Scotland to meet Scottish needs while being part of a shared UK toolset.
7. That improved feedback to recorders and ongoing access to their own records is key for effective engagement and recognition of recorders.

Redrawn Data Flow Model (based on *Model B* plus key business changes)

Note 1: though Data Services are shown, they and any other services will only be defined in more detail in Workshop 2.

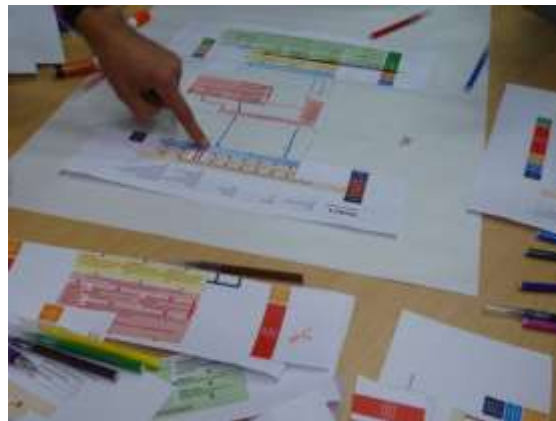
Note 2: For full details of all suggested business changes, please see Session 9 outputs.



Workshop participants



Left to right: Colin McLeod, Ellen Wilson, Brian Boag, Alistair Lavery, Jim McIntosh, Ashleigh Whiffin, Philip Briggs, Rachel Stroud, Elspeth Haston, Andy Musgrove, Sydney Gauld, Vlad Blagoderov, Stephen Moran, Tim Foster, Claire Seymour, Oli Grafton, Christine Oines, Kelly Ann Dempsey (entirely behind Christine), Andy Tharme, Christine Johnston, Will George, Mike Kerr, Jackie Stewart, Oli Pescott, Gill Dowse, Mark Pollitt, Liz Edwards, Jo Judge (entirely behind Liz), Zoe Randall. [Battleby Conference Centre, 18 September 2017]



Workshop Objectives

- To inform attendees about the SBIF Review and progress towards a sustainable biological recording infrastructure
- To harness the expertise of participants in furthering the business case for change
- To develop a vision for effective data flows by:
 - Identifying potentially valid models based on agreed design principles
 - Comparing the effort and value of each model
 - Recommending our preferred model(s) to take forward to future workshops

Workshop sessions

1. Icebreaker question
2. SBIF Review so far...
3. Review of the case for change
4. Review of the data flow design principles
5. Data flow model variants and their pros and cons
6. Adapted data flow models
7. Feedback on adapted models
8. The preferred model
9. Business changes needed
10. Workshop feedback

Icebreaker

SESSION 1

1a) Icebreaker question: Given the workshop objectives, what is the biggest benefit for you in identifying a new model for data flow?

LERCs, Recording Groups, SBIF Review Working Group:

Reduced duplication of effort

One place to submit records and one place to get them back

Traceability

Get records from field to end user as quickly as possible

System that can cope with information of many formats. Standardisation.

Local face as a way to ease into flow.
Point of contact

More people using biological data

Always know where something has come from, where it's been and what has been done

Space for biological 'side notes' in submission + curation of records

1b) Icebreaker question: Given the workshop objectives, what is the biggest benefit for you in identifying a new model for data flow?

Museums, Botanic Gardens/SBIF Advisory Group:

Better buy-in from across different sectors.
Confidence in data and in the process.

Free people up to add value rather than spend time dealing with data flow issues.

Data visibility.

Linking data effectively

Publish once, use many times

Transparency of data flows

Traceability: being able to track the origin/path of individual records

Ease of getting data into the data flow

Achieve alignment and elegant data flow so less pressure on individuals; make it easy

Verification: to be able to ensure records have been verified (reliable records)

Removing access controls so data flow is easy

Maintaining the data in the data flow (i.e. updates, annotations, corrections etc)

Recognition that data used for regional and national purposes is the same essentially

1c) Icebreaker question: Given the workshop objectives, what is the biggest benefit for you in identifying a new model for data flow?

Local Government, National Government, Academia:

Easier to ID data gaps

Better flow, fewer
bottlenecks

Simpler for collectors and
end users

Efficient, better value

Consistency, e.g. for
planning.

Identifying key points for
investment

Increased trust – between
sectors/organisations and in
the data

Clearer, more balanced
focus for investment

1d) Icebreaker question: Given the workshop objectives, what is the biggest benefit for you in identifying a new model for data flow?

NGOs and Recording Schemes:

More usable data in the long term (i.e. sustainable)

Effective feedback mechanisms to maintain volunteer involvement

Add non-conservation data – soil type etc

Can enter data where I want and know all those who need it can get it

Obvious and simple verification process

Include info on species and method for each record so data users can assess reliability of record

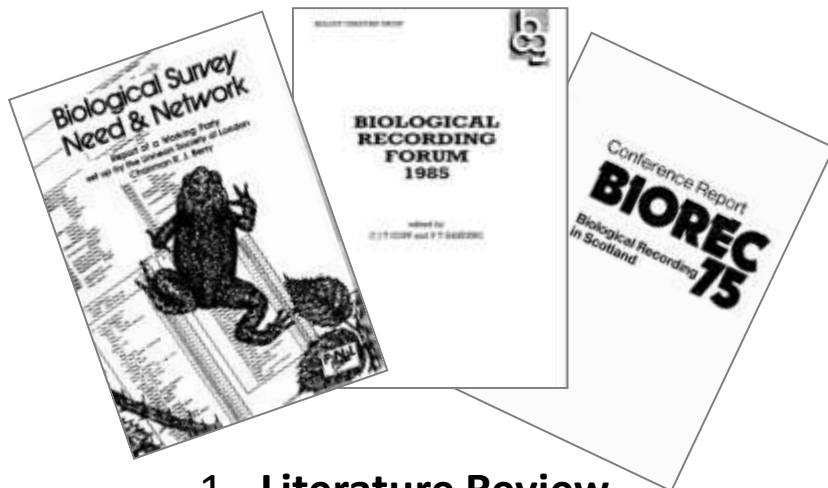
A system which supports and values data providers

Schemes and societies could get more data + that data is validated and made available to end users

SBIF Review so far...

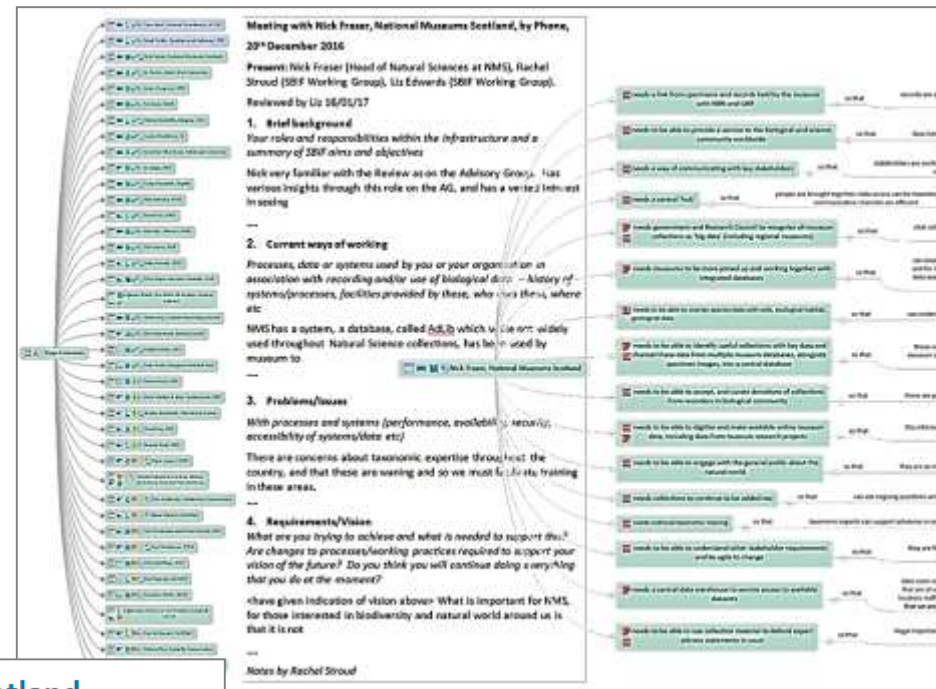
SESSION 2

2: Findings of the SBIF Review so far



1 - Literature Review

2 - Interviews



SBIF Review of the Biological Recording Infrastructure in Scotland

Welcome to the SBIF Review Questionnaire

Dear Questionnaire Participant

Many people are involved in the collection or use of biological records - together we are a vital network with a shared desire to understand, enjoy and protect the biodiversity around us. We are needed more than ever as pressures on the environment are growing and biological records are essential for monitoring species and habitat change, informing planning and conservation decision making and bringing people closer to the natural world. Yet the complexity of our biological recording communities and infrastructure for collecting and sharing biological records, along with the difficulties of securing long term funding, may mean that we are less effective collectively than we could be.

3 - Questionnaire

For full details of the findings of the SBIF Review questionnaire, Interviews and literature review, please visit the SBIF Review web pages:

<https://nbn.org.uk/about-us/where-we-are/in-scotland/review/>

Review of the case for change

SESSION 3

3) The Case for Change: review of drivers, objectives and benefits



Specific comments

- Number each box for ease of reference (done)
- Add “of known quality” to the driver for open data (done)
- **Consider adding a driver** relating to “the need for participants to feel confidence in the overall data flow model so that they can then buy into it” (added as Objective 7)
- Not just a demand for open biological data but also environmental and sociological data too (left generic)
- A register of data products could be useful (as well as schemes and services) (added ‘data providers and data products’ to Objective 3)
- **Consider adding a driver** relating to the need to reduce bottlenecks in verification or to increase the number of verifiers (added ‘and fast-flowing through the NBN DFP’ to Benefit 7)
- **Consider adding a driver** relating to the need for increased collaboration across the network (added to Benefit 5)
- Consider the need for an objective to increase the level of feedback to recorders (added to Benefit 1)

Workshop attendees happy overall with the case for change

ENABLING TECH

ENABLING PROJECTS

BUSINESS CHANGES

BENEFITS

OBJECTIVES

DRIVERS

1 To follow once all four workshops have taken place

1 To follow once all four workshops have taken place

1 To follow once all four workshops have taken place

1 Clear data flows and submission points, with duplication of effort eliminated, and **with feedback for Recorders on how records are used**

1 To agree the preferred models for data flow, service provision, governance and funding

1 Insufficient sustainable funding

2 All stages of the NBN Data Flow Pathway are supported by efficient, high quality services that are easily accessible at the point of use

2 To establish sufficient and sustainable funding for the preferred models for data flow, service provision and governance

2 Demand for timely access to Open Data of **known quality**

3 Sustainable funding allows long term planning and continued provision of services

3 To establish a service catalogue and a register of recording schemes, **data providers and data products** aligned and affiliated with the NBN Data Flow Pathway

3 Demand for clarity on how and where to submit data

4 Sustainable funding allows consistent provision of services across all geographic areas

4 To maximise the use of technology and automation to support the preferred models for data flow and service provision

4 Demand for complete coverage for service provision

5 Roles and responsibilities are clear, **collaboration is encouraged and duplication of effort is eliminated**

5 To showcase the ease, efficacy and enjoyment of biological recording

5 Availability and capability of new technology

6 Recorders have consistent access to high quality training, are supported and feel valued

6 To maximise the number of biological records that are openly available, especially for use in environmental decision -making

6 Desire to empower people to discover, enjoy and protect nature

7 All data are easily discoverable, of known quality **and fast-flowing through the NBN Data Flow Pathway**

7 **Desire to build energy, confidence and buy-in with an improved data flow that delivers the SBIF Vision**

10 Taxonomic skills base is increased with more people being engaged in recording biological data and caring about nature

8 Recorders are connected with their data throughout the whole NBN Data Flow Pathway

11 Environmental decisions are properly informed by the consistent use of high quality biological data

9 Technology and automation are fully utilised to free up time for added-value activities

Revisions following Workshop 1 shown in yellow; further revisions to follow Workshops 2-4...

Review of the data flow design principles

SESSION 4

4a) Design principles: review of the Data Flow Design principles

Session 3: Data Flow Design Principles

Each model must:

1. Facilitate a single master version of each record
2. Facilitate single submission and curation of records per route/scheme
3. Facilitate access to, and management of, own records
4. Facilitate full coverage (geography/species/habitats)
5. Facilitate open data, allowing for sensitivity restrictions
6. Provide one place where all data for a given use can be found
7. Make available records of known quality (verified + unverified)
8. Facilitate equal access for all (local, national and central)
9. Facilitate prompt progress through the six Data Flow Pathway stages
10. Minimise duplication of effort (and acronyms!)



- P6: queries on whether it is possible to provide one place for all data for a given use (agreed this is desirable per use; but not possible across all uses)
- P7: queries on should the principles allow for some records to have their provenance attributed for IPR reasons (they already can under open licences e.g. CC BY; add the words '**and provenance**' to 'of known quality')

Specific queries and notes

- P1: queries on where the master record is held/by whom (which is likely to be with/by its curator); each record submitted needs a permanent unique identifier whether or not it is the master record; are voucher specimens separate to a record? (sometimes, but can be linked as an attribute but depends on museum/business rules for vouchers)
- P2: queries as to why single submission would be 'per route/scheme' (which is because each scheme/route may have different attributes)
- P3: queries as to definition of 'own records' and who is considered to be an owner (which is the person submitting them in this case) and how you define this for historic records (we won't be able to retrospectively but we will be able to record this in future; 'own' in most cases relates to the recorder)
- P3: can we interpret management to mean that records can be edited or deleted at a later point by the recorder? (no – we need to use the word '**access**' rather than manage as we don't mean to infer 'curation' at this point)
- P5: queries on how to reference sensitive records and access permissions and whether different principles should apply for sensitive records (all principles apply to all records whether or not they are sensitive)

Workshop attendees happy overall with the design principles

4b) Design principles: review of the Data Flow Design principles

Revised Design Principles:

Each model must:

1. Facilitate a single master version of each record
2. Facilitate single submission and curation of records per route/scheme
3. Facilitate access for each Recorder to view the records they have submitted
4. Facilitate full coverage (geography/species/habitats)
5. Facilitate open data, allowing for sensitivity restrictions
6. Provide one place where all data for a given use can be found
7. Make available records of known quality and provenance
8. Facilitate equal access for all (local/national/central)
9. Facilitate prompt progress through the six Data Flow Pathway stages
10. Minimise duplication of effort (and acronyms!)



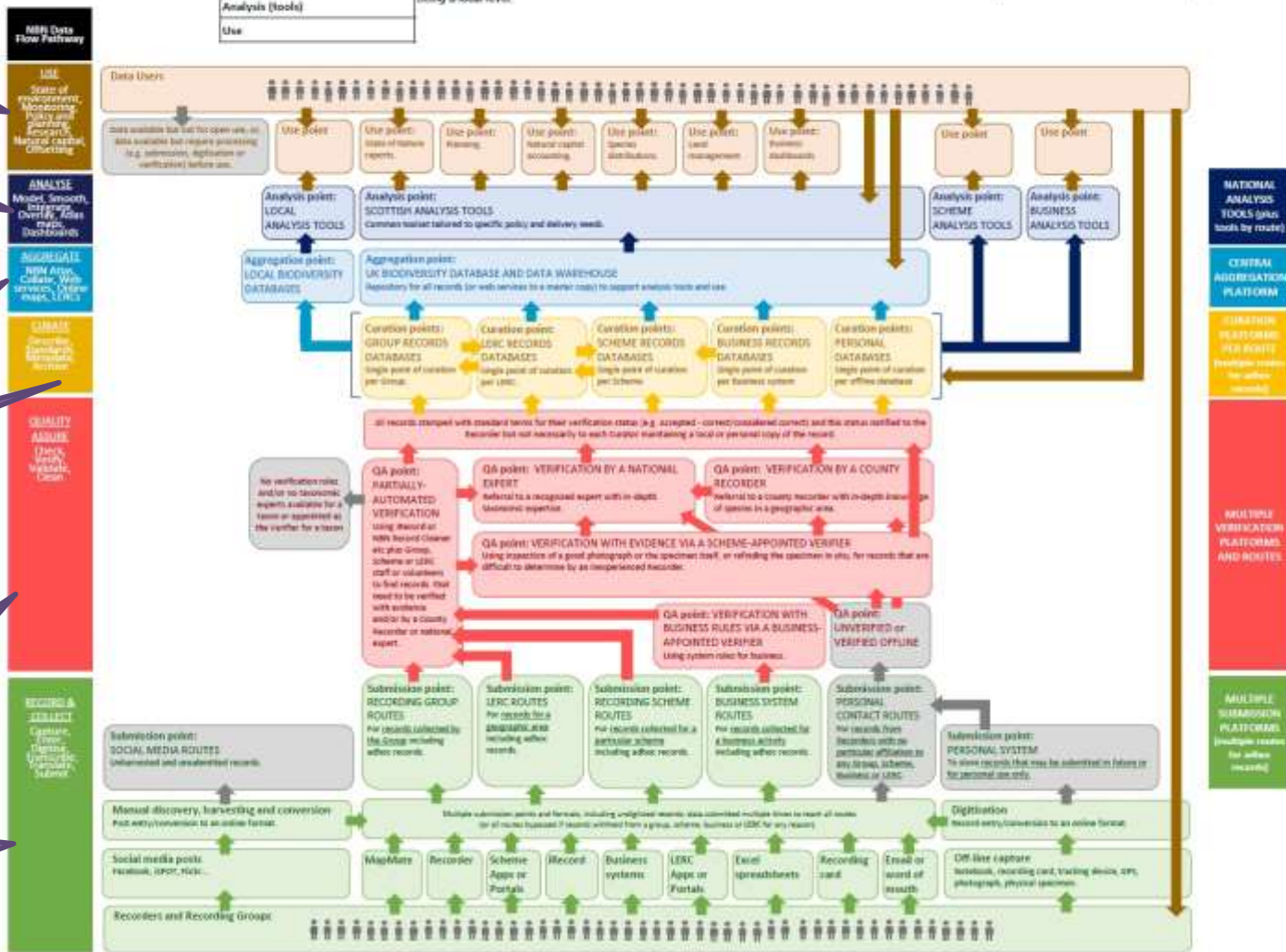
Introduction to the data flow models

SESSION 5

5a) Data flow models: Model 0 – current situation

Model 0 CURRENT SITUATION

LOCAL	REGIONAL	NATIONAL	CENTRAL	SCHEME/BUSINESS
Record & Collect (capture)	Use	Analysis (tools)	Aggregate (data warehouse)	Record & Collect (submit: adhoc/scheme/business)
Record & Collect (submit: adhoc)		Use	Use	Quality Assure (scheme/business)
Quality Assure (adhoc)	NB: the difference between Regional and Local is blurred and so for simplicity this diagram shows activities that may be regional as being a local level.			Curate (scheme/business)
Curate (adhoc)				Analysis (tools)
Aggregate (data warehouse)				Use
Analysis (tools)				
Use				



Overall, in the current situation resources are in play at the local level and within schemes; little is centralised or nationalised so efficiency is minimised but local value is maximised.

Currently, users need to go to multiple data providers (including recorders, LERCs, schemes, organisations, groups, social media, etc) to ascertain and access all possible sources of data

Currently, analysis tools are shaped by national government or LERC service user needs; schemes and organisations also provide analytical tools.

Currently, the NBN Atlas is the primary aggregation point for the UK but there are others which also aggregate (potentially a different version of) records for local and regional use.

Currently, records may be held by more than one curator and/or in databases that aren't openly available. There is effort involved in exchanging records between curators and aggregation points.

This is a generalised view of verification; there are bottlenecks and verification rules may not be comprehensive or current. It can be hard to contact recorders as there is no register of contact details nor contact made.

Currently, lots of different formats are welcomed whether on or offline; social media posts require manual 'harvesting'. Recorders may submit a record multiple times to ensure it reaches all curators or users.

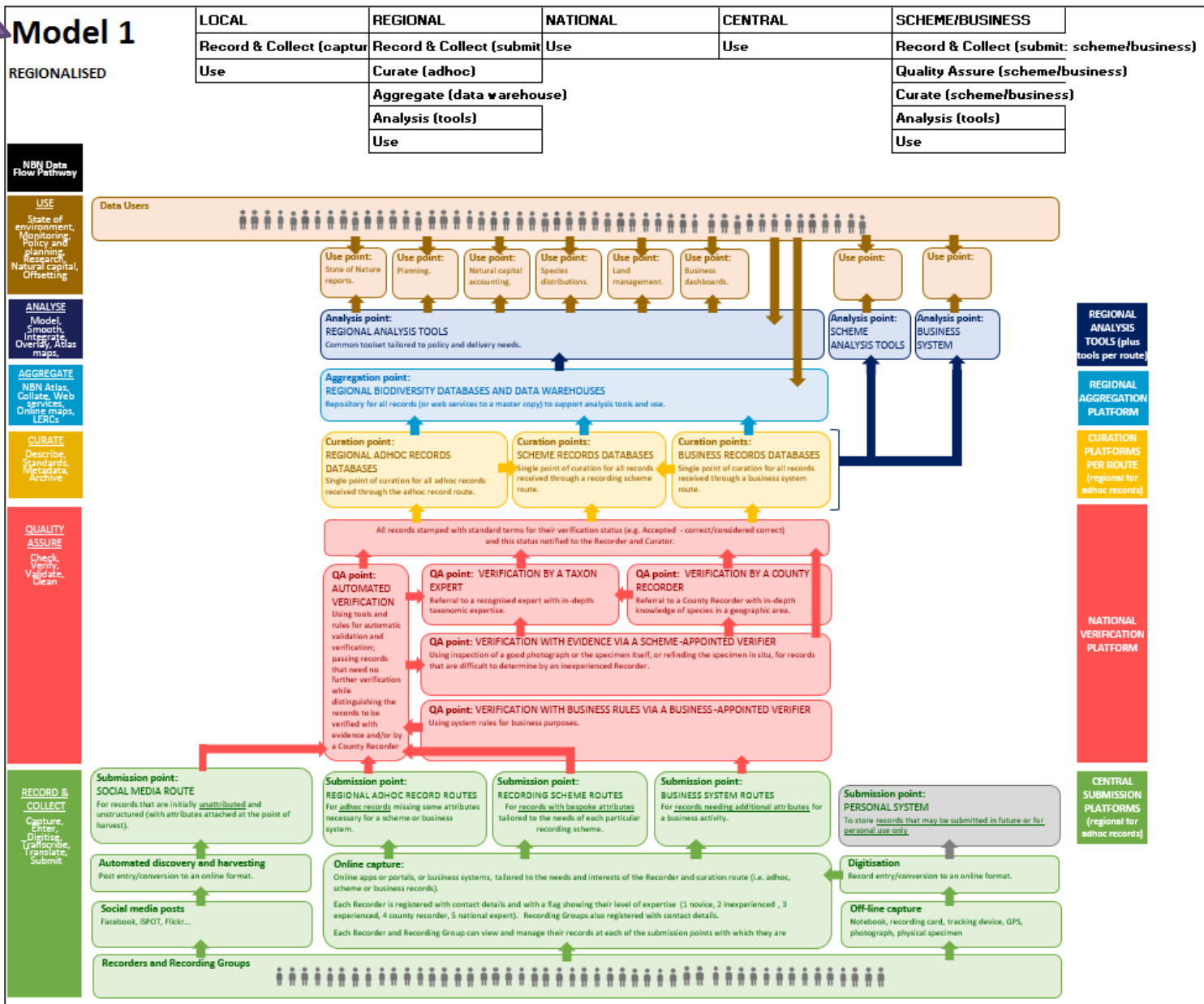
5b) Data flow models: Model 1 – Regional version

Overall, in the regionalised situation resources are in play at the regional level and within schemes or organisations; there are some efficiencies from being regional and still some local value.

Because of the need to meet all of the Data Flow Design Principles, all of the models are broadly the same. All that differs is the LOCATION at which each step of the Data Flow Pathway happens.

A key difference between each model is the route for adhoc records; it is proposed that adhoc records are submitted through a single route. In this model, the submission point is through REGIONAL adhoc records databases.

All of the models show a generalised route for verification and for the harvesting of social media records.

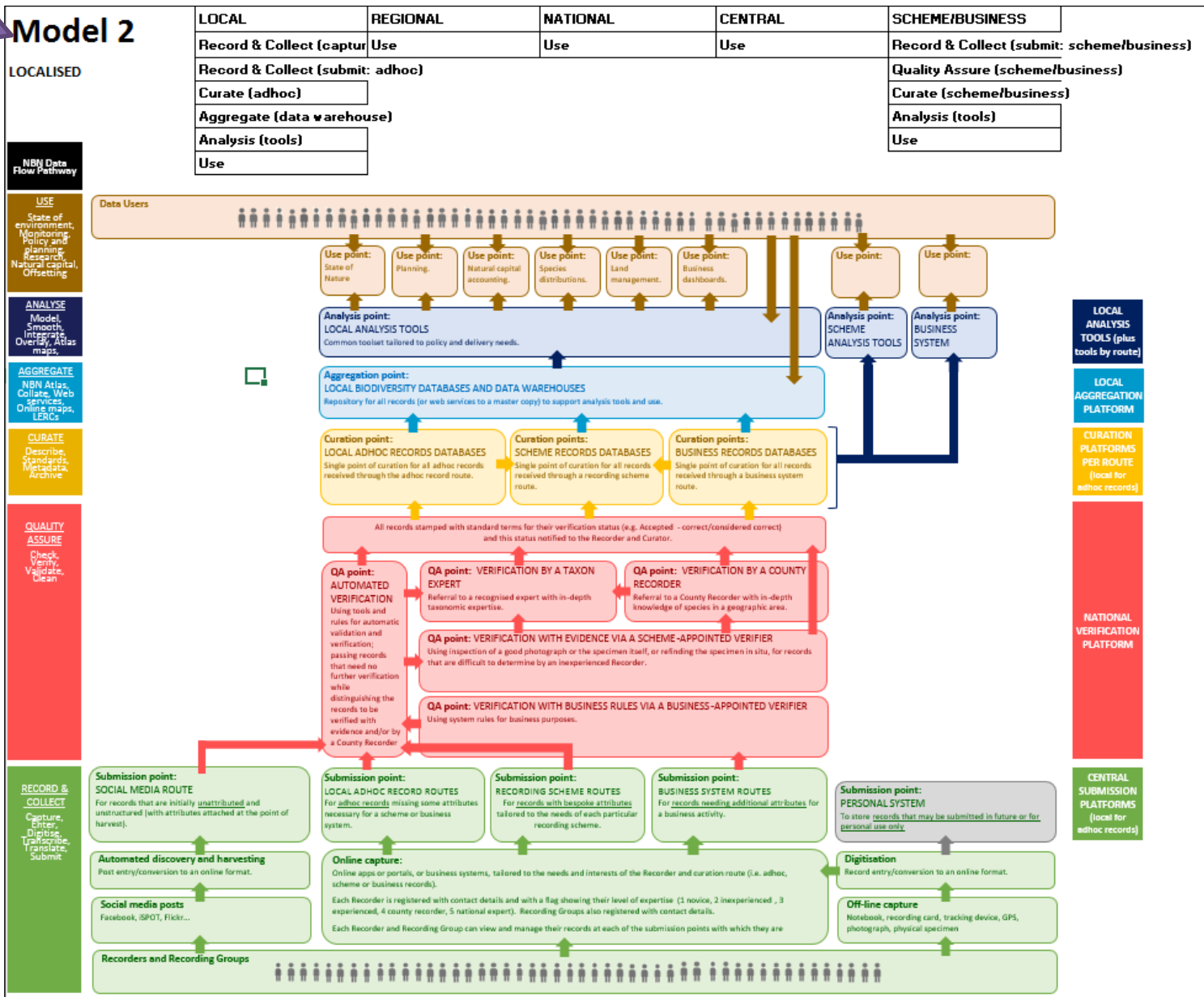


5c) Data flow models: Model 2 – Local version

Overall, in the localised situation resources are in play at the local level and within schemes or organisations; there are no efficiencies from being regional, national or central but local value is maximised.

In this model adhoc records are submitted through a single route via LOCAL adhoc records databases.

The aggregation point and analysis points are also LOCAL (with some tools also provided within schemes and organisations).

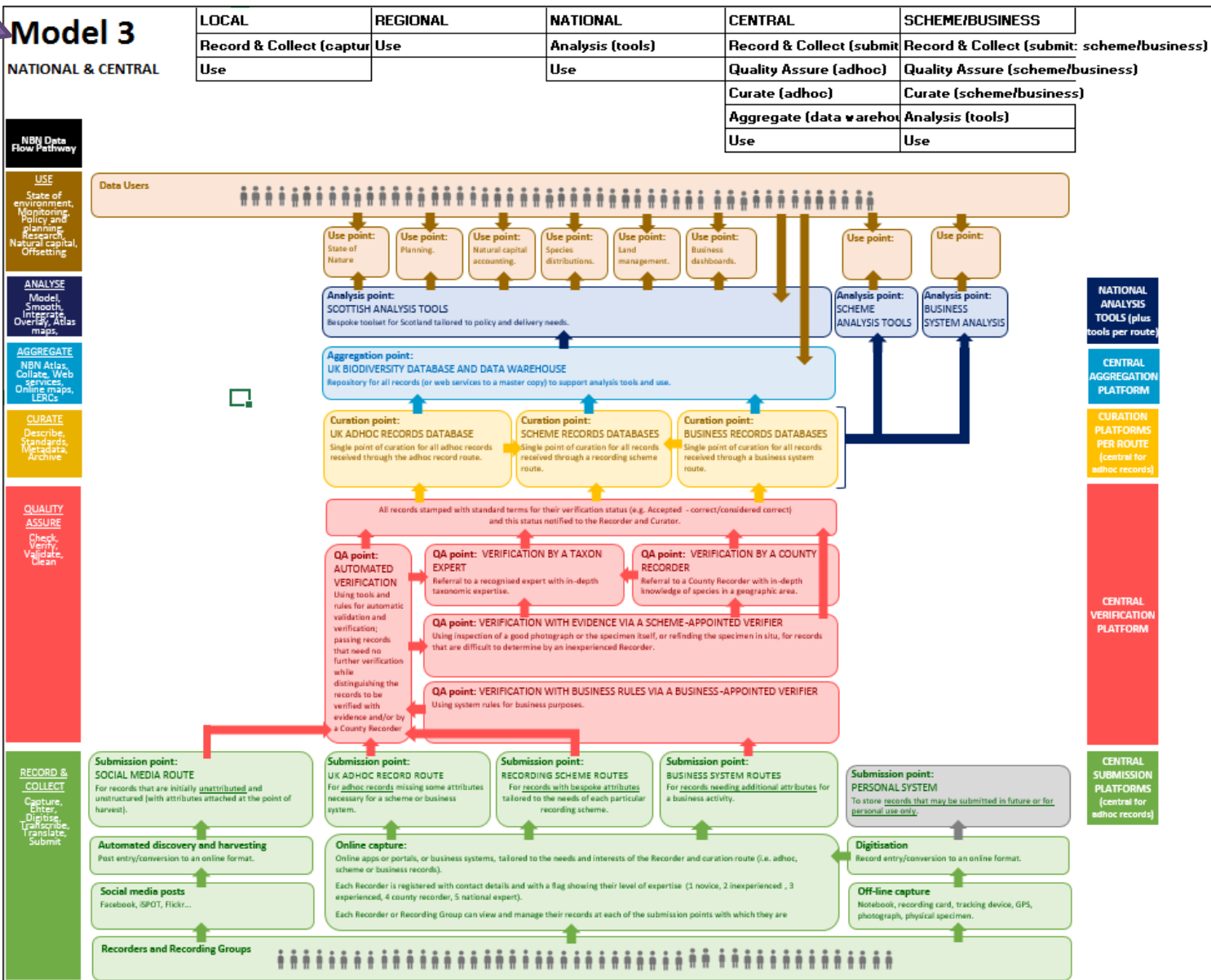


5d) Data flow models: Model 3 – National and central version

Overall, in a national and central situation resources are in play at the national and central level and within schemes or organisations; efficiencies are realised but local value could be diminished.

In this model adhoc records are submitted through a single route via a single CENTRAL (i.e. UK) adhoc records database.

The aggregation point is also centralised, but analysis tools may be primarily national, or central or within schemes and organisations).



5e) Data flow models: Model 4 – Central version

Overall, in a central situation resources are in play at the central level and within schemes or organisations; efficiencies are realised but local value could be diminished.

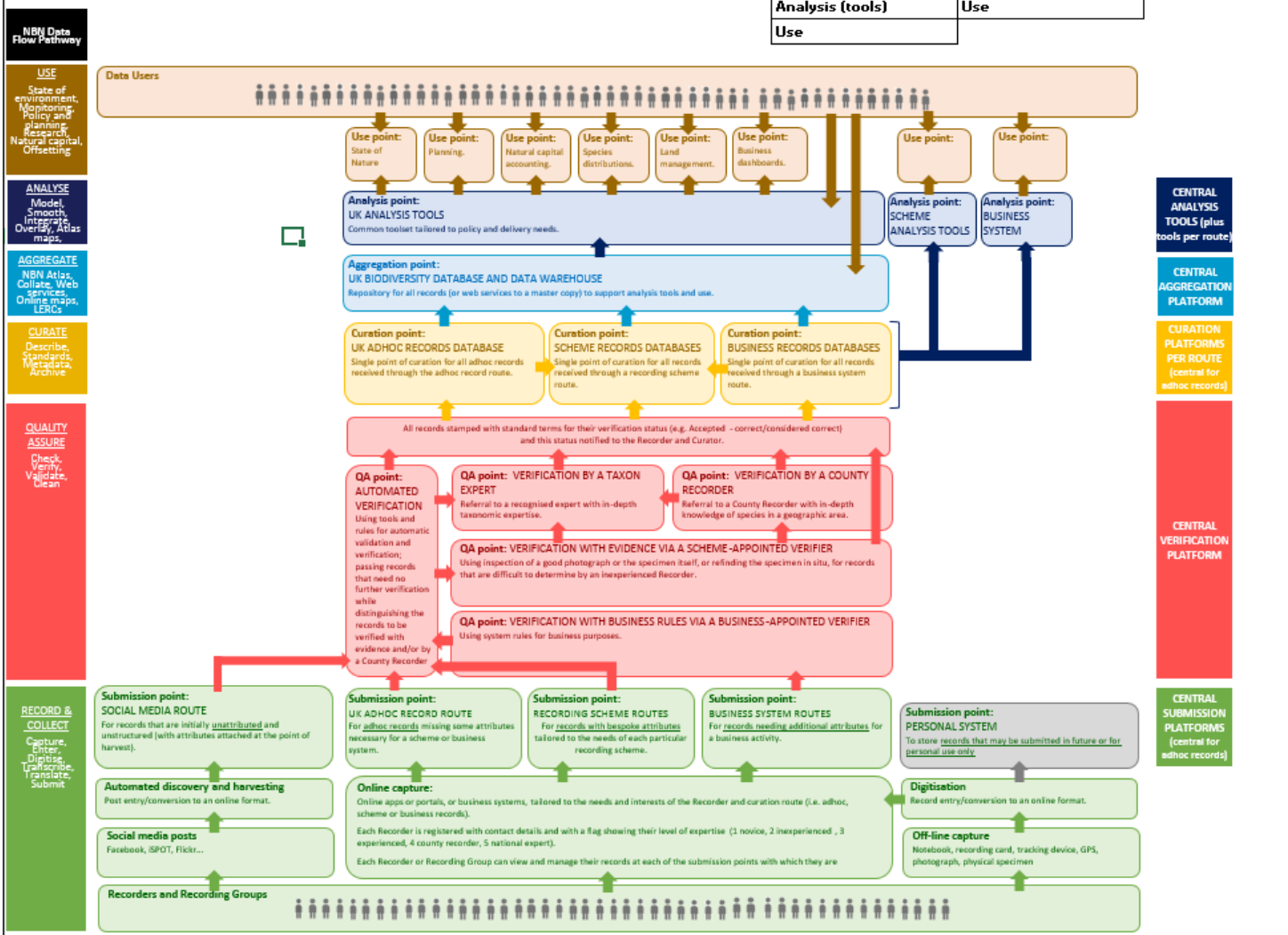
In this model adhoc records are submitted through a single route via a single CENTRAL (i.e. UK) adhoc records database.

The aggregation point is also centralised, as are analysis tools which are CENTRAL or within schemes and organisations.

Model 4

CENTRALISED

LOCAL	REGIONAL	NATIONAL	CENTRAL	SCHEME/BUSINESS
Record & Collect (capture)	Use	Use	Record & Collect (submit)	Record & Collect (submit: scheme/business)
Use			Quality Assure (adhoc)	Quality Assure (scheme/business)
			Curate (adhoc)	Curate (scheme/business)
			Aggregate (data warehouse)	Analysis (tools)
			Analysis (tools)	Use
			Use	

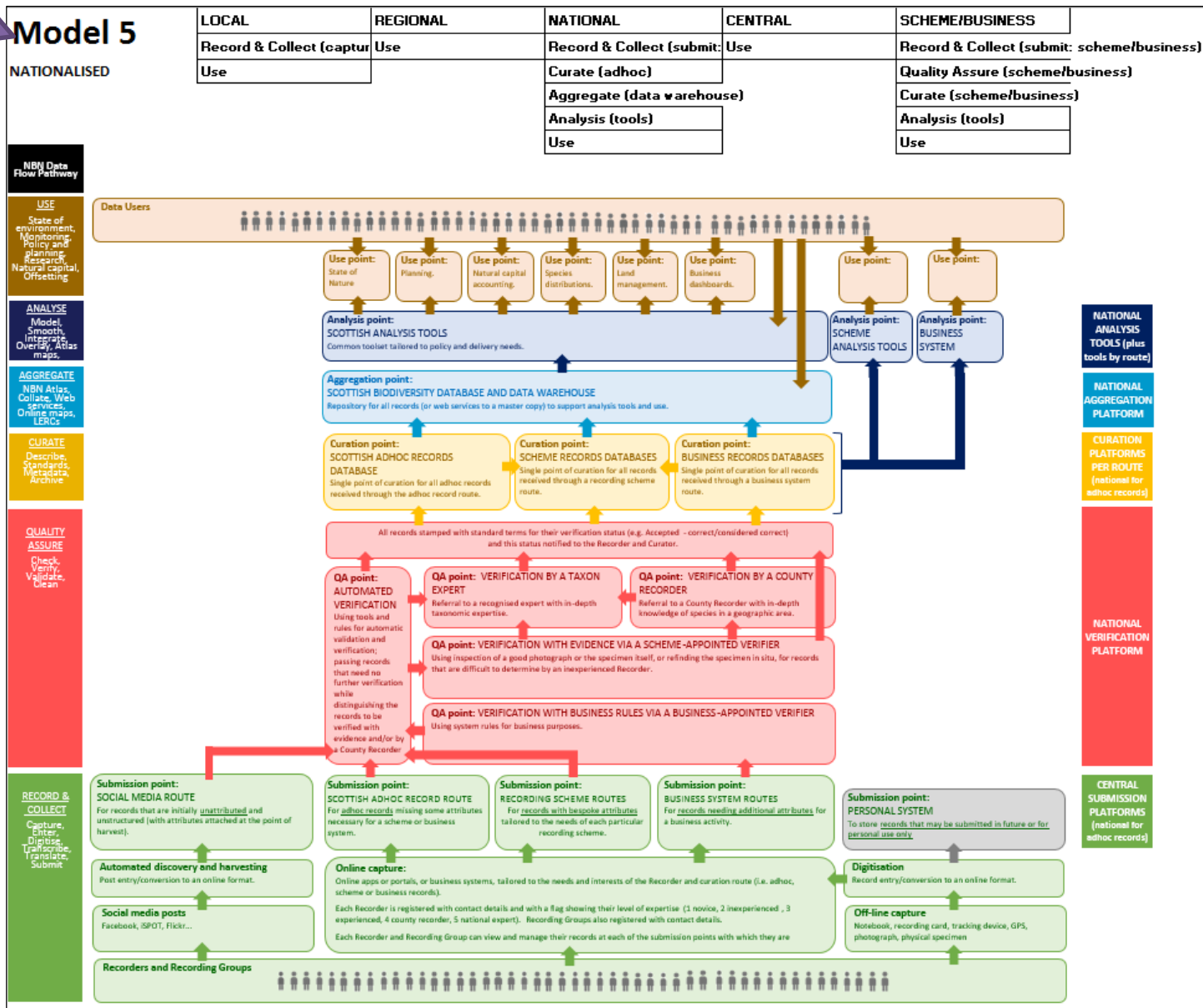


5f) Data flow models: Model 5 – National version

Overall, in a national situation resources are in play at the national level and within schemes or organisations; efficiencies are realised but local value could be diminished.

In this model adhoc records are submitted through a single route via NATIONAL adhoc records databases.

The aggregation point is also nationalised with a NATIONAL data warehouse; analysis tools are nationally provided too, or within schemes and organisations.



5g) Model feedback: pros and cons (rough notes from our discussion)

Group 1: Diagram of Models – Central along bottom. Cost on Y axis. Cost goes down as you aggregate, and reduce duplication of infrastructure etc. What you lose in Local Relevant. The more local a system is, the more relevant you can be locally – feedback to recorders – and you can also be relevant to local authorities and planners so can provide tailored reports. However, if you go to the central point, trying to interpret this in a local way and make interesting to recorders becomes very difficult. If you had to have a system where everyone is happy – is somewhere in the middle. But what we want is a central pot of data, which can be interpreted locally. The more centralised a system becomes, if you can't present data into local/county level for i.e. verifiers at a county level then this becomes an issue. Need good systems to be able to group data. Engagement of people is a key part of having the local aspect of the data model. Engagement is a better word than 'relevance'.

Group 2: Agreed that Model 3 would be preferred model. Interpreted Record and Collect differently to Group 1. Recorders and recording groups would have local engagement but would be slightly simplified from now. Pros allowed national identity of Scotland within a centralised database and could therefore be politically futureproof. Gives the efficiencies and standardisation that are needed to make data flow more easily. A fully centralised model would not be politically acceptable. A nationalised model would cause difficulties for recorders in the borders, and looking at conservation for statutory reporting. Localised would be great for local engagement but lots of complexity, inefficiencies, may isolate people if they don't have access to what others do have elsewhere

Group 3: Didn't look at Current as lots of problems. Model 5: Nationalised model is realistically good, lose of engagement at local level but gain efficiency.

Model 4: quite liked centralised. Some similarities between central and nationalised. Pros in efficiency as can centralise the technical expertise and save cost in upgrading etc. Con is that in its basic form you have one set of analysis tools and each country may need different reporting tools, which also applied to National model. Also maybe lose local creativity that you get from local based models. National/Central: Lots of pros and cons same – efficiency of scale, reduced duplication, ease of sharing internationally (i.e. to GBIF) as this is difficult if you have lots of local databases. Potential for country specific analysis. National models make it harder to create relationships with recorders at local scale.

Model 2: Localised. Pros as can get local engagement, support and knowledge but heavily outweighed by duplication of effort, and lack of expertise.

Model 1: Regional. Pros – more standardised data input. Cons loss of local level interaction and engagement

Issues with localised and regional model is for people on the borders and for those who are travelling and moving around the country when recording.

Group 4: Model 2: Strong links with the local community. Maybe more responsive to local needs so getting things changed would happen quicker. But is expensive and lots of duplication. Maybe gaps in expertise. Maybe creates diversion in approaches so creates problems when aggregating downstream

Central/National: Efficient for verifiers. Could put national badge so easier for funding routes. Easier to implement changes – need cross group agreement before making changes. May be a lack of flexibility. Change for some groups.

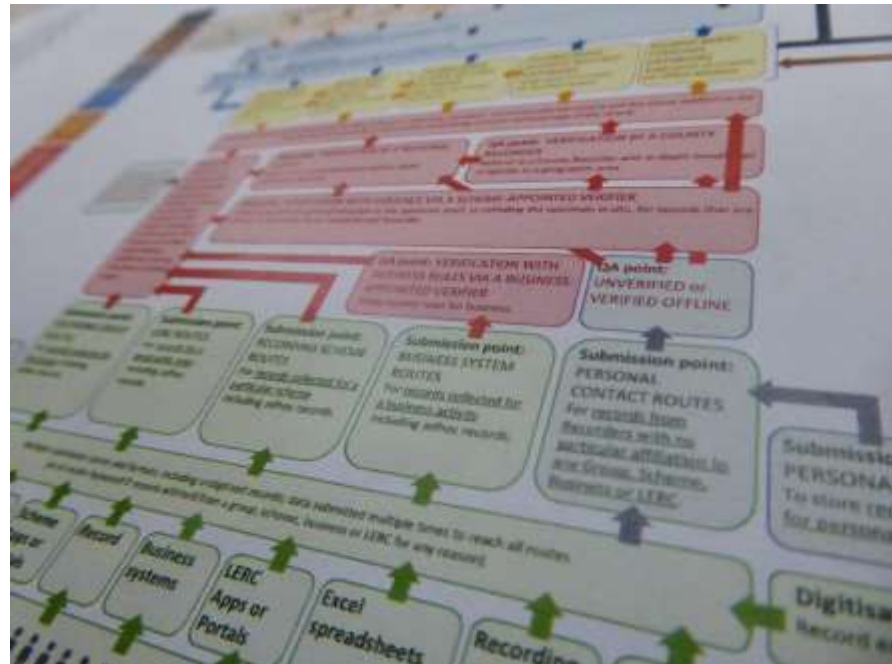
The system we have now people are used to so need to put support in system

Central: More standardised, but potentially would overlook the national level in terms of funding and analysis of results. May be pros and cons to this but might miss something out at national level. May be a vulnerability to everything being at a central level – system down for everyone if something happens. Less flexibility

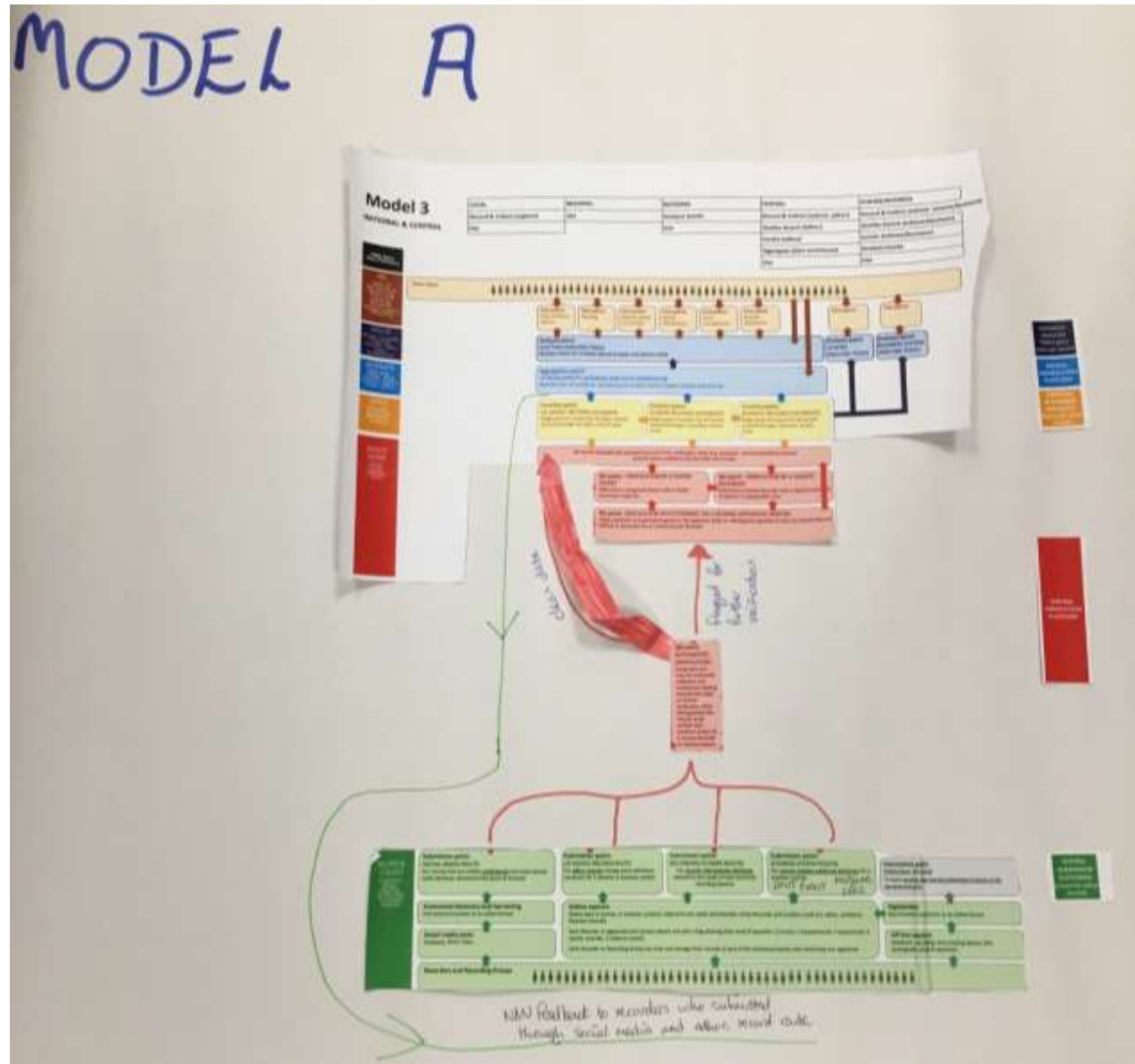
National: Good for potentially resources at national level and funding at national level. System would address national needs. But there would be duplication across the countries in the UK so potentially more bureaucracies. Cross board data submission may be an issue too.

Adapted data flow models

SESSION 6



6a) Adapted data flow models: Model A



Having considered the pros and cons of each model, and each stage of each model, workshop attendees agreed to reject the Current Situation model and to consider it no further. Working in groups, the workshop attendees then looked at how to create an improved model by using the stages in models 1-5 as jigsaw pieces that could be put together in any combination.

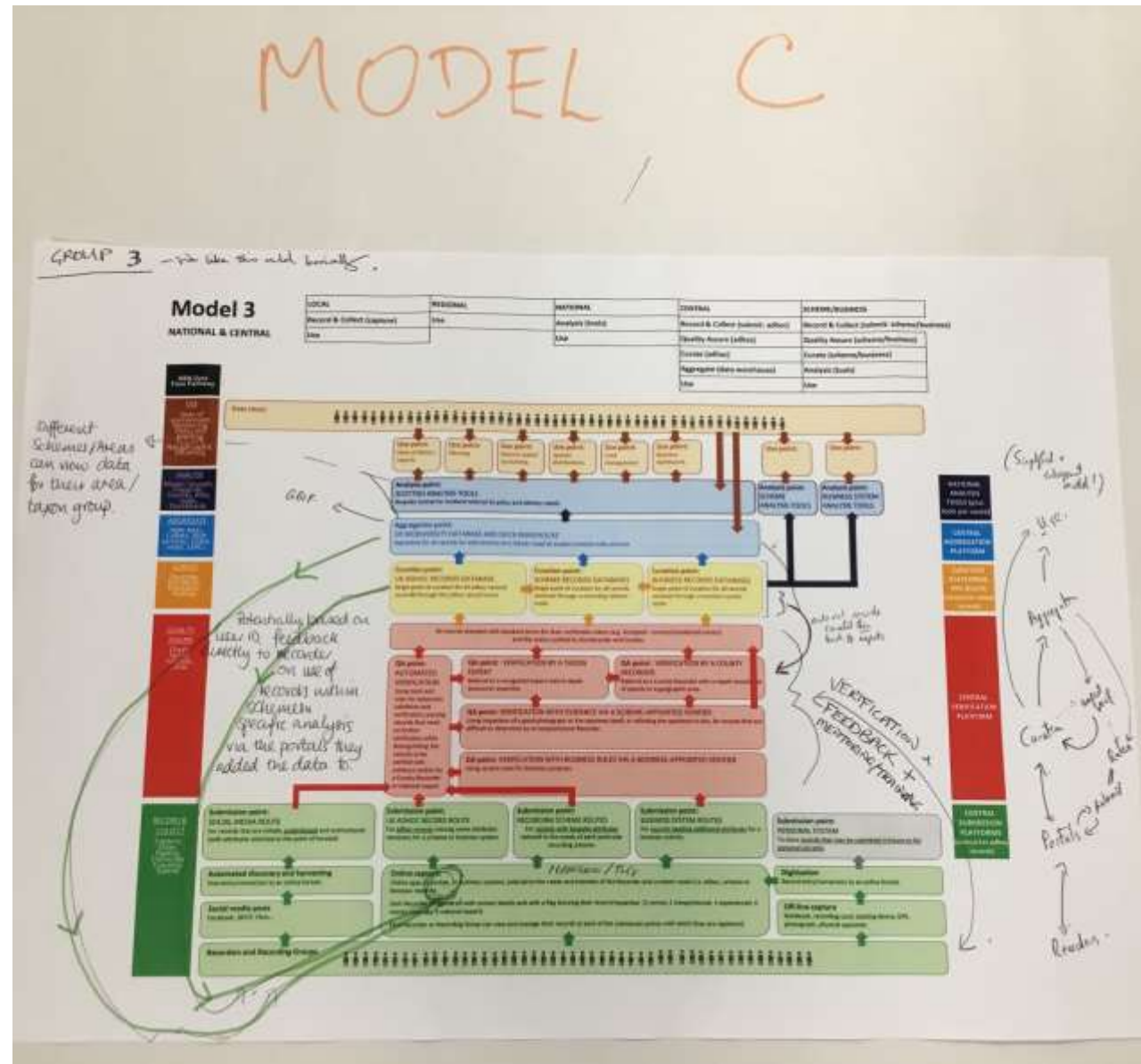
Group 1 preferred Model 3 and chose to adapt this model as the basis for their improved model. A feedback loop was added to recognise the importance of feedback on records received via either an adhoc or social media route.

6b) Adapted data flow models: Model B

Group 2 also preferred Model 3 but made a number of adaptations to show aggregation happening earlier, verification being an ongoing process, and the involvement of recorders more clearly in verification.



6c) Adapted data flow models: Model C



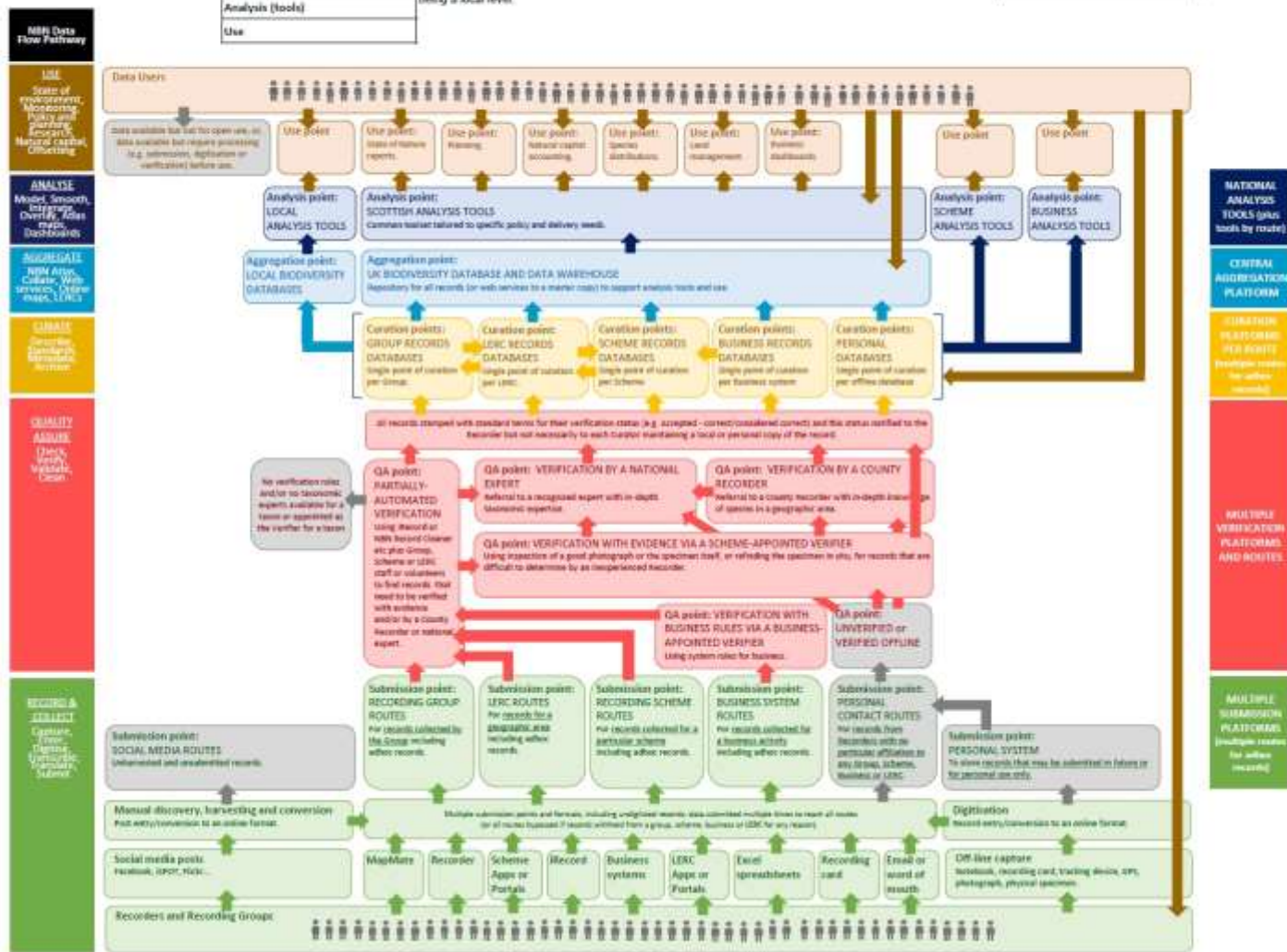
Groups 3 and 4 preferred Model 3 and did not adapt it substantially; however a simplified model was later proposed (see annotated drawing right).

6d) Model D – Current Situation for comparison



CURRENT SITUATION

LOCAL	REGIONAL	NATIONAL	CENTRAL	SCHEME/BUSINESS
Record & Collect (capture)	Use	Analysis (tools)	Aggregate (data warehouse)	Record & Collect (submit: adhoc/scheme/business)
Record & Collect (submit: adhoc)		Use	Use	Quality Assure (scheme/business)
Quality Assure (adhoc)	NB: the difference between Regional and Local is blurred and so for simplicity this diagram shows activities that may be regional as being a local level			Curate (scheme/business)
Curate (adhoc)				Analysis (tools)
Aggregate (data warehouse)				Use
Analysis (tools)				
Use				



Feedback on adapted models

SESSION 7

7a) Model feedback: for adapted models (rough notes from our discussion)

Group 1

Like Model 3 without chopping it up! Put on some feedback loops for additional detail – data need to flow back from certain points for example from curation point and aggregation point of UK database, you expect data to feedback to the capture portals at the bottom. There needs to be two way flow between verifiers and observers to do the verification job. The verification job is not just about quality control, it is also about mentoring, training and feedback. If done properly it is an excellent way of training that person and developing skills, so we put that in. We added on permissions, terms and conditions as an important term not to be forgotten, so as to ease data flowing through so don't lose this. We had an out arrow at the top going towards GBIF and other similar aggregations of data.

Group 2

Very similar comments to first table, focused on model three. Had some feedback from curation back to record and collect. Also discussed curation – the diagram cuts across horizontally but actually it happens all the way through so it is almost vertical so another feedback loop too. Could not agree on aggregation patterns as this was happening after curation, some felt this was ok, some felt it should happen before record and collect and quality assurance phases. We discussed roles and responsibilities in terms of who has access at various points in the process and who has access at what point of the process, so agree with first table.

Group 3

Settled on a version of model 3, did think radically about the simplest flow being straight in from online into the aggregation point but decided that the best option was that after collect & record, everything goes through automated verification and clean data goes straight into curation and aggregation points. Need to add feedback point from aggregation down to social media and adhoc as these people aren't getting support back from a scheme....

Group 4

Agree with other groups – we brought aggregation to an early stage as it has to happen early on so you can share the data out into the verification process; so we have curation closely associated with each submission path, then as soon as it gets aggregated, it goes through the verification path and gets an immediate label including needs to be verified; verification is an ongoing process that includes recorders. Verification can happen at an appropriate scale if it is aggregated as it can be served to those who need it. Recorders should be thought about in terms of analysis products. Recorders wrap around our data flow model. Hub that wraps around central pool!

7b) Model feedback: value and effort of adapted models (rough notes)

Group 1

Model A. Made certain assumptions based on what people have laid out. Assumed that A had more automated data checking, and implicated that there is less feedback. Because of this gave use and Analysis low value, and higher value for quality assurance. Based on what they saw in the model.

Model B. Started on this – perhaps biased as a result. Gave 5 across board as a Value and used as reference. Liked the speed of data, use and analysis. Early aggregation is valuable. Only one that got 5 in value across the board.

Model C. similar to A but were not assuming any reduced value for validation. Lower scores for aggregation and curation as happening at a later stage. Relatively high effort

Model D. Quite low scores for Value across board, and Effort was uniformly higher.

Group 2

Model A, B and C gave 5 across board for value as value same across all. But effort differs though similar. Though in model a there was a lower amount of effort in aggregate than other two, because of feedback from aggregators to curators. Model C and model A gave one for effort based on assumption that ... was automated. Model C got more effort for verification because verifiers would have more data coming though to verify. Model D – values were 2 or 3 for all 6 steps and effort was high for all of them. For use the value was 2 as data not used effectively and takes multiple effort to get data. Analyse 4 for effort as individual organisation developing own analyses tools rather than having suite available. Record and collect and quality assure has 5 for effort as volume of records high, backlog, feedback issues and lots of different ways in.

Group 3

Model A was pretty good so gave 4 for value for all, apart from use and analyse which got a 5. Models A and C came out pretty much the same as they were essentially doing the same thing in terms of data flow and not huge amount of difference between them – 4 for value and 2-4 for effort. Model B – by bringing curation and aggregation before quality assure kept value the same as first 4 stages but reduced the effort because curation easier to do before quality assure, and quality assurance easier because doing it on an aggregated set of data.

Group 4

Model D scored poorly – as to be expected, 2 and 3 for value. 4 and 5 for effort. Model A and C were similar and so scored identically. Noted against model C that there was some annotation around feedback and curation. Struggle with curation as starts at point of submission and sees data throughout. This is ownership and management of the data and hard to recognise in the data flow models. Quality assurance, wanted to recognise that feedback is important. Model B, maybe harsher than others but did score B highest. Felt that curation was not quite fixed. 5 for QA, aggregation and assuming benefits for analysis and use as a result of QA and aggregation. For record and collect scored the same as model C.

The preferred model

SESSION 8

8) Hearts and minds evaluation approach



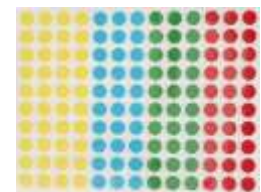
Value/Effort assessment

- Each group was asked to assess each of the Data Flow Pathway Stages within each model
- Assessment of the **Value** gained from each stage considered: benefits, efficiencies, data accessibility, data quality and speed of flow
- Assessment of the **Effort** involved to run each stage considered: people, services, governance, funding and technology
- Very High value and effort scored 5, High scored 4, Medium scored 3, Low scored 2 and Very Low scored 1
- The Value score of each stage was divided by the Effort score to give the V/E per stage for each model



'Hearts' assessment

- Each person was provided with 18 coloured dots (3 per stage) so that they could indicate which stage they preferred based on their own feelings and preferences.
- For each stage, each person could either place all three dots on the stage in one model, or one dot on the stage in each of three models.
- Only Models A-C were included as Model D was the 'current situation' which had been ruled out as a valid model.

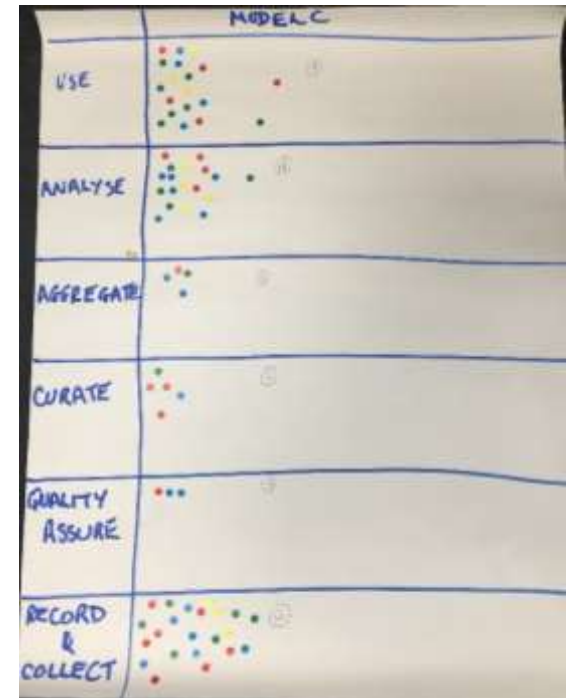
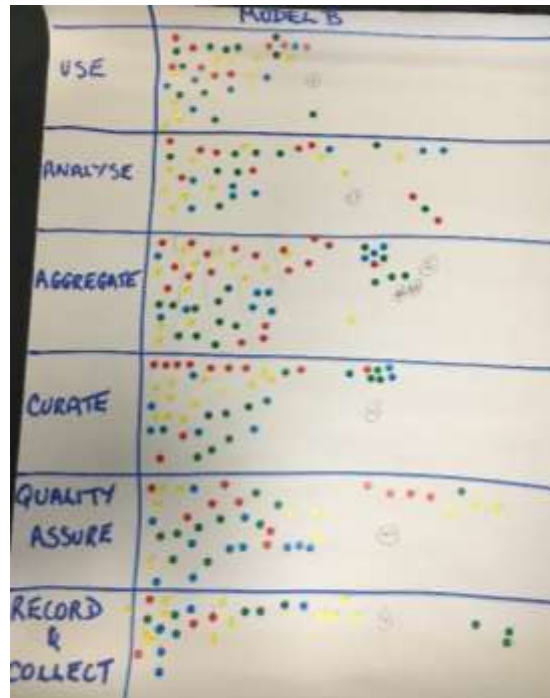
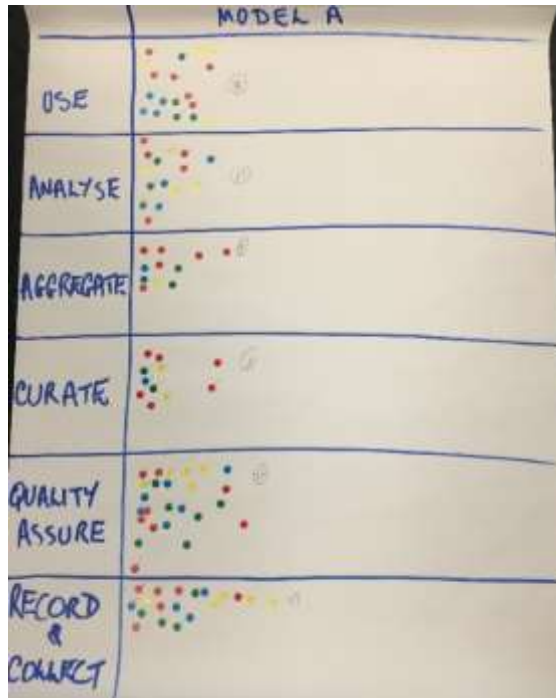


8b: Hearts and Minds Model Evaluation

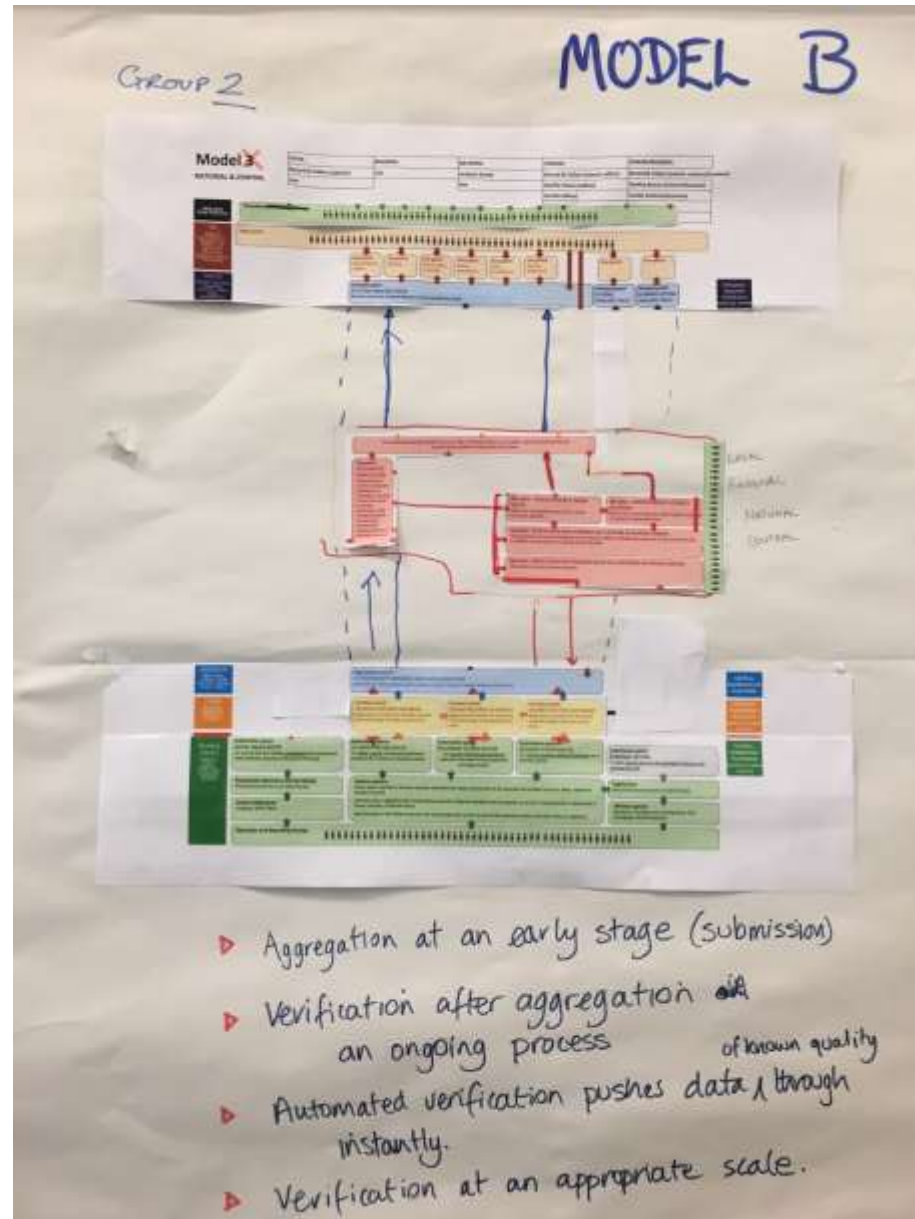
Model		A			B			C			D			BEST
Stage	V/E	♥	Σ	V/E	♥	Σ	V/E	♥	Σ	V/E	♥	Σ		
USE	9.7	18	27.7	10.2	36	46.2	10.2	19	29.2	3.32			B	
ANALYSE	8.3	15	23.3	8.8	37	45.8	8.8	18	26.8	3.08			B	
AGGREGATE	10.2	11	21.2	8.8	56	64.8	9.8	4	13.8	2.65			B	
CURATE	7.8	11	18.8	9.6	43	52.6	9.6	5	14.6	2.85			B	
QUALITY ASSURE	4.3	25	29.3	6.3	42	48.3	6.3	3	9.3	1.9			B	
RECORD & COLLECT	4.9	19	23.9	5.3	31	36.3	5.3	22	27.3	1.9			B	
Final scores														
Combined														
RANK														

V and E = 5 Very high, 4 High, 3 Medium, 2 Low, 1 Very low; ♥=3 dots/votes per person

8: Voting flip charts



8: The preferred Model: Model B



After assessing the value and effort of each stage for each model, and after voting to consider which stages felt most acceptable, Model B scored most highly and so is our preferred model.

The Workshop noted that Model B should also include a yellow 'point of curation' box for harvested social media records (or, at least a link to the 'curation box' for adhoc records).

The green strip illustrating recorders alongside the verification process shows the importance of liaison with recorders during verification.

Business changes needed

SESSION 9

8) Business changes needed to achieve Model B (rough notes)

RECORD AND COLLECT In the stop box a previous group had written, stop accepting non affiliated schemes, which we thought was harsh – there has to be a transition. Another group thought we should stop feeding data and records into schemes that don't share their data, maybe harsh again as it does take time to changes recorders thinking and opinion. In terms of start, we definitely think we need to feedback better to recorders and the feedback we are thinking about might include a rank of the records submitted, a detailed breakdown by species status, the dates when these species were task recorders in the area. It needs to be quick. Too much feedback is currently too late or non-existent. A couple of points about making records unique. Each record needs to have a unique identifier. Some groups also said each recorder needs to have a unique identifier. What should continue without change: encourage the use of standard recording protocols, arguably we should improve them and provide better guidance on the records we are after, to ensure they are complete records and people enter all the data that are required. We should continue to encourage data from any source including social media. Continue with changes: there needs to be clear guidance to the user about how the data will be used.

CURATE Stop – stop using local databases. Think that we added clarification to this as sounded a bit harsh to stop using local databases without some sort of transition phase. Had some questions around time delayed and protected sensitive species records (i.e. beaver records) – there needs a way to temporarily store some data and share later. Still questions around how this will work. Curation of local adhoc records – stop this adhoc curation of records. Start: social media harvesting and create a UK adhoc social media database. Link specimens with records. Start aggregating data from local databases. Continue without changes: continue to keep individual engaged. Continue with changes: maintain standards throughout systems. Maintain unique ID throughout process

Ellen Wilson: Local access to data is really important and should be maintained so 'removal' of local databases does not mean loss of local access as local access will be facilitated from the records being available online for all. Direction of travel is for this facilitation but it will take time to achieve this. We would encourage those who are doing the right things. The carrot is on doing the right thing. Hopefully the arrangements put into place will be so powerful and useful that people want to sign up and use it

AGGREGATE This had fewest suggestions. Stop reduce development of new aggregation databases, stop aggregating data from multiple capture points to organisations own databases. Start need a route from aggregation to verification. Automated flow needs to be developed from databases to aggregation point. Make a directory of schemes and organisations using databases which can be showcased. Continue with changes all uk data available on GBIF. Continue without changes all data with licences.

Andy Musgrove: automated verification happens best at point of data entry. Seems nonsensical for this to happen three stages down from data entry. Agreed- it happens at every stage. Vlad: QA is not one stage process, happens at all stages – bulk happens once you have a whole dataset and can filter etc to identify and bulk change mistakes. Oli G: Curation, and QA infiltrate everything – some QA happens at point of data entry, other points that need the full pot of data for the verification to work properly.

Recognition in room that QA is non-linear. Accepted that have parallel activities occurring.

ANALYSE Stop ... analysis should stop on local databases. Stop there being a reliance on people storing between quality offline copies of data. Start sharing best practise, new business analysis for all analysis tools – service analysis to understand audience to be able to develop right tools. Developing case studies and guidance for tools. Sharing tools for different scales that could be developed locally and shared elsewhere through national platforms. Finally, developing encouragement of tools to be open source on central platforms. Continue without change – available of scheme tools and business tools, communication of results. This also overlaps with things that might need change – feedback to recorders, more integration with other datasets. Changing in new tools, new Scottish analysis tools, local interpretation from a central dataset. The ability to download 'snapshots' for detailed analysis and the ability to share analysis that has already been done on the central platform. Rather than always hitting NBN server for all data requests, could keep a local copy that stayed in sync with NBN and stayed up to date.

8) Business changes needed to achieve Model B continued (rough notes)

USE In terms of start and stop this had a bit of a contradiction... stop the multiple access points for use. Start... having more access points! Take this as more access into the same point. Start use specific user reporting tools, but allowing individuals parameters to be set with those. Being able to distinguish between verified and unverified data and choose what you want to use. Something in here about acknowledgements and feedback to recorders – feedback to recorders on use of data, and ability to track between record used and original recorder. Access to data using web services. Ability to take data out of the system to use elsewhere. Start a mechanism for licencing use. Had discussion about how we licence for use, what we might want to licence, and what needs to be licences in a more controlled way. Huge list for continue with changes... being able to interpret data at different scales, addressing issues of data being in one place but being able to view it at the level you need. Guidance on sensitive data, guidance on tools. More and better metadata. Not really touched on metadata. More data on truly open licences. Empowering user to use data, and also acknowledging who we allow to use data and want we want them to do with the data. All about change.

QUALITY ASSURE Stop independent verification before aggregation. Stop blocking unverified records, just flag as needing verification. Stop sending lots of data to county recorders for verification – create more flexible tools for verifiers to manage their workloads... Start using image recognition where possible and using machine learning. Start validating and verifying data in an aggregated pot. Feedback outcome to curator and the recorders. Start public (?) register of verification experts. Start quality flagging every record. Start pushing records to verifiers automatically (with exceptions!). Increase the transparency of the process. Start quantifying the problem of verified records – use this as a call for increase funding and training. Start calling in the IT data wizards. Ensure a link between observation and specimens for ultimate quality. Create a centralised system to generate, manage and serve up auto-verification rules and start trying to gain wide spreads acceptable of verification terminology and vocab. Continue with changes invest in experts, training new experts. Train in ID skills and data management – though don't go hand in hand. May be useful to have these roles in different people. Continue bring people into verification fold. Improve auto-verification through data driven rules. Improve ruleset and governance. Improve verifier disagreement resolution. Continue to improve succession planning. Use verification process as means of upskilling and mentoring recorders. Continue without changes Support verifiers.

OTHER DISCUSSION

Licences – will they be needed at all to use the data, are the data not open in the model? Answer- it depends, some of the funding models may require licencing. It is unknown at this point. In terms of access and use, it is looking to be as open as possible. But some records do need licencing. It should be as simple and open as possible. Not fully shaped. Some recorders a bit wary of making their records available to everyone, especially if can be used for commercial purposes. Infrastructure needs funding, so there will need to be some T&C around licences where commercial. Data should be open at point of use. **Use level** – should you have single point of access or multiple point of access. This is still unclear which one you should have. This may tie into the services provided. Quite a few organisations around the table have their own current systems. There is an interesting question about how they will run in the future with any new system. Would there be a transition period. This is all to be dealt with in the detail – costings and efficiencies.

Business changes for RECORD & COLLECT

STOP

- Accepting from non-affiliated schemes
- Developing new routes for data capture; instead feed into central data curation/aggregation
- Feeding records into schemes that do not share their data
- Reinventing wheels

CONTINUE WITH CHANGES

- Clear guidance to user on licence/use of data
- Promotion and development of existing online reporting and recording tools
- Broaden training to other non-charismatic taxa
- Increase/promote online data entry
- Increase harvesting of records from museum/herbaria collections (better links to collection catalogues)
- Improve automatic feedback to Recorders (quicker, more detailed, mapped, broken down by species status (RDL, NR/NS etc), last recorded)

START

- Feedback on records that are new for a county etc
- Affiliated recording scheme register
- Unique record identifiers
- Support for affiliated schemes
- Coordinate/increase digitisation services
- Experiment with automated workflow (e.g. image recognition, DNA, sonar...)
- Harvesting records from social media
- Widget for recording via social media

CONTINUE WITHOUT CHANGES

- Encourage more use of standard protocols e.g. for data entry
- Continued encouragement of data from any source (including social media)
- Training for Recorders
- Feedback to Recorders is vital; Recorders will not put their records into a pot endlessly without getting some feedback/recognition

Business changes for QUALITY ASSURE

STOP

- “Independent” verification before aggregation
- Blocking un-verified records (just flag them as needing verification)
- Sending ‘shed loads’ of data to county recorders for verification/create more flexible tools for verifiers to manage their workloads

CONTINUE WITH CHANGES

- Increase investing in expert verifiers/determiners
 - Training in ID and in data management
- Accepting more people from community into the verification expert fold
- Further development of verification tools for aggregated data
- Improve auto-verification (using data-driven rules) both ruleset and governance improvement needed
- Manage recorder/user/verifier resolution
- Succession planning
- Use verification process as means of mentoring/upskilling recorders

START

- Use image recognition where possible/machine learning
- Validating/verifying data in aggregated pot
- Feedback outcome of verification to curator of master record and recorder [synchronisation]
- Public (?) register of verification experts
- Quality flag every record
- Push records to verifiers (automatic)
- Transparency of process
- Quantifying level/problem of un-verified records – use as call for increased training/funding etc
- Call in the IT/Data Wizards!
- Ensure link between observation and physical specimen for ultimate quality!
- Create centralised system to generate/manage/serve up auto-verification rules
- Gain a widespread acceptance of standard verification terms

CONTINUE WITHOUT CHANGES

- Training new experts
- Support verifiers

Business changes for CURATE

STOP

- Local databases (e.g. CMRs, LRCs)
- Local group databases
- Better offline databases (may be time delays to protect sensitive species?)
- Local ad-hoc record curation

CONTINUE WITH CHANGES

- Increase training/standardisation
- Ensure links about records are maintained through process and systems
- Increase resources and security of resourcing (sustainable funding model)
- Maintain unique ID for records throughout process

START

- Social media record harvesting
- Create ad-hoc UK database
- BA – to examine core attribution
- Looking at “random” collections (specimens)
 - Linking of records/collections
- Aggregate local databases
- Improve confidence in management of sensitive data
- Curating throughout the data model
- Sync curated data with aggregated database
- Community verification process

CONTINUE WITHOUT CHANGES

- Continue to involve groups/individuals. Keep engaged
- Keep up to date with taxonomic changes

Business changes for AGGREGATE

STOP

- Reduce the development of new aggregation databases
- Pushing records for expert verification pre-aggregation?
- Aggregating data from multi-capture points into your organisation's database

CONTINUE WITH CHANGES

- All UK data available on GBIF
- Don't detach curation/aggregation

START

- Recognition of a single aggregation point
- Need a route for data to flow from aggregation to verification [note this model suggest auto-verification happens after aggregation but not sure this is the case; it happens at the same time]
- Create system to harvest social media data
- Automated flow approaches to be developed from curated datasets to aggregation point
- Auto-validation before aggregation; auto-validate as records go into aggregation point
- Clear roles and responsibilities
- Make a directory of organisations/schemes/using web services so can be showcased!
- Pulling data from global systems (e.g. eBirds, iNaturalist, Observado)

CONTINUE WITHOUT CHANGES

- All data with licences

Business changes for ANALYSE

STOP

- Analysis on local databases
- Reliance on offline better copies of data

CONTINUE WITH CHANGES

- More integration with other environmental datasets (habitat, climate, soils etc)
- Training with new tools
- Scottish analysis tools – incorporating local analysis tools
- Incorporate national data into business analyses
- Improve comms re currently available tools and support
- Note: whilst analysis can now use centralised datasets and tools, interpretation of results will still benefit from local interpretation
- Recording schemes/groups/LRCs continue to analyse their data using tools and systems implemented by new model – and they will want to do so if the analysis capability is very powerful
- Ability to download snapshot for detailed analysis and potential to synchronise
- Feedback on analyses to recorders especially and other data providers

START

- Promotion/guidance for analysis tools e.g. case studies; sharing best analysis practice
- Enabling locally developed analytical tools to be shared elsewhere via a ‘national’ platform
- Facilitate R package to analyse data (data export) API services
- Tools developed based on service analysis; understand audience; BA [business analysis] on all analysis tools
- Sharing tools developed for specific purposes
- Encourage development of tools to be open source and on a single platform

CONTINUE WITHOUT CHANGES

- Scheme tools and business tools
- Widespread use of data for analysis
- Communication of results

Business changes for USE

STOP

- Multiple access points for use

CONTINUE WITH CHANGES

- Better presentation tools e.g. for mapping and looking at data within SSSIs for example
- Promote availability of different use points
- More guidance on tools/downloads available
- Develop download tools
- Further develop access to sensitive data (with agreement and clarity)
- More data on truly open licences
- Interpretation at different scales (local, regional etc)
- More data products – e.g. interpretation of data – and allow people to make their own
- More and better metadata

START

- More access points
- Develop use-specific reporting tools where you can add individual parameters
- More explicit choice on use of verified and unverified data (including sub-divisions e.g auto-verified data)
- Better acknowledgement of data sources to dataset and individual record level
- Access to data using web services (e.g. Ias, FES)
- Publicising the benefits of the data for the sectors e.g. showcase projects which use the data
- Feedback to recorders/schemes/groups how their data are being used
- System to enable tracing of record used back to the org/recorder
- Mechanism for licencing use

CONTINUE WITHOUT CHANGES

- Users' desire for data
- Multiple uses

Workshop feedback and vision

SESSION 10

Our vision for improved data flows

So that we could draft a '100 word vision' of the preferred model for data flow, and the changes needed to achieve it, we brainstormed words that could be included:

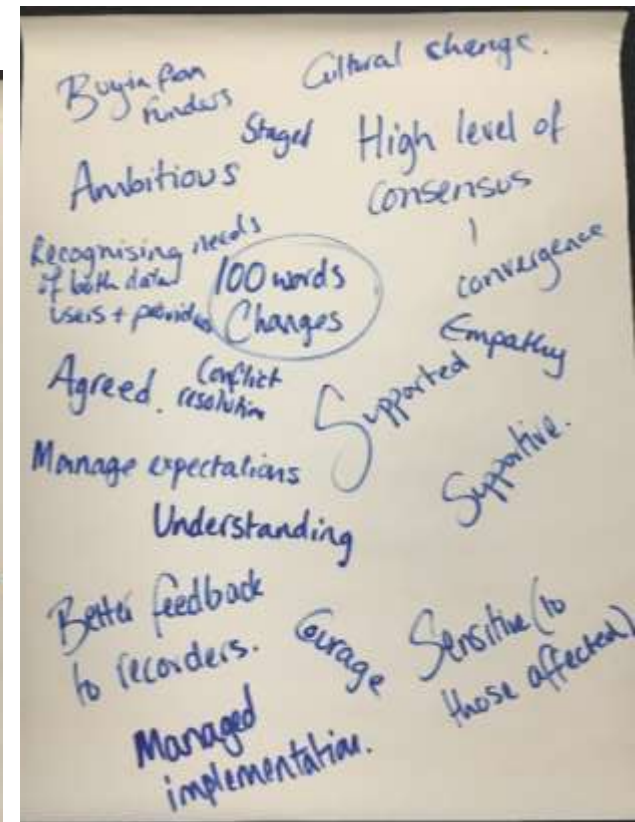
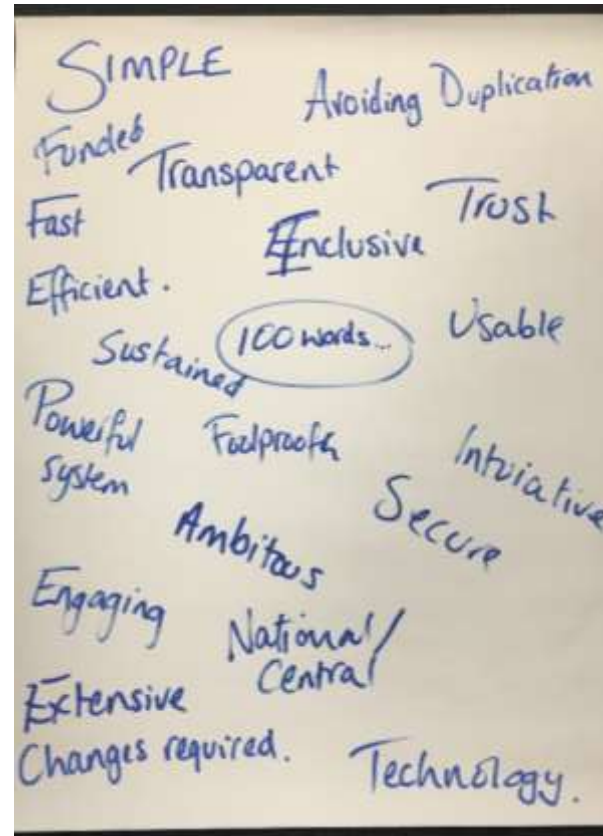
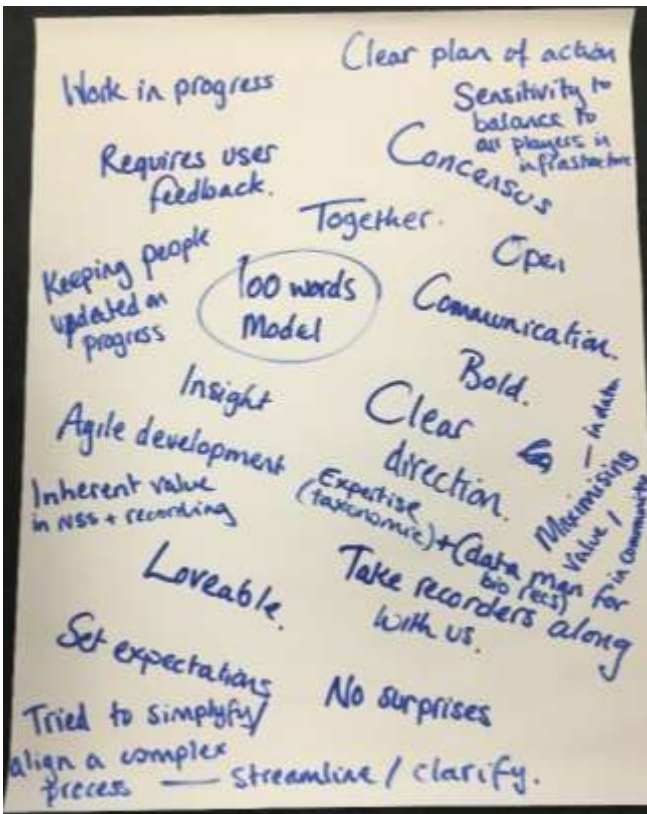
On the transition needed:

- Buy-in from funders
- Ambitious
- Staged
- Recognising needs of both data users and providers
- Agreed
- Conflict resolution
- Manage expectations
- Understanding
- Better feedback to Recorders
- Managed implementation
- Courage
- Supported
- Supportive
- Empathy
- Sensitive (to those affected)
- High level of consensus – convergence
- Cultural change

On the model:

- Simple
- Funded
- Fast
- Efficient
- Transparent
- Sustained
- Inclusive
- Avoiding duplication
- Trust
- Usable
- Powerful system
- Ambitious
- Engaging
- Extensive changes required
- National/Central
- Technology
- Secure
- Foolproof
- Intuitive
- Work in progress
- Requires user feedback
- Insight
- Keeping people updated on progress
- Agile development
- Inherent value in NSS and recording
- Lovable
- Set expectations
- Tried to simplify/align a complex process
- Streamline/clarify
- Clear plan of action
- Sensitivity to balance in players in infrastructure
- Consensus
- Open
- Together
- Communication
- Bold
- Clear direction
- Taxonomic and data management expertise
- Take Recorders along with us
- No surprises

10: Brainstorm flipcharts



Workshop feedback

- Creativity in tasks was really appreciated
- Getting some of the material out earlier so time to absorb and reflect and people can be more prepared
- Need a bit more time on tasks – frustration when hurried on
- Split tasks and put tea break in the middle – time to think and discuss over tea and come back fresh
- Copies of presentation, hand out and resources for attendees (also will be on NBN website)
- Safe space really worked well – strong team working

Ellen Wilson: everyone pulling in same direction so workshop has been really productive. Everyone here can show leadership in their sector - pulling in the same direction should make implementation effortless.