DIYALA PROJECT

Clemens D. Reichel

At some stage, every successful project develops the momentum that will carry its idea towards success beyond the point of no return. I believe the Diyala Project successfully passed this point during this year.

Is it surprising to hear that occasionally one doubts whether a project can be successfully completed? Perhaps the Diyala excavators themselves had doubts about the feasibility of publishing the ambitious final volume of their work, the *Miscellaneous Objects from the Diyala Region*. Once they had completed their excavations at Tell Agrab, Tell Asmar, Ishchali, and Khafaje, the excavators started a large-scale publication project, in which most of the architecture and some of the key finds, such as the sculpture, seals, and pottery, were published in Oriental Institute Publication volumes. Despite numerous attempts, the publication of over 15,000 items found during the Diyala expedition never went past a planning stage until 1994, when the Diyala Project was launched by McGuire Gibson. Even this project encountered a major crisis last year, when news reached the Oriental Institute that the Iraq Museum had been looted, leading us to believe that most of the priceless objects that we had hoped to study in person in Baghdad had been irretrievably lost. Could our project still be brought to a meaningful conclusion? One-and-a-half years later we know that almost all the 600 seals from the Diyala excavations assigned to the Iraq Museum were stolen, and the presence of most of the other 8,000 Diyala objects in the Iraq Museum largely remains to be confirmed. In the end, however, this tragedy only hardened our resolve to get all of the Diyala materials published online as soon as possible.

Another more positive lesson we learn again and again is that, in a project of this size, it often takes a new person with a new idea to get things going again. By 1994, when McGuire Gibson took on the responsibility of publishing the Diyala objects, the idea of publishing them in a book had essentially been declared dead — the large number of objects (15,000) would have required an immense number of drawings and photographs for adequate illustration, making a paper publication prohibitively expensive. Gibson therefore decided to launch the project as a computer-based relational database. Initially we intended to put the data out on CDs, but with the rapid development of the World-Wide Web we later on opted for a Web-based dissemination of the data. The Web-based database remained more of a vague idea than a concrete plan until 2000, when George Sundell joined our project as a volunteer and data architect (fig. 1). Four years later we are well into executing this plan, but George's input has also started to modify the principal layout of the project. A data architect in his professional life, data management was George's bread and butter, and over the next two years he examined the relation-

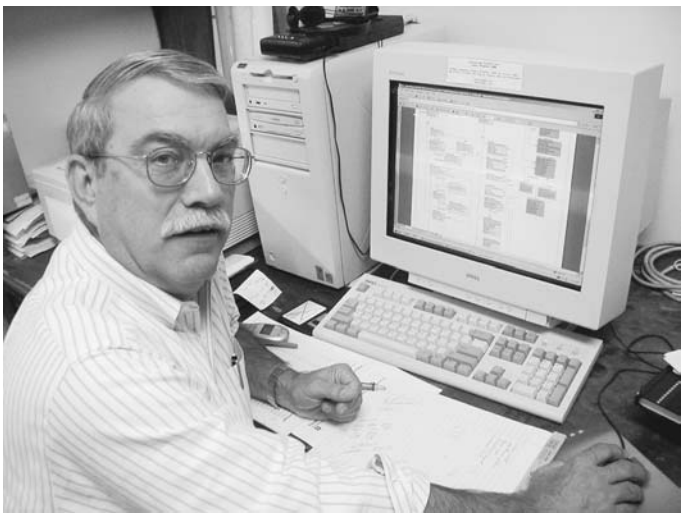
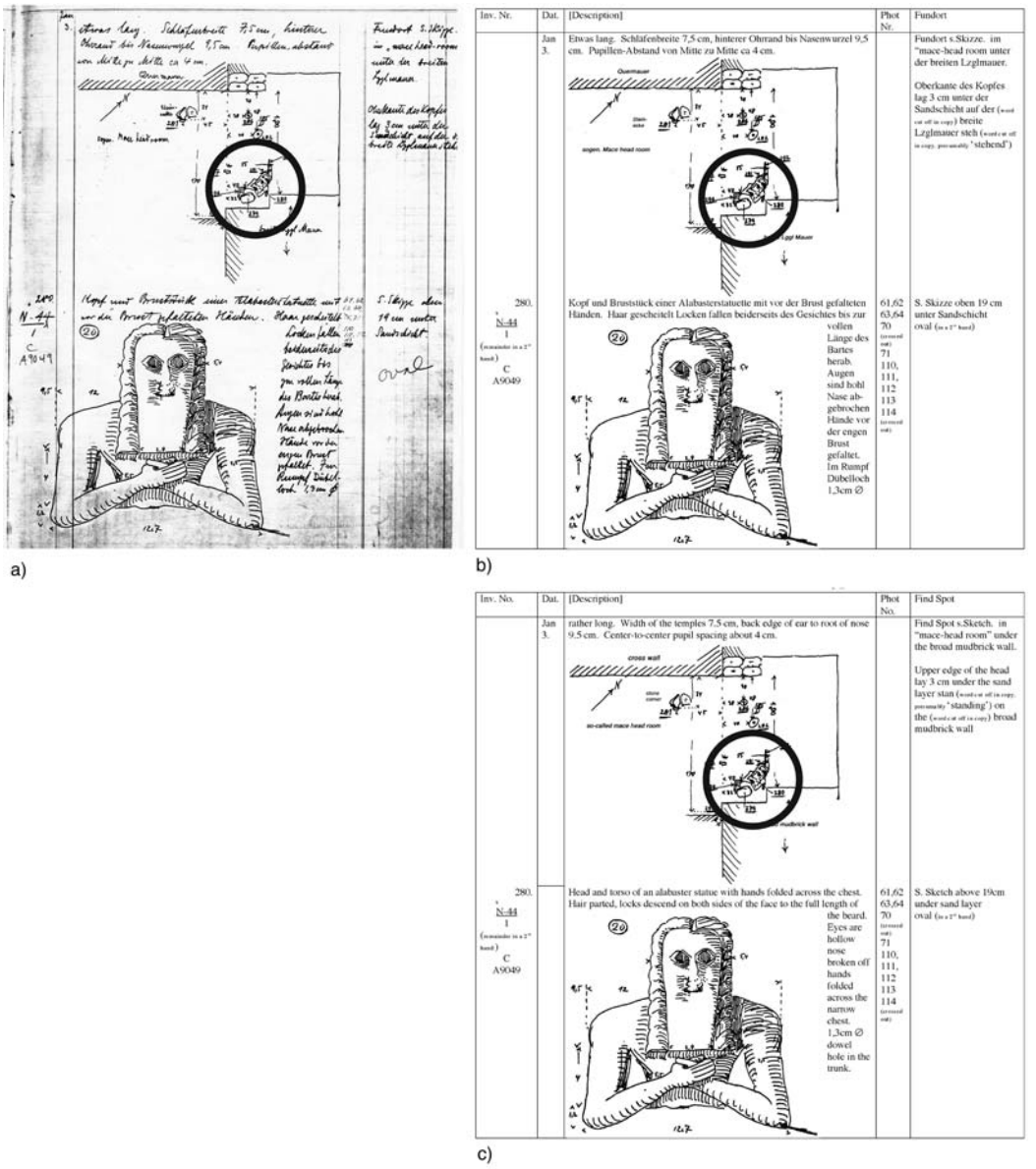


Figure 1. Another table done! George Sandell, data architect for the Diyala Project, checking another item on the database load plan

ship of information elements defined by us and drastically refined the layout of our data structure. One of the weak points in our database was the absence of systematic source indicators. In other words, while we provide a full list of both published and unpublished sources for each object, it was not clear from which source a certain descriptive element had been taken and whether or not it had been verified for accuracy. In George's new design we distinguish between *primary* and *interpretive information*. By *primary information* we refer to sources such as field diaries, object registers, and catalog cards with descriptions given by the excavators, an example of which is shown in figure 3. Whether the information in them is correct can be verified in some cases, for example, when we have a chance of physically reanalyzing an object described in such a source. Other cases, however, such as a description of the archaeological context in which an object was excavated, cannot be verified independently. *Interpretive information*, by contrast, reflects our own work. This information can be updated and improved upon as needed. These two information sources therefore require very different treatments in a database. Our own *interpretive* descriptions are stored as computer-based entries; should any ambiguity or typo be found, it can be fixed with the approval of the project director and the responsible collaborator. The *primary* data from the excavators, however, is a very different matter. This data is on paper, most of it in handwritten notes. Frequent use of abbreviations and incorrect use of words and phrases (the native tongues of the Diyala excavators included Danish, Dutch, English, German, and Ukrainian among others) make a data transfer by typing highly interpretive and subject to mistake. Some notes were kept in languages other than English. Conrad Preusser, the site director during the initial 1930/31 season at Khafaje, for example, kept all his notes, including the object register, in German; subsequently, he translated many of his notes into English, but frequent idiomatic mistakes can be found in his translations. A direct transfer of excavators' notes into text files is complicated further by the fact that notes were often revised and annotated using colored pencils; in many cases different handwritings can be discerned, making it clear that more than one mind was at work. Throughout the years we found a number of these changes to be unsustainable or wrong (some of them even made it into print), so we deemed it vital to record both the original entry and how it was subsequently changed. To complicate matters even further, many notes were enhanced with sketch drawings that were impossible to render in a textual description. We tried to accommodate all these idiosyncrasies in a more and more refined layout, but whenever I encountered an ambiguity I eventually found myself checking the original source, not relying on our own database entry. It finally became clear to us that neither we nor anyone else could really do analytical work without seeing the original notes. This is when the idea of a "Virtual Diyala Archive," described in more detail in previous *Annual Reports*, was born. Instead of typing up every word we decided to index these sources by searchable key words, scan the original records, and make them accessible through the Web-based database. End users would therefore have the same access to published and unpublished records that we enjoy at the Oriental Institute, draw their own conclusions, and be in a better position to question our interpretations. George began implementing the structural changes in 2002. By early 2004, our database was ready to accept these additional materials.

Simply talking about "scanning" data, however, is insufficient. As indicated above, many annotations were made in color. If we really wanted to make consultations of the paper originals redundant we had to scan all the notes in color. The scanning resolution turned out to be an even more difficult issue to resolve. To determine the right resolution we first had to decide the ultimate purpose of the scan. A screen display image (that is, 72–96 dots per inch [dpi]) is at a fairly low resolution, which would require very little time to capture. Since online dissemination of this data was the primary objective, we could have settled for a low resolution, but these scans



a)

b)

Inv. Nr.	Dat.	[Description]	Phot Nr.	Fundort
	Jan 3.	Etwas lang. Schläfenbreite 7,5 cm, hinterer Ohrrand bis Nasenwurzel 9,5 cm. Pupillen-Abstand von Mitte zu Mitte ca 4 cm.		Fundort s.Skizze. im "mace-head room unter der breiten Lzglmauer. Oberkante des Kopfes lag 3 cm unter der Sandschicht auf der (word cut off in copy) breite Lzglmauer steh (word cut off in copy, presumably 'stehend')
280. N-44 1 (remainder in a 2" box) C A9049		Kopf und Brustsitück einer Alabasterstatuette mit vor der Brust gefalteten Händen. Haar gescheitelt Locken fallen beiderseits des Gesichtes bis zur vollen Länge des Bartes herab. Augen sind hohl Nase abgebrochen Hände vor der engen Brust gefaltet. Im Rumpf Dübelloch 1,3cm ∅	61,62 63,64 70 (erased out) 71 110, 111, 112 113 114 (erased out)	S. Skizze oben 19 cm unter Sandschicht oval (in a 2" box)

c)

Inv. No.	Dat.	[Description]	Phot No.	Find Spot
	Jan 3.	rather long. Width of the temples 7.5 cm, back edge of ear to root of nose 9.5 cm. Center-to-center pupil spacing about 4 cm.		Find Spot s.Sketch. in "mace-head room" under the broad mudbrick wall. Upper edge of the head lay 3 cm under the sand layer stan (word cut off in copy, presumably "standing") on the (word cut off in copy) broad mudbrick wall
280. N-44 1 (remainder in a 2" box) C A9049		Head and torso of an alabaster statue with hands folded across the chest. Hair parted, locks descend on both sides of the face to the full length of the beard. Eyes are hollow nose broken off hands folded across the narrow chest. 1,3cm ∅ dowel hole in the trunk.	61,62 63,64 70 (erased out) 71 110, 111, 112 113 114 (erased out)	S. Sketch above 19cm under sand layer oval (in a 2" box)

Figure 2. Digital clean-up: Example of an entry from the Khafaje field register: (a) the original by Conrad Preusser, (b) cleaned and retyped by Robert Wagner in German, and (c) translation into English. The upper sketch renders the find context in which the piece of sculpture, sketched below, was found

would have had severe limits. What if, for example, someone requested a printable version of this scan? Printable images require a much higher resolution than screen displays — at least 300 dpi with present-day printers. Some records also required enlargements to be readable or in any way usable, which again required a higher a scanning resolution. With screen and possibly printer resolutions likely to increase in the next years, we wondered how permanent a record we would create if we scanned these records at the lower end of possible resolutions. To create a virtual facsimile of the original record we had to think past today's requirements — there seemed no point in doing this now if someone had to do it all over again in ten years! The need to define *archival standards* became all too clear. I will not draw out the technical discussion

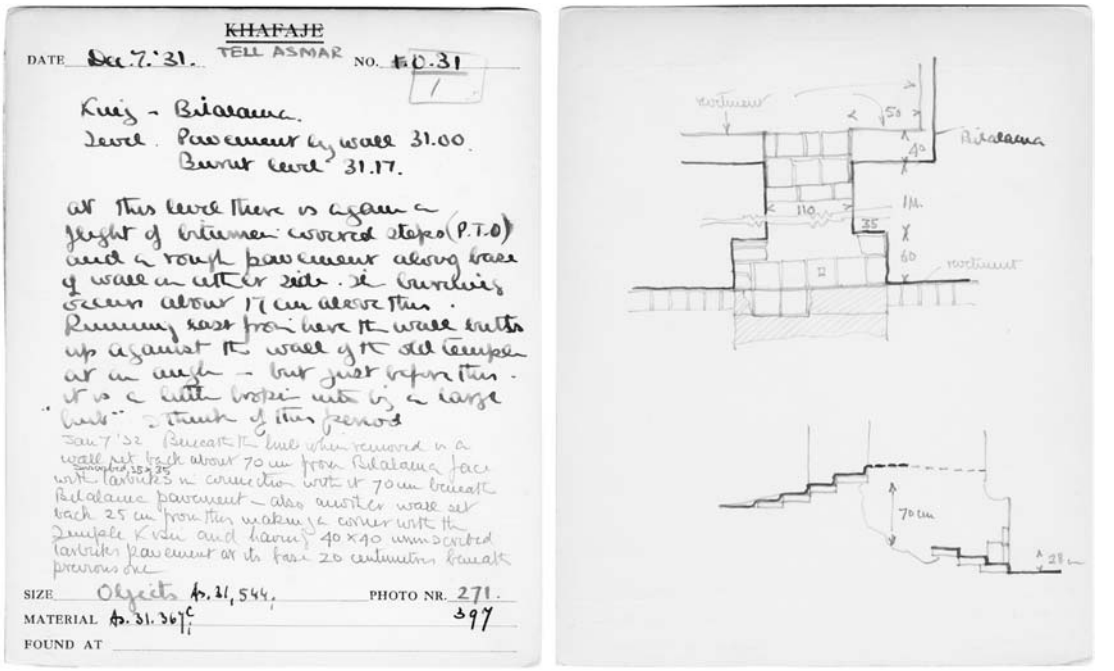


Figure 3. Example of a locus card for the Palace of the Rulers at Tell Asmar, showing sketches and successive additions and annotations

any further, but in a nutshell, we decided to scan all records at a standard resolution of 400–600 dpi in color. Files would be saved in a lossless, uncompressed format, Tag Image File Format (TIFF). Smaller versions would be created for display on screen in the database, while retaining the larger files for archival purposes.

We had solved the theoretical issues. Now we had to face the practical ones. Scanning all the field notes, object cards, and plans is an enormous undertaking. With records still turning up every now and then in unsuspected places, it remains difficult for me to put down a final figure for the total number of scans necessary to accomplish our task, but I estimate that it may well exceed 100,000! This creates enormous logistical problems — a 400 dpi color scan of an average Diyala catalog card is about 15 megabytes in size. Many of them have annotations on both front and back. Mathematically, if every scan were the size of a catalog card (and many of them are much larger) we would have to store 3,000 gigabytes (3,000,000,000,000 bytes) worth of data. To back up all this data on CDs, we would have to burn 4,615 CDs, or 9,230 to have an essential second copy! It was clear that we faced a data storage crisis. But leaving issues of required hardware, software, and data storage aside, who was going to do all of this work, and where?

Once more, the right person showed up at the right time. One day in spring I received a phone call that put things in motion. Robert Wagner had worked as a translator of German to English, mostly dealing with industrial patents and, following his retirement, was looking for new things to do. Having someone else on board aside from myself who reads German was an exciting prospect, especially since it offered the possibility for me to avoid dealing with the dreaded German Khafaje object registers mentioned above. I decided to put Robert's skills to the test and asked him to enter Conrad Preusser's notes, kept in beautiful but sometimes almost indecipherable handwriting. In short, Robert mastered the job — in addition to the original field registers, we now have versions of these registers typed up in German as well as in English (fig. 2b–c).

Robert faithfully recreates the original layout of the registers, adding their sketches by scanning and annotating them. His entries have been added to our database in a searchable form, but we will also add his translations as Portable Document Format (PDF) files to our site. Having finished this work, I told him about our scanning plans, and he readily agreed to take it on. Calling someone "compulsive" seems to have developed a negative undertone, but I would be hard-pressed to find a better term to describe Robert's attention to detail. After a period of experimentation on which devices to use to hold documents in place on the scanner and what backgrounds or which color profiles to use, he got started and has been working on it since spring. By last count he has already created some 16,000 scans so far, equivalent to about 10,000 catalog cards! Since May, Robert has been joined by Karen Terras in the scanning efforts. A veteran of the Iraq Museum Project, she joined the Diyala Project and has worked through field diaries and pottery sheets. By July, both volunteers had created about 250 gigabytes of data to be stored. The storage crunch was alleviated with the purchase of a DVD burner. DVDs can hold up to 4.7 gigabytes of data each, which reduces our storage problem significantly, even if the number of DVDs to be burned remains sizable. Having seen the evolution of data storage devices from 800 kilobyte double-density disks, to 100 megabyte zip disks, 650 megabyte CDs, and now 4.7 gigabyte DVDs, I am confident that our storage crunch is a temporary one, soon to be made irrelevant by technological innovation.

Figure 4. Vessel bearer from Tell Agrab, of limestone, dated 2600 B.C.

Eventually Robert (and maybe others as well) will tackle another challenge — the scanning of the original Diyala large-format negatives. We already have scans of contact prints from many of these negatives, made by our volunteer Joyce Weil between 1996 and 1998 on a flatbed scanner. A few test scans of the negatives, however, have shown us how much better the quality and resolution of scans would be from the negatives. These negatives are large (5×7 inches), but many of them contain more than one object. Modern scanners would allow us to scan them at a very high resolution (for example, 2,000 dpi), which would produce printable high-quality close-ups of even small objects. Many of these objects had been allocated to the Iraq Museum and might have disappeared or been damaged during the museum looting of 2003. Getting as good an image of them as possible seems of vital importance. Once more, data storage is going

to be a challenge — a scan of a 5 × 7 inch black-and-white negative at 2,000 dpi creates about a 135 megabyte TIFF file. With about 8,000 negatives to scan, data storage is once again going to be a challenge.

Quite obviously, one of our most essential needs for the next few years boils down to money — money to buy faster computers, more external hard drives and DVDs, and to hire graduate students to add, edit, and review data. While writing this report I have received the joyful news that we have been awarded $100,000 under the National Endowment of Humanities "Preserve Iraq's Cultural Heritage" initiative, to be paid out over the next two years. This grant will help us to get the necessary computer upgrades and storage devices and to hire at least one student as a helper. To expand and maintain the site, however, we will have to find other long-term funding sources. After the site is launched (we hope by summer 2006), work will go on for years. We will continue to add and edit data and mark up building plans with object findspots to allow interactive data queries for spatial analyses. And finally, we will add new photographs of objects accessioned by the Oriental Institute Museum that have been taken over the past few years. Between 1997 and 2002, about 7,000 photographs were taken on black-and-white film by our volunteer Betsy Kremers and myself; these negatives have already been scanned at a high resolution. Since May 2003 we have been working with a Nikon digital SLR camera, which has allowed us to add color photographs, obtain close-up shots and details of many objects with its 1:1 macro lens, and enjoy the added benefit of instant quality control (fig. 4).

Finally, I want thank those donors who have supported us so loyally in the past. The bad news about a computerized database project is that it is never done, but the good news is that it can always be improved upon. I hope that the standards chosen by us will keep the Diyala database — in the face of an ever-developing computer world — a viable research and publication tool for a long time to come.